

Article

Diversity and Functional Genome Analysis of Lactic Acid Bacteria Isolated from Trás-os-Montes Artisanal Alheira

Nathalia Fernandes ¹, Alessandra De Cesare ², Valentina Indio ², Ursula Gonzales-Barron ¹
and Vasco Cadavez ^{1,*}

¹ Mountain Research Center (CIMO), Associate Laboratory for Sustainability and Technology in Mountains Regions (LA SusTEC), Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal; nathalia@ipb.pt (N.F.); ubarron@ipb.pt (U.G.-B.)

² Department of Veterinary Medical Sciences, Alma Mater Studiorum, University of Bologna, Ozzano Emilia, 40064 Bologna, Italy; alessandra.decesare@unibo.it (A.D.C.); valentina.indio2@unibo.it (V.I.)

* Correspondence: vcadavez@ipb.pt

Abstract

Alheira is a traditionally smoked, fermented, non-ready-to-eat meat sausage produced in the Trás-os-Montes region of Portugal, where lactic acid bacteria (LAB) are major determinants of product safety and quality. This study used whole-genome sequencing to characterize 59 LAB isolates collected from artisanal alheira produced in six municipalities. Eight species belonging to six genera were identified, and average nucleotide identity analysis resolved 24 non-redundant strain groups. Functional annotation revealed broad repertoires of carbohydrate-active enzymes, proteases, lipases, transport systems, and genes associated with tolerance to acid, oxidative, bile, heat, and salt stress. Genotype–phenotype inference was performed with a subset of 22 strain groups represented by unflagged genome assemblies. After false-discovery-rate correction, carbon-metabolism, transport, energy-metabolism, total functional-gene, acid-stress, bile-stress, and total-stress counts were positively associated with total pH decline. A mixed-effects model showed a positive association between standardized total functional-gene count and acidification ($b = 0.137 \pm 0.059$, $p = 0.049$; marginal $R^2 = 0.266$, conditional $R^2 = 0.582$), but within-between decomposition indicated that the association was detectable between species rather than within species. Regularized regression, Random Forest, AICc model comparison, and partial least-squares regression converged on carbon metabolism, transport, energy metabolism, and acid- or bile-stress functions as the principal exploratory signals; however, sparse LASSO regression for variable selection was unstable, and prediction of unobserved species was limited. None of the five targeted pathways: acetaldehyde, citrate, diacetyl, exopolysaccharide, and lactose metabolism, remained significant after multiple-testing correction. These findings provide a genomic and phenotypic basis for selecting native starter-culture candidates while emphasizing the need for independent functional and safety validation.

Keywords: lactic acid bacteria; whole-genome sequencing; fermented sausage; acidification; starter culture; food safety



Academic Editors: Talha Demirci and Enes Dertli

Received: 30 July 2026

Revised: 26 August 2026

Accepted: 27 August 2026

Published: 7 September 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Alheira is a traditional smoked and fermented meat sausage produced in the Trás-os-Montes region of northeastern Portugal, where it holds significant cultural and economic importance as part of the regional gastronomic heritage [1,2]. Its production involves the

mixing of cooked meats (e.g., mainly pork and poultry) with bread, spices, and condiments, stuffed into intestine casings, and subjected to curing via a cold-smoking process followed by drying [1,3]. Notably, artisanal alheira is produced without the addition of starter cultures, meaning that fermentation is driven entirely by the indigenous microbiota naturally present in the raw ingredients and production environment [4]. Lactic acid bacteria (LAB) are the dominant microorganisms in alheira and play a central role in determining its safety, shelf life, and sensory characteristics [4,5].

Despite its cultural significance, alheira presents considerable challenges for the food industry in terms of microbiological safety and shelf life extension. As a high-moisture product with a water activity typically exceeding 0.98, raw alheira sold at retail may be susceptible to the growth of foodborne pathogens, including *Listeria monocytogenes*, *Salmonella enterica*, and *Staphylococcus aureus* [1,6]. To address these challenges, the processed meat industry has traditionally relied on freezing, vacuum packaging, and the addition of chemical preservatives, most notably nitrites, to extend the shelf life of the raw product. However, all these methods have limitations, especially the use of nitrites in processed meat, which has come under increasing scrutiny following the classification of processed meat as a Group 1 carcinogen by the International Agency for Research on Cancer (IARC), with nitrite-derived N-nitroso compounds identified as key contributors to colorectal cancer risk [7].

In response to this challenge, there is growing pressure from consumers, industry, and regulatory agencies worldwide to develop clean-label preservation strategies, that is, solutions that extend shelf life and ensure food safety without the use of synthetic chemical additives [8]. This movement is not limited to the European markets, but reflects a global trend, with regulatory agencies across North America, Asia, and beyond increasingly incentivizing the reformulation of processed meat products to reduce or eliminate chemical preservatives. In this way, LAB continues to represent one of the most promising natural alternatives, given their capacity to promote fermentation, as well as their long history of safe use in foods [4,9,10].

Recent advances in food microbiology have also shifted LAB research from descriptive screening toward genome-informed prediction of industrially relevant traits. The rapid expansion of whole-genome sequencing datasets has enabled genotype–phenotype association analyses and microbial genome-wide association studies (GWAS), which are increasingly being used to identify the genetic determinants of fermentation performance, carbohydrate utilization, and stress adaptation in food-associated bacteria [11,12].

Despite these advances, it remains unclear whether variation in functional genomic content can quantitatively explain technologically relevant phenotypes of indigenous LAB from traditional fermented foods, particularly after accounting for the strong taxonomic structure of bacterial genomes. For alheira, previous work characterized the identity and technological phenotypes of indigenous LAB, but the relationship between strain-level functional genomic diversity and fermentation performance has not yet been systematically evaluated. The present study extends our previous work [4] by (1) characterizing the functional and safety-related genomic profiles of LAB from alheira using whole-genome sequencing (WGS), and (2) integrating these genomic predictions with previously measured phenotypes, including proteolytic activity, lactic acid production, antimicrobial activity, and pH kinetics, to evaluate genotype–phenotype relationships relevant to starter-culture selection.

2. Materials and Methods

2.1. Strain Isolation

A total of 59 LAB isolates previously recovered from artisanally produced alheira sausages were included in this study. The isolates originated from six municipalities in

the Trás-os-Montes region of northeastern Portugal: Mirandela ($n = 20$), Valpaços ($n = 16$), Vinhais ($n = 12$), Vimioso ($n = 6$), Bragança ($n = 4$), and Mogadouro ($n = 1$). The LAB isolates used in this study are a subset of those reported in [4], where isolation procedures and initial phenotypic characterization are described in detail.

Briefly, the isolates were selected on the basis of previously measured antimicrobial activity against *Listeria monocytogenes* WDCM 00019, *Salmonella enterica* serovar Typhimurium ATCC 43971, and *Staphylococcus aureus* ATCC 6538. These indicator strains were the reference strains used in the previous phenotypic screening of the isolate collection [4], thereby allowing direct continuity between the earlier technological characterization and the present genomic analysis. The selected isolates were maintained at $-80\text{ }^{\circ}\text{C}$ in 50% (v/v) glycerol until further analysis.

2.2. Sample Preparation

Each isolate was plated on the medium used for primary isolation (MRS or M17 agar) and incubated anaerobically at $30\text{ }^{\circ}\text{C}$ for 48 h. A single colony was resuspended in 4 mL of Brain Heart Infusion broth and incubated overnight at $30\text{ }^{\circ}\text{C}$. Cells were harvested by centrifugation, and the resulting pellets were stored at $-20\text{ }^{\circ}\text{C}$ until DNA extraction.

2.3. Genomic DNA Extraction

Genomic DNA was extracted using the MagAttract HMW DNA Kit (QIAGEN, Bologna, Italy), following a workflow optimized for high-molecular-weight DNA recovery. Briefly, pellets were resuspended in Buffer P1 (50 mM Tris-HCl, pH 8.0; 10 mM EDTA). Cell-wall digestion was performed with lysozyme (100 mg/mL) at $37\text{ }^{\circ}\text{C}$ and agitation at 900 rpm for 1 h, followed by proteinase K digestion at $56\text{ }^{\circ}\text{C}$ for 30 min and RNase A treatment (100 mg/mL). DNA was bound to MagAttract Suspension G magnetic particles in the presence of Buffers AL and MB, washed sequentially with Buffers MW1 and PE and twice with distilled water, and eluted in Buffer AE. MagAttract Suspension G and Buffers AL, MB, MW1, PE, and AE were supplied as components of the MagAttract HMW DNA Kit (QIAGEN, Bologna, Italy). Vigorous pipetting was avoided to minimize mechanical shearing. DNA quality and purity were assessed spectrophotometrically using the 260/280 ($\approx 1.8\text{--}2.0$) and 260/230 ($\approx 2.0\text{--}2.2$) absorbance ratios.

2.4. Library Preparation and Sequencing

Libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina, Milan, Italy). Genomic DNA was tagged using Bead-Linked Transposomes in Tagmentation Buffer 1 at $55\text{ }^{\circ}\text{C}$ for 15 min. Tagmentation was stopped with Tagment Stop Buffer at $37\text{ }^{\circ}\text{C}$ for 15 min, followed by two washes with Tagment Wash Buffer on a magnetic stand. Tagmented DNA was amplified by limited-cycle PCR using Enhanced PCR Mix and dual-index adapters (i7 and i5); the number of cycles was adjusted to DNA input (5–12 cycles). Thermocycling comprised 3 min at $68\text{ }^{\circ}\text{C}$ and 3 min at $98\text{ }^{\circ}\text{C}$, followed by cycles of 45 s at $98\text{ }^{\circ}\text{C}$, 30 s at $62\text{ }^{\circ}\text{C}$, and 2 min at $68\text{ }^{\circ}\text{C}$, with a final extension of 1 min at $68\text{ }^{\circ}\text{C}$. Libraries were purified by double-sided size selection with Sample Purification Beads, washed twice with 80% ethanol, and eluted in Resuspension Buffer. Sequencing was performed on a NextSeq 6000Dx platform (Illumina Inc., San Diego, CA, USA). Sequence data were deposited in the National Center for Biotechnology Information under BioProject PRJNA1463802.

2.5. Raw-Read Processing and Genome Assembly

Initial quality control of paired-end FASTQ files was performed with Falco [13]. Low-quality bases and adapter sequences were removed with fastp [14], and post-filtering reports were inspected before assembly. Quality-filtered reads were assembled de novo

with Shovill [15], a SPAdes-based pipeline optimized for bacterial genomes. Assembly graphs were inspected with Bandage [16], and assembly quality metrics, including contig number and N50, were calculated with QUAST [17].

2.6. Taxonomic, Phylogenetic, and Functional Annotation

Pairwise average nucleotide identity (ANI) was calculated with fastANI [18]. Isolates from the same producer that shared $\geq 99.9\%$ ANI were assigned to the same strain group, and one representative assembly was retained for non-redundant analyses. Taxonomic assignment and phylogenetic placement were performed with GTDB-Tk v2 using the bacterial set of 120 conserved marker genes [19]. KEGG-oriented metabolic annotation was obtained with kbDRAM [20] in the KBase environment [21]. Phylogenetic trees were visualized with the *phytools* R package [22]. Structural annotation and targeted mining of technological traits were performed with Prokka v1.14.6 [23], focusing on carbohydrate and energy metabolism, transport systems, proteases and peptidases, lipolytic enzymes, and genes associated with acid, salt, heat, and oxidative stress tolerance.

2.7. Safety Assessment

The genomic safety profile was assessed using three complementary approaches. Antimicrobial-resistance-associated genes were screened with AMRFinderPlus v3.11 using the *-plus* option [24]. Virulence-associated genes were screened with Abricate [25] against the Virulence Factor Database, using minimum identity and coverage thresholds of 80%. A targeted search was also conducted for histidine, tyrosine, ornithine, and lysine decarboxylase genes associated with the potential formation of biogenic amines.

2.8. Phenotypic Measurements and Acidification Descriptors

Acidification and other technological phenotypes were measured previously [4]. Briefly, pH was recorded at 0, 3, 4, and 6 h. Acidification descriptors were defined as $\Delta\text{pH}_{03} = \text{pH}_0 - \text{pH}_3$, $\Delta\text{pH}_{36} = \text{pH}_3 - \text{pH}_6$, and $\Delta\text{pH}_{06} = \text{pH}_0 - \text{pH}_6$; therefore, larger positive values represent greater pH decline. A decay model was fitted to each pH profile to obtain the acidification rate constant (k) and acidification power (P). Although k and P were retained as complementary descriptors of acidification dynamics, ΔpH_{06} was selected a priori as the primary phenotype for genotype–phenotype modelling. Unlike model-derived kinetic parameters, ΔpH_{06} is a directly observed and readily interpretable measure of the magnitude of acidification over the common 0–6 h experimental window, allowing robust comparison among strains without introducing additional uncertainty from nonlinear parameter estimation. Previously measured lactic acid production, proteolytic activity, and spot-on-the-lawn antimicrobial activity were also integrated into the analysis.

2.9. Statistical Analysis

Built-in scripts were tailored for the automated assembly and functional annotation of the LAB genomes. Analyses in Python (v3.12.14) used pandas [26], NumPy [27], SciPy [28], matplotlib [29], and seaborn [30] for data management and the generation of correlation matrices within a Snakemake v9.22.0 workflow [31].

All downstream statistical analyses were performed in R v4.5.1 [32]. Data were imported using the readxl package [33] and processed using dplyr [34], tidyr [35], and stringr [36].

Statistical analyses were conducted at the non-redundant strain level ($n = 22$), after excluding genomes flagged because of assembly-quality concerns. Phenotypic measurements from isolates belonging to the same non-redundant strain were aggregated by calculating their arithmetic mean, whereas genomic counts were obtained from the representative genome of each cluster.

Associations between acidification variables and genomic predictors were initially assessed using Pearson correlation coefficients. The Benjamini–Hochberg procedure was applied to control the false discovery rate (FDR) across multiple comparisons. Differences among species were evaluated using Kruskal–Wallis tests, followed by FDR correction of the resulting p -values. These analyses were implemented using functions from the R stats package [32].

Linear mixed-effects models were fitted using the lme4 package [37]. The statistical significance of fixed effects was evaluated using Satterthwaite’s approximation for denominator degrees of freedom, as implemented in lmerTest [38]. Species was included as a random-intercept grouping factor, and the standardised total genomic gene count was included as a fixed-effect predictor. Separate models were fitted for total acidification (ΔpH_{06}), lactic acid production, and proteolytic index:

$$Y_{ij} = \beta_0 + \beta_1 Z_{\text{total},ij} + u_j + \varepsilon_{ij}, \quad (1)$$

where Y_{ij} represents the phenotypic outcome for strain i belonging to species j , $Z_{\text{total},ij}$ is the standardized total genomic gene count, $u_j \sim N(0, \sigma_{\text{species}}^2)$ is the species-level random intercept, and $\varepsilon_{ij} \sim N(0, \sigma_{\text{residual}}^2)$ is the residual error.

Model assumptions and goodness-of-fit statistics were assessed using the performance package [39]. Intraclass correlation coefficients and marginal and conditional coefficients of determination were obtained from the fitted mixed models. Predicted marginal effects were calculated usingggeffects [40].

Complementary exploratory analyses were conducted to identify predictors of acidification. Random Forest regression was used to evaluate possible non-linear relationships and predictor interactions, with model performance summarised by the out-of-bag coefficient of determination, root mean squared error, and permutation-based variable importance [41]. Candidate regression models were compared using the small-sample corrected Akaike Information Criterion (AICc), ΔAICc values, and Akaike weights, calculated with the MuMIn package [42]. A sensitivity analysis was additionally performed after excluding species represented by a single non-redundant strain, resulting in a reduced dataset of $n = 20$ strains.

Statistical graphics were generated using ggplot2 [43], and multi-panel figures were assembled using patchwork [44]. Unless otherwise stated, statistical significance was evaluated at $\alpha = 0.05$.

3. Results

3.1. Ecological and Taxonomic Diversity

Taxonomic classification assigned all 59 isolates to the order *Lactobacillales* (phylum *Bacillota*, class *Bacilli*), resolving eight species across six genera (Figure 1; Table 1). The taxonomic composition varied markedly among the sampled municipalities, with each location presenting a distinct species profile.

Enterococcus lactis was the numerically dominant species overall ($n = 20$; 33.9% of all isolates) and was almost exclusively concentrated in Mirandela, from which 18 of the 20 *E. lactis* isolates were recovered, with single representatives in Valpaços and Bragança. Despite this numerical dominance, pairwise ANI analysis identified only two non-redundant *E. lactis* genomes across all 20 isolates (Tables 1 and A1), indicating a highly clonal population structure.

Table 1. Species identification and genome characteristics of LAB isolates (DC4755–DC4813) from artisanal alheira sausages. *n*: total isolates; *n_{nr}*: non-redundant strains retained after duplicate removal (pairwise ANI $\geq$ 99.9%). Accession superscripts refer to footnotes.

Species	<i>n</i>	<i>n_{nr}</i>	Closest Reference
<i>Enterococcus lactis</i>	20	2	<i>E. lactis</i> CCM 8412 ^a
<i>Leuconostoc mesenteroides</i>	13	8	<i>L. mesenteroides</i> ATCC 8293 ^b
<i>Lactiplantibacillus plantarum</i>	8	3	<i>L. plantarum</i> WCFS1 ^c
<i>Latilactobacillus sakei</i>	8	3	<i>L. sakei</i> 23K ^d
<i>Pediococcus pentosaceus</i>	4	3	<i>P. pentosaceus</i> DSM 20336 ^e
<i>Lacticaseibacillus paracasei</i>	4	3	<i>Lb. paracasei</i> BD-II ^f
<i>Weissella viridescens</i>	1	1	<i>W. viridescens</i> DSM 20410 ^g
<i>Enterococcus durans</i>	1	1	<i>E. durans</i> NBRC 100479 ^h

^a GCF_015751045.1; ^b GCF_000014445.1; ^c GCF_014131735.1; ^d GCF_002370355.1; ^e GCF_001437285.1; ^f GCF_000829035.1; ^g GCF_001437355.1; ^h GCF_001544215.1. Duplicates identified by pairwise ANI (fastANI $\geq$ 99.9%) among isolates from the same collection. Full strain inventory in Table A1.

Phylogenetic Relationships

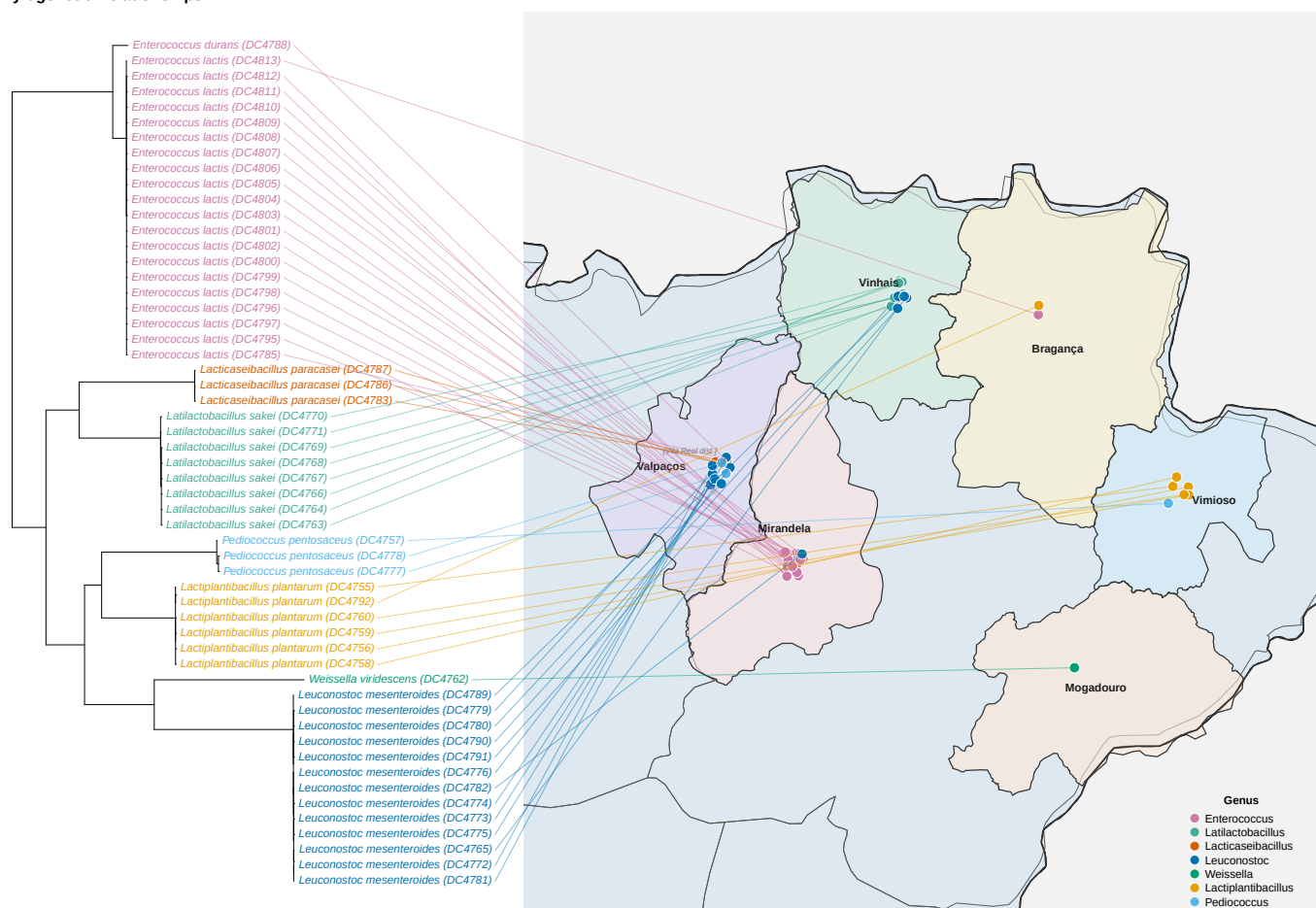


Figure 1. Phylogenetic relationships and geographic origins of the 59 LAB isolates recovered from artisanal alheira sausages. The phylogenetic tree is linked to the municipality of origin, and colors denote genus-level assignments.

Leuconostoc mesenteroides was the second most abundant species (*n* = 13; 22.0%), recovered predominantly from Valpaços (*n* = 8) and Vinhais (*n* = 4), with a single isolate from Mirandela. The remaining species each displayed a clear geographic signature. *Latilactobacillus sakei* (*n* = 8) was recovered exclusively from Vinhais, where it represented

two-thirds of all isolates from that municipality, and formed a phylogenetic clade with three non-redundant genomes. *Lactiplantibacillus plantarum* ($n = 8$) was predominant in Vimioso ($n = 5$) and Bragança ($n = 3$), likewise forming a phylogenetic cluster with three non-redundant genomes.

Lacticaseibacillus paracasei ($n = 4$) was isolated exclusively from Valpaços, with three non-redundant genomes. *Pediococcus pentosaceus* ($n = 4$) was distributed across Valpaços, Mirandela, and Vimioso. Valpaços was the most species-rich collection site, with five distinct species recovered across 16 isolates, while Vinhais yielded a two-species community (*Lt. sakei* and *Ln. mesenteroides*) and Mirandela was dominated almost entirely by *E. lactis*. *Weissella viridescens* ($n = 1$) and *E. durans* ($n = 1$) were represented by single isolates from Mogadouro and Valpaços, respectively; although both were placed within the reference phylogeny, their low abundance precluded within-species diversity assessment.

3.2. Functional Genomic Characterization

To evaluate the metabolic potential and technological relevance of the LAB collection, the functional analysis considered carbon and energy metabolism, carbohydrate-active enzymes, transport systems, stress-tolerance mechanisms, proteolytic and lipolytic functions, and safety-relevant genomic features.

3.2.1. Carbon Utilization

In the alheira matrix, bread serves as the primary carbohydrate source, providing starch that is broken down into fermentable sugars (i.e., glucose and maltose). Several key metabolic housekeeping genes encoding glycolytic enzymes were identified (Table 2) across our strain collection, including glucose-6-phosphate isomerase [EC: 5.3.1.9], 6-phosphofructokinase [EC: 2.7.1.11], fructose-bisphosphate aldolase [EC: 4.1.2.13], triosephosphate isomerase [EC: 5.3.1.1], phosphoglycerate kinase [EC: 2.7.2.3], phosphoglycerate mutase [EC: 5.4.2.11], enolase [EC: 4.2.1.11], pyruvate kinase [EC: 2.7.1.40], L-lactate dehydrogenase and D-lactate dehydrogenase, characteristic of the Embden–Meyerhof–Parnas (EMP) pathway, see Figure A1.

Moreover, the pentose phosphate pathway (PPP) was broadly conserved across the strain collection. Genes encoding the oxidative phase enzymes glucose-6-phosphate 1-dehydrogenase [EC: 1.1.1.49, 1.1.1.193], 6-phosphogluconolactonase [EC: 3.1.1.31], and 6-phosphogluconate dehydrogenase [EC: 1.1.1.44, 1.1.1.343] were identified in all strains. However, transketolase [EC: 2.2.1.1] was absent in 16 strains (43/59), specifically in *Lt. sakei*, *P. pentosaceus*, *W. viridescens*, and most *Lb. paracasei* strains. Transaldolase [EC: 2.2.1.2] showed the most restricted distribution, being present in only 8/59 strains, all belonging to *L. plantarum*.

Genes associated with the Entner–Doudoroff (ED) pathway were partially conserved across the collection. Glucose-6-phosphate 1-dehydrogenase [EC: 1.1.1.49, 1.1.1.193] and 6-phosphogluconolactonase [EC: 3.1.1.31] were present in all strains, consistent with their shared role in the oxidative phase of the pentose phosphate pathway. However, 2-dehydro-3-deoxyphosphogluconate aldolase [EC: 4.1.2.14], which is specific to the ED pathway and catalyses the cleavage of 2-dehydro-3-deoxy-6-phosphogluconate into pyruvate and glyceraldehyde-3-phosphate, was identified in only 30/59 strains. Its presence was restricted to *E. lactis*, *E. durans*, *Lb. paracasei* strains, strains of *P. pentosaceus* (S18, S19), and *L. plantarum* (S13).

Table 2. Presence (+) and absence (–) of genes associated with carbon-metabolism functions and carbohydrate-active-enzyme families across the 24 non-redundant LAB strain groups. A plus sign denotes detection of one or more annotated components and does not necessarily indicate a complete functional pathway.

Pathway	<i>E. Lac.</i>				<i>Ln. Mes.</i>					<i>L. Pla.</i>				<i>Lt. Sak.</i>			<i>P. Pen.</i>			<i>Lb. Par.</i>			<i>W. Vir.</i>	<i>E. Dur.</i>
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24
<i>Central carbon metabolism</i>																								
EMP-associated genes	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PPP-associated genes	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ED-associated genes	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Galactose (Leloir pathway)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Galacturonate degradation	+	+	+	+	+	+	+	+	+	+	–	–	+	+	+	+	–	+	+	+	+	+	–	+
<i>Pyruvate metabolism</i>																								
Acetate pathway	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Pyruvate → acetyl-CoA	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	–	+
<i>Sugar utilisation</i>																								
Mannose degradation	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Galactose degradation	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Lactose degradation	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	–	+	+	–	+
Sucrose degradation	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	–	+	–	+	+	+	–	–
Xylose (isomerase pathway)	–	–	+	+	+	+	+	+	+	+	–	–	–	–	–	–	–	+	+	+	–	+	–	+
Xylose (oxidoreductase)	–	–	+	+	+	+	+	+	+	+	–	–	–	–	–	–	–	+	+	+	–	+	–	–
<i>CAZymes</i>																								
Glycoside hydrolases	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Glycosyltransferases	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Polysaccharide lyases	+	+	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	+	–	+	–	+

EMP, Embden–Meyerhof–Parnas; PPP, pentose phosphate pathway; ED, Entner–Doudoroff pathway. Presence indicates at least one annotated component; pathway completeness must be evaluated from the individual enzyme profile. Strain groups are arranged by species.

The lac operon was also investigated. At the isolate level, the lactose permease (*lacY*)-BetaGal pathway was strongly species-structured (Figure 2). However, after clonally redundant isolates were collapsed to the non-redundant strain level used for statistical modelling, *lacY* was present only in the two *E. lactis* strain groups. Consequently, *lacY* status was completely confounded with species and could not be interpreted as an independent predictor of acidification.

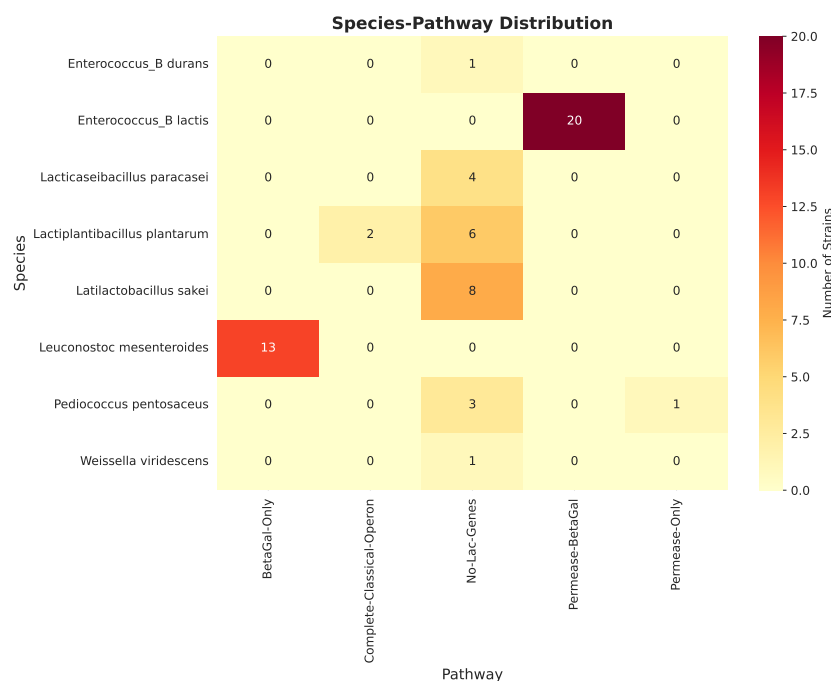


Figure 2. Distribution of lac operon pathway types across LAB species isolated from alheira ($n = 59$). Heatmap showing the number of strains per species assigned to each lac operon pathway category, based on presence/absence of *lacZ/lacX* (β -galactosidase), *lacY* (permease), and *lacA* (transacetylase) genes. Cell values indicate strain counts; color intensity is proportional to the number of strains (scale bar, right).

3.2.2. Carbohydrate-Active Enzymes

All 59 isolates encoded diverse repertoires of carbohydrate-active enzymes (CAZymes) (Table 2 and Figure A2). Among the Glycoside Hydrolases (GHs), GH13, which encodes alpha-amylase [EC: 3.2.1.1] and pullulanase [EC: 3.2.1.41], was the most widely distributed family, present in all strains. GH43, encoding beta-xylosidase [EC: 3.2.1.37] and alpha-L-arabinofuranosidase [EC: 3.2.1.55], was present in 37/59 strains, being absent in *Lt. sakei*, *Lb. paracasei*, *W. viridescens*, *P. pentosaceus* strains S17 and S19, and *L. plantarum* strains S11 and S12. GH5, encoding endo-beta-1,4-glucanase/cellulase [EC: 3.2.1.4], was the most restricted GH family (13/59), with its presence confined exclusively to *Ln. mesenteroides* strains.

Among the Glycosyl Transferases (GTs), GT2, encoding cellulose synthase [EC: 2.4.1.12] and chitin synthase [EC: 2.4.1.16], was nearly ubiquitous across the collection (58/59), suggesting a broadly conserved capacity for exopolysaccharide biosynthesis. Regarding the Polysaccharide Lyases (PLs), PL9, encoding pectate lyase [EC: 4.2.2.2] and exopolygalacturonase [EC: 3.2.1.67], was present in 23/59 strains, restricted to *E. lactis*, *E. durans*, strain S21 of *Lb. paracasei*, and strain S13 of *L. plantarum*. PL12, encoding heparan-sulfate lyase [EC: 4.2.2.8], and PL8, encoding hyaluronate lyase [EC: 4.2.2.1] and chondroitin AC lyase [EC: 4.2.2.5], showed the most restricted distributions, present in only 3/59 and 1/59 strains

respectively; PL12 was identified in *Lb. paracasei* and *E. durans*, while PL8 was confined to *E. durans*.

3.2.3. Stress Tolerance

Acid tolerance was supported through multiple complementary mechanisms (Tables 3 and A5). The F_0F_1 -ATPase was universally present. The arginine deiminase system (*arcA*) was present in most species, absent only in *L. plantarum* S11 and S12 and *Lb. paracasei* S20 and S22, while glutamate decarboxylase (*gadB*), an alternative acid resistance mechanism, was restricted exclusively to *L. plantarum* (S11–S13).

Oxygen-related metabolic capacity, encompassing NADH oxidases and alkyl hydroperoxide reductases, was restricted to *L. plantarum* (8/8), *Ln. mesenteroides* (13/13), *Lb. paracasei* (4/4), and *W. viridescens* (1/1). Thioredoxin reductase (*trxB*) and the glutathione transport system (*cydD/kefB*) were universal, forming the basal oxidative defence shared by all strains. The peroxide-scavenging route, comprising alkyl hydroperoxide reductase (*ahpC*), was present in *E. lactis*, *Ln. mesenteroides*, and *Lb. paracasei*, but absent in *Lt. sakei*, *P. pentosaceus*, and *W. viridescens*.

The superoxide dismutation route, through *sodA*, was present in *E. lactis*, *Lt. sakei*, and *Lb. paracasei*, but absent across all eight *Ln. mesenteroides* strains. Vegetative catalase (*katA*) showed the most restricted distribution, detected exclusively in *L. plantarum* (all three strains), *Lt. sakei* (all three strains), and *W. viridescens*, and fully absent in *E. lactis*, *Ln. mesenteroides*, and *Lb. paracasei*. This complementary pattern between *ahpC* and *katA*, where the former tends to be present when the latter is absent, suggests functionally divergent oxidative stress strategies across the collection.

Choloylglycine hydrolase (*cbh*) was uniquely present in both *E. lactis* strains (S1 and S2), while bile salt hydrolase (*bsh*) was detected in *E. lactis* S1, all three *L. plantarum* strains, *Lb. paracasei* S21, *W. viridescens* S23, and *E. durans* S24. All *Ln. mesenteroides*, *Lt. sakei*, and *P. pentosaceus* strains lacked both bile tolerance genes entirely.

Moreover, heat shock response genes were broadly conserved. The DnaK/DnaJ/GrpE chaperone triad was detected in all 24 non-redundant strain groups, and the GroEL/GroES chaperonin system was similarly ubiquitous, with the sole exception of *E. lactis* S2, where both genes were absent. The disaggregase ClpB was present in most species but absent across all eight *Ln. mesenteroides* strains (S3–S10) and in *W. viridescens* (S23). Salt stress adaptation was uniformly supported by the glycine betaine transporters *betL/opuC* and *opuCD*, which were conserved across all strains.

3.2.4. Transporter Profile

The transporter profiles were strongly species-structured, reflecting distinct ecological adaptations across the 24 non-redundant strain groups. A partial hemin transport system (*hrtA*, K09814) was detected in 5/24 strains: both *E. lactis* strains (S1–S2), one *L. plantarum* strain (S13), one *Lb. paracasei* strain (S21), and *E. durans* (S24). Cobalt/nickel transport (*cbiM*, K02007) was detected in 7/24 strains, shared between all three *L. plantarum* strains (S11–S13) and all three *Lt. sakei* strains (S14–S16), with one additional *P. pentosaceus* strain (S18). Compatible solute transport (glycine betaine/proline ABC transporter, *proV*/K02000) was near-universal, present in 20/24 strains across all eight species, with the exception of two *L. plantarum* strains (S11 and S12) and two *P. pentosaceus* strains (S17 and S19).

Table 3. Presence (+) and absence (−) of energy metabolism pathways and stress tolerance genes across the 24 non-redundant LAB strain groups.

Gene/Pathway	Function	E. Lac.				Ln. Mes.					L. Pla.				Lt. Sak.				P. Pen.			Lb. Par.			W. Vir.	E. Dur.
		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18 †	S19	S20	S21	S22 ‡	S23	S24	
Respiratory chain complexes																										
Fumarate reductase	Anaerobic respiratory complex	+	+	+	+	−	−	+	+	+	+	+	+	+	−	−	−	+	+	+	+	+	+	−	+	
F-type ATPase	F-type proton-translocating ATPase	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
V/A-type ATPase	V/A-type proton-translocating ATPase	+	+	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	+	
Cytochrome bd	Terminal quinol oxidase (aerotolerance)	−	−	+	+	+	+	+	+	+	+	+	+	+	−	−	−	−	−	−	+	+	+	+	−	
Metal/inorganic reduction																										
arsC	Arsenate reduction	+	+	−	−	−	−	−	−	−	−	−	−	+	−	−	−	+	+	+	−	−	−	−	+	
Heat shock																										
clpP	Clp protease proteolytic subunit	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
clpB	Chaperone ClpB (disaggregase)	+	+	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+	+	+	+	+	−	+	
dnaK	Chaperone DnaK (Hsp70)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
dnaJ	Co-chaperone DnaJ (Hsp40)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
grpE	Nucleotide exchange factor GrpE	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
groEL	Chaperonin GroEL (Hsp60)	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
groES	Co-chaperonin GroES (Hsp10)	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Oxidative stress																										
trxB	Thioredoxin/thioredoxin reductase	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
cydD/kefB	Glutathione transport system	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
ahpC	Alkyl hydroperoxide reductase C	+	+	+	+	+	+	+	+	+	+	−	−	+	−	−	−	−	−	−	+	+	+	−	+	
sodA	Superoxide dismutase [Mn]	+	+	−	−	−	−	−	−	−	−	−	−	+	+	+	+	−	+	−	+	+	+	−	+	
katA	Vegetative catalase	−	−	−	−	−	−	−	−	−	−	+	+	+	+	+	+	−	+	−	−	−	−	+	−	
Acid tolerance																										
F ₀ F ₁ -ATPase	Proton-translocating ATPase	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
arcA	Arginine deiminase system	+	+	+	+	+	+	+	+	+	+	−	−	+	+	+	+	+	+	+	−	+	−	+	+	
gadB	Glutamate decarboxylase	−	−	−	−	−	−	−	−	−	−	+	+	+	−	−	−	−	−	−	−	−	−	−	−	
Bile tolerance																										
bsh	Bile salt hydrolase	+	−	−	−	−	−	−	−	−	−	+	+	+	−	−	−	−	−	−	−	+	−	−	+	
cbh	Choloylglycine hydrolase	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	
Salt tolerance																										
betL/opuC	Glycine betaine transporter	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
opuCD	Betaine/carnitine/choline transport	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
proV	Proline/betaine transport ATPase	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
opuA	Glycine betaine ABC transporter OpuA	+	−	+	+	+	+	+	+	+	+	−	−	+	−	−	−	−	+	−	−	+	−	−	+	

Strains are grouped by species. F-type and V/A-type ATPases are proton-translocating complexes. Cytochrome *bd*: terminal quinol oxidase associated with aerotolerance. Arsenate reduction encoded by *arsC*. Stress categories: heat shock, oxidative stress, acid tolerance, bile tolerance, and salt tolerance. † Representative assembly had insufficient amino acids in the GTDB-Tk multiple-sequence alignment. ‡ Representative assembly had more than 37.5% of marker genes with multiple hits.

3.2.5. Proteases

Proteolytic genes were broadly distributed across the 24 non-redundant strain groups, but their detailed presence/absence patterns were not strongly associated with the phenotypes examined here. The complete protease and peptidase inventory is provided in Figure 3 and Tables A2–A4. Core Clp-system components were widely conserved, while selected stress-linked and cell-envelope proteases showed species-dependent distributions. These annotations were treated as indicators of technological potential rather than as direct predictors of acidification.



Figure 3. Heatmap of annotated protease and peptidase products across the 24 non-redundant strain groups. Cell intensity represents the annotation count for each product category.

3.2.6. Lipases

Lipolytic annotations were also species-structured, spanning carboxylesterases, putative and true lipases, phospholipases, and bile salt hydrolases. The complete strain-by-product distribution is shown in Figure 4 and the Appendix A. *E. lactis* displayed the broadest annotated lipolytic repertoire, whereas several other species showed more restricted profiles. As with the proteolytic annotations, these genes were interpreted as potential technological functions requiring phenotypic confirmation.

3.2.7. Safety Profile

Two antibiotic resistance-associated transporter systems were identified (Table 4). The bacitracin resistance ABC transporter BceAB (*bceAB*) was present in 7/24 strains: both *E. lactis* strains (S1–S2), all three *Lt. sakei* strains (S14–S16), one *L. plantarum* strain (S13), and one *P. pentosaceus* strain (S18). The tetracycline efflux pump TetA (*tetA*) was detected in 11/24 strains, comprising both *E. lactis* strains (S1–S2), all three *Lt. sakei* strains (S14–S16), all three *Lb. paracasei* strains (S20–S22), one *L. plantarum* strain (S13), one *P. pentosaceus* strain (S18), and *E. durans* (S24). The *VraDE* transporter (*vraDE*), which mediates a complementary bacitracin resistance mechanism, showed the broadest distribution of the three AMR systems detected, present in 19/24 strains and absent only in *L. plantarum* S11–S12, *P. pentosaceus* S17 and S19, and *W. viridescens* S23.

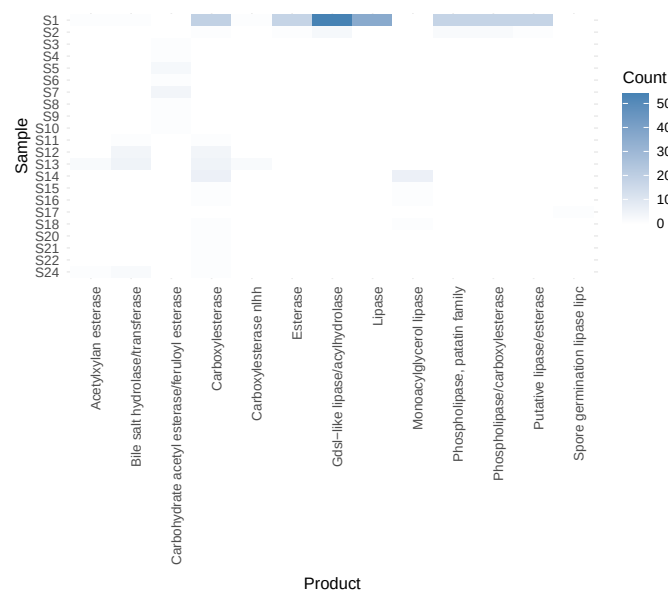


Figure 4. Heatmap of annotated lipolytic products across the 24 non-redundant strain groups. Cell intensity represents the annotation count for each product category.

Virulence-associated genes were restricted to three strains: the collagen adhesin *acm* was detected in both *E. lactis* strains (S1–S2) and one *L. plantarum* strain (S13), while the collagen/fibronectin-binding protein *sgrA* was additionally detected in S2 and S13. The remaining 21 strains had no virulence genes detected.

Biogenic amine biosynthesis genes were detected in five strains: histamine decarboxylase (*hdcA*) in two *Lb. paracasei* strains (S21–S22) and *E. durans* (S24); tyramine decarboxylase (*tdcA*) and lysine decarboxylase (*ldcA*) in both *E. lactis* strains (S1–S2); and ornithine decarboxylase (*odc*) in all three *Lb. paracasei* strains (S20–S22). The co-occurrence of *bceAB*, *tetA*, and *vraDE* in both *E. lactis* strains and all *Lt. sakei* strains, combined with the QPS non-listed status of *E. lactis*, *W. viridescens*, and *E. durans*, indicates that phenotypic minimum inhibitory concentration (MIC) testing in accordance with EFSA guidelines is required before any of these strains are considered for starter culture applications.

Table 4. Safety-relevant genomic features of the 24 non-redundant LAB strains (DC4755–DC4813) from artisanal alheira sausages.

Strain	Species	AMR-Associated Genes	Virulence Genes	QPS Status	Biogenic Amines
<i>Enterococcus lactis</i>					
S1	<i>E. lactis</i>	<i>bceAB, tetA, vraDE</i>	<i>acm</i>	Not listed	Tyramine (<i>tdcA</i>); cadaverine (<i>ldcA</i>)
S2	<i>E. lactis</i>	<i>bceAB, tetA, vraDE</i>	<i>acm, sgrA</i>	Not listed	Tyramine (<i>tdcA</i>); cadaverine (<i>ldcA</i>)
<i>Leuconostoc mesenteroides</i>					
S3	<i>Ln. mesenteroides</i>	<i>vraDE</i>	nd	QPS listed	nd
S4	<i>Ln. mesenteroides</i>	<i>vraDE</i>	nd	QPS listed	nd
S5	<i>Ln. mesenteroides</i>	<i>vraDE</i>	nd	QPS listed	nd
S6	<i>Ln. mesenteroides</i>	<i>vraDE</i>	nd	QPS listed	nd
S7	<i>Ln. mesenteroides</i>	<i>vraDE</i>	nd	QPS listed	nd
S8	<i>Ln. mesenteroides</i>	<i>vraDE</i>	nd	QPS listed	nd
S9	<i>Ln. mesenteroides</i>	<i>vraDE</i>	nd	QPS listed	nd
S10	<i>Ln. mesenteroides</i>	<i>vraDE</i>	nd	QPS listed	nd

Table 4. Cont.

Strain	Species	AMR-Associated Genes	Virulence Genes	QPS Status	Biogenic Amines
<i>Lactiplantibacillus plantarum</i>					
S11	<i>L. plantarum</i>	nd	nd	QPS listed	nd
S12	<i>L. plantarum</i>	nd	nd	QPS listed	nd
S13	<i>L. plantarum</i>	<i>bceAB, tetA, vraDE</i>	<i>acm, sgrA</i>	QPS listed	nd
<i>Latilactobacillus sakei</i>					
S14	<i>Lt. sakei</i>	<i>bceAB, tetA, vraDE</i>	nd	QPS listed	nd
S15	<i>Lt. sakei</i>	<i>bceAB, tetA, vraDE</i>	nd	QPS listed	nd
S16	<i>Lt. sakei</i>	<i>bceAB, tetA, vraDE</i>	nd	QPS listed	nd
<i>Pediococcus pentosaceus</i>					
S17	<i>P. pentosaceus</i>	nd	nd	QPS listed	nd
S18 [†]	<i>P. pentosaceus</i>	<i>bceAB, tetA, vraDE</i>	nd	QPS listed	nd
S19	<i>P. pentosaceus</i>	nd	nd	QPS listed	nd
<i>Lactocaseibacillus paracasei</i>					
S20	<i>Lb. paracasei</i>	<i>tetA, vraDE</i>	nd	QPS listed	Putrescine (<i>odc</i>) Histamine (<i>hdcA</i>); putrescine (<i>odc</i>) Histamine (<i>hdcA</i>); putrescine (<i>odc</i>)
S21	<i>Lb. paracasei</i>	<i>tetA, vraDE</i>	nd	QPS listed	
S22 [‡]	<i>Lb. paracasei</i>	<i>tetA, vraDE</i>	nd	QPS listed	
<i>Weissella viridescens</i>					
S23	<i>W. viridescens</i>	nd	nd	Not listed	nd
<i>Enterococcus durans</i>					
S24	<i>E. durans</i>	<i>tetA, vraDE</i>	nd	Not listed	Histamine (<i>hdcA</i>)

Abbreviations. AMR: antimicrobial resistance; QPS: Qualified Presumption of Safety; LAB: lactic acid bacteria; nd: not detected. **AMR-associated genes.** *bceAB*: bacitracin resistance ABC transporter; *vraDE*: bacitracin resistance VraDE transporter; *tetA*: tetracycline efflux pump. **Virulence genes.** *acm*: collagen adhesin; *sgrA*: collagen/fibronectin-binding MSCRAMM protein. **Biogenic amine decarboxylase genes.** *hdcA*: histidine decarboxylase; *tdcA*: tyrosine decarboxylase; *ldcA*: putative lysine decarboxylase; *odc*: inducible ornithine decarboxylase. [†] Representative assembly had insufficient amino acids in the GTDB-Tk multiple-sequence alignment. [‡] Representative assembly had more than 37.5% of marker genes with multiple hits.

3.3. Integration of Genomic and Phenotypic Profiles

Genomic predictors were integrated with acidification and other technological phenotypes. Descriptive species-level summaries in Table 5 use isolate-level phenotypic observations after exclusion of the five assemblies flagged in Table A1; genomic counts were assigned from the corresponding ANI-group representative. Inferential analyses were conducted at the non-redundant strain-group level ($n=22$), as described in Section 2.9.

Descriptive Species-Level Patterns

E. lactis isolates showed the greatest mean total pH decline ($\Delta\text{pH}_{06} = 0.84 \pm 0.05$) and the lowest mean count of genes within the five targeted metabolic pathways (8.25 ± 1.37). *Lb. paracasei* also showed a comparatively large mean pH decline (0.76 ± 0.04) and the highest targeted-pathway gene count (27.5 ± 0.7), although only two unflagged isolates contributed to this descriptive summary. *L. plantarum* displayed moderate acidification and high targeted-pathway gene counts, while *Lt. sakei*, *P. pentosaceus*, and *Ln. mesenteroides* showed smaller mean pH declines.

Table 5. Descriptive species-level summary of acidification and targeted metabolic-gene counts. Values are means $\pm$ standard deviations. $\Delta\text{pH}_{06} = \text{pH}_0 - \text{pH}_6$; larger positive values indicate greater acidification.

Species	<i>n</i>	pH_0	pH_6	ΔpH_{06}	Mean Genes
<i>Enterococcus lactis</i>	20	6.63 $\pm$ 0.02	5.79 $\pm$ 0.05	0.84 $\pm$ 0.05	8.25 $\pm$ 1.37
<i>Leuconostoc mesenteroides</i>	13	6.66 $\pm$ 0.06	6.36 $\pm$ 0.12	0.30 $\pm$ 0.14	21.0 $\pm$ 3.6
<i>Lactiplantibacillus plantarum</i>	6	6.67 $\pm$ 0.04	6.33 $\pm$ 0.24	0.34 $\pm$ 0.22	24.8 $\pm$ 1.6
<i>Latilactobacillus sakei</i>	8	6.70 $\pm$ 0.03	6.44 $\pm$ 0.06	0.26 $\pm$ 0.07	18.0 $\pm$ 3.7
<i>Pediococcus pentosaceus</i>	3	6.66 $\pm$ 0.06	6.42 $\pm$ 0.05	0.24 $\pm$ 0.06	13.0 $\pm$ 0.0
<i>Lacticaseibacillus paracasei</i>	2	6.57 $\pm$ 0.03	6.13 $\pm$ 0.43	0.76 $\pm$ 0.04	27.5 $\pm$ 0.7
<i>Weissella viridescens</i>	1	6.69	6.45	0.24	18.0
<i>Enterococcus durans</i>	1	6.59	5.79	0.79	24.0

Five isolate assemblies (DC4761, DC4784, DC4787, DC4793, and DC4794) were flagged. The unflagged assembly DC4786 represented ANI group S21 in strain-level inference.

Antimicrobial activity varied by pathogen, temperature, and species (Table 6). At 10 °C, the largest mean inhibition zones against *L. monocytogenes* were observed for *L. plantarum* (19.54 mm) and *Ln. mesenteroides* (19.23 mm). Against *S. enterica* serovar Typhimurium, inhibition zones were generally smaller; the largest means were observed for *Ln. mesenteroides* at 10 °C (10.90 mm) and *E. durans* at 37 °C (13.17 mm). *E. lactis* showed little inhibition of *S. aureus*, whereas *Lb. paracasei* maintained similar mean inhibition at both temperatures (9.37 and 9.24 mm).

Table 6. Species-level summary of inhibition capacity against foodborne pathogens at 10 °C and 37 °C. Values are mean inhibition-zone diameters (mm) $\pm$ standard deviation.

Species	<i>n</i>	<i>Lm</i> 10 °C	<i>Lm</i> 37 °C	<i>Sal</i> 10 °C	<i>Sal</i> 37 °C	<i>Sta</i> 10 °C	<i>Sta</i> 37 °C
<i>Enterococcus lactis</i>	20	11.65 $\pm$ 1.87	8.33 $\pm$ 0.90	4.51 $\pm$ 1.35	0.59 $\pm$ 2.65	5.15 $\pm$ 1.35	0.60 $\pm$ 1.95
<i>Leuconostoc mesenteroides</i>	13	19.23 $\pm$ 2.34	10.54 $\pm$ 1.18	10.90 $\pm$ 0.78	12.63 $\pm$ 1.50	10.38 $\pm$ 1.19	6.42 $\pm$ 0.75
<i>Lactiplantibacillus plantarum</i>	8	19.54 $\pm$ 1.74	10.38 $\pm$ 1.11	10.32 $\pm$ 0.75	11.24 $\pm$ 1.30	10.97 $\pm$ 1.81	8.13 $\pm$ 0.96
<i>Latilactobacillus sakei</i>	8	15.29 $\pm$ 2.48	11.59 $\pm$ 1.01	8.69 $\pm$ 0.99	9.31 $\pm$ 1.21	8.43 $\pm$ 1.51	6.41 $\pm$ 0.62
<i>Pediococcus pentosaceus</i>	4	17.44 $\pm$ 4.58	10.15 $\pm$ 0.74	9.87 $\pm$ 1.58	9.76 $\pm$ 1.61	9.84 $\pm$ 1.99	6.82 $\pm$ 0.41
<i>Lacticaseibacillus paracasei</i>	4	17.68 $\pm$ 0.70	9.96 $\pm$ 0.59	9.68 $\pm$ 1.18	11.02 $\pm$ 0.84	9.37 $\pm$ 2.13	9.24 $\pm$ 0.70
<i>Weissella viridescens</i>	1	18.21	11.10	10.43	9.88	11.24	6.18
<i>Enterococcus durans</i>	1	16.73	11.87	10.68	13.17	8.07	8.63

Lm, *Listeria monocytogenes*; *Sal*, *Salmonella enterica* serovar Typhimurium; *Sta*, *Staphylococcus aureus*. Standard deviations are not applicable to species represented by one isolate.

3.4. Genotype–Phenotype Associations and Model Robustness

3.4.1. Strain-Level Correlations

At the strain-group level, carbon-metabolism, transport, energy-metabolism, total functional-gene, acid-stress, bile-stress, and total-stress counts were positively associated with ΔpH_{06} after FDR correction (Table 7). The strongest associations were observed for transport genes ($r = 0.793$) and acid-stress genes ($r = 0.763$). Transport-gene count was strongly associated with ΔpH_{36} ($r = 0.814$) and negatively associated with pH at 6 h ($r = -0.794$). Diacetyl-gene count showed a nominal negative association ($r = -0.520$, raw $p = 0.011$), but its FDR-adjusted value was 0.055.

Table 7. Pearson correlations between genomic predictors and total pH decline in the 22 non-redundant strain groups used for inference. FDR-adjusted values below 0.05 are shown in bold.

Predictor	<i>r</i>	Raw <i>p</i>	FDR <i>p</i>
Transport genes	0.793	<0.001	<0.001
Carbon-metabolism genes	0.704	<0.001	0.001
Energy-metabolism genes	0.601	<0.001	0.006
Total functional genes	0.648	<0.001	0.003
Diacetyl pathway	−0.520	0.011	0.055
Lactose pathway	−0.404	0.056	0.140
Acid-stress genes	0.763	<0.001	<0.001
Bile-stress genes	0.628	0.001	0.003
Total stress genes	0.579	0.003	0.006

3.4.2. Primary Mixed Model and Within–Between Decomposition

In the principal mixed-effects model, standardized total functional-gene count had a positive association with acidification ($b = 0.137 \pm 0.059$, $t = 2.34$, $p = 0.049$; Table 8). The species-level intraclass correlation coefficient was 0.430. The fixed effect explained 26.6% of the variance, whereas fixed and random effects together explained 58.2%. No detectable within-species association ($b = 0.057$, $p = 0.227$) was found, but rather, a positive between-species association ($b = 0.190$, $p = 0.038$). Thus, the overall association was primarily attributable to differences in average genomic profiles among species rather than to fine-scale variation among strains within the same species.

Table 8. Primary mixed-effects model and within–between decomposition for total pH decline.

Parameter	Estimate	<i>p</i>
Standardized total functional-gene count	0.137 ± 0.059	0.049
Within-species deviation	0.057	0.227
Between-species mean	0.190	0.038
Species-level SD	0.149	–
Residual SD	0.172	–
Intraclass correlation coefficient	0.430	–
Marginal R^2	0.266	–
Conditional R^2	0.582	–

A quadratic term was not significant ($p = 0.268$), and a likelihood-ratio comparison of linear and quadratic models was also non-significant ($p = 0.239$). The generalized additive model estimated one effective degree of freedom for the smooth term, supporting an approximately linear relationship over the observed range.

3.4.3. Regularization, Ensemble, and Latent-Component Analyses

At the minimum cross-validation error, LASSO restricted to the five broad functional categories retained carbon-metabolism and transport-gene counts. When all 16 predictor groups were considered, acid- and bile-stress gene counts were retained. Under the more conservative one-standard-error rule, no predictors were selected in either analysis, indicating that sparse selection was not stable in this small dataset. Elastic Net optimization selected $\alpha = 0$ (Ridge; minimum leave-one-out mean squared error 0.0378), with $\alpha = 0.25$ yielding a nearly identical error (0.0379).

Moreover, the Random Forest analysis identified carbon-metabolism, energy-metabolism, and acid-stress gene counts as the most influential predictors by permutation importance (Figure 5). Partial-dependence curves were broadly monotonic over the observed predictor ranges (Figure 6). Out-of-bag performance was modest ($R^2 = 0.245$, RMSE=0.214), and these outputs were interpreted as exploratory.

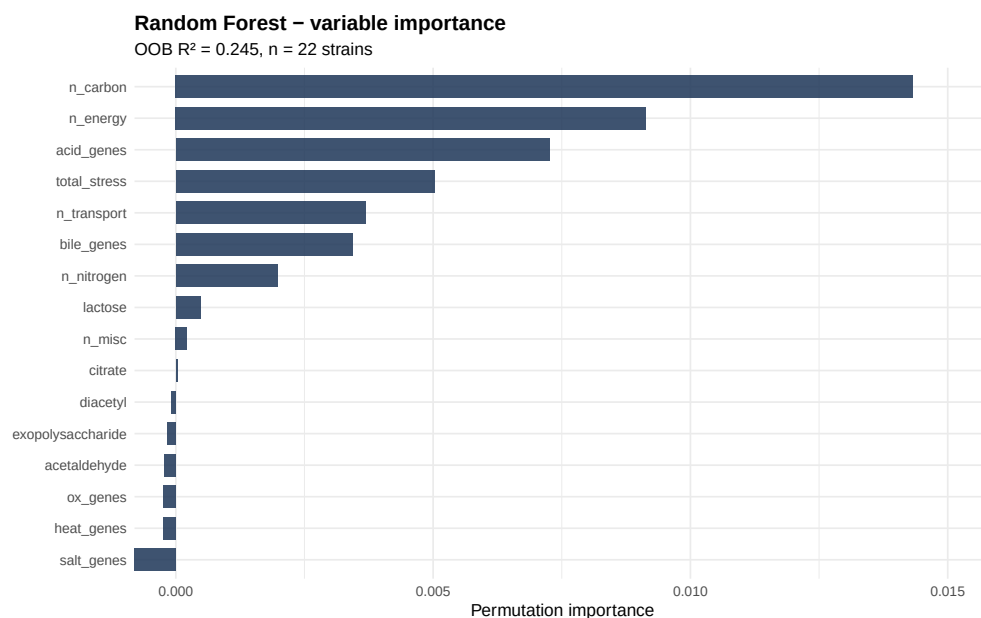


Figure 5. Permutation importance of genomic predictors in the Random Forest model for ΔpH_{06} ($n = 22$ strain groups; 2000 trees). Out-of-bag $R^2 = 0.245$.

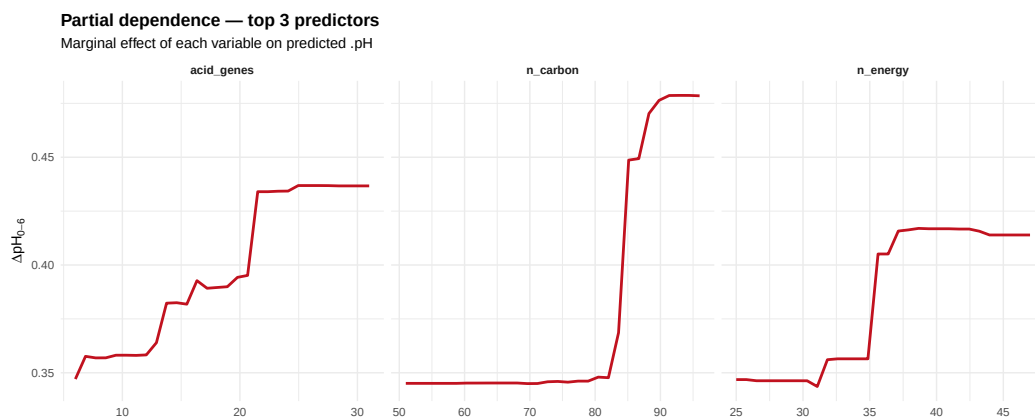


Figure 6. Partial-dependence curves for the three leading Random Forest predictors of ΔpH_{06} . Curves are exploratory because of the small strain-level sample size.

PLS regression selected one component by leave-one-out cross-validation (RMSEP = 0.1973, compared with 0.2516 for the intercept-only model). This component explained 42.3% of predictor variance and 38.5% of ΔpH_{06} variance. The reduction in RMSEP was not improved by adding a second or third component (0.2033 and 0.2094, respectively). A summary of these exploratory multivariable and validation analyses is presented in Table 9.

Table 9. Summary of exploratory multivariable and validation analyses. Cross-validation schemes differ, and performance values should not be compared as if obtained from a common external validation set.

Analysis	Selected Signal	Validation/Performance	Interpretation
LASSO, functional groups	Carbon; transport at $\lambda_{\min}$	No predictor at λ_{1SE}	Selection unstable
LASSO, 16 predictors	Acid stress; bile stress at $\lambda_{\min}$	No predictor at λ_{1SE}	Selection unstable
Elastic Net	Ridge ($\alpha = 0$)	LOOCV MSE = 0.0378	Distributed correlated signal
Random Forest	Carbon; energy; acid stress	OOB $R^2 = 0.245$; RMSE = 0.214	Modest internal prediction
PLS, one component	Multivariate latent profile	LOOCV RMSEP = 0.1973	38.5% of response variance
Linear total-diversity model	Total functional genes	LOSO-CV $R^2 = 0.078$; RMSE = 0.244	Weak prediction of unseen species

3.4.4. AICc Model Comparison

AICc comparison assigned nearly equal support to the carbon-plus-transport and stress models ($\Delta AICc = 0.01$; weights 0.422 and 0.420, respectively; Table 10). The total-diversity model retained some support, whereas the null, full functional-category, and five-pathway models were less supported.

Table 10. Small-sample-corrected Akaike information criterion comparison of candidate acidification models.

Model	k	logLik	AICc	$\Delta AICc$	Weight
Carbon + transport	5	7.54	−1.33	0.00	0.422
Stress	8	14.20	−1.32	0.01	0.420
Total diversity	4	4.58	1.19	2.52	0.120
Null	3	1.89	3.55	4.88	0.037
Functional groups	8	8.41	10.25	11.58	0.001
Targeted pathways	8	6.65	13.78	15.11	<0.001

3.4.5. Module-Level and Sugar-Degradation Screens

No individual KEGG module emerged as a robust predictor in the high-dimensional screening of 141 retained modules. The largest unadjusted correlations were observed for histidine biosynthesis ($r = 0.468$), heme biosynthesis ($r = 0.430$), L-cystine transport ($r = 0.425$), threonine biosynthesis ($r = 0.403$), and assimilatory sulfate reduction ($r = 0.402$). LASSO with 141 module-level predictors selected no module.

Among Woodcroft sugar-degradation categories, galacturonate-degradation gene count showed the largest exploratory correlation with acidification ($r = 0.732$; Table 11). Because this screen involved multiple pathways and no multiplicity-adjusted significance value was available, this association is treated as a hypothesis for future validation rather than as confirmatory evidence.

Table 11. Exploratory correlations between sugar-degradation gene counts and total pH decline.

Pathway	<i>n</i> Present	Mean Genes	<i>r</i>
Galactose degradation	22	2.95	0.123
Galacturonate degradation	18	1.27	0.732
Lactose degradation	20	0.95	−0.169
Mannose degradation	22	4.27	−0.174
Sucrose degradation	18	1.36	−0.081
Xylose, isomerase pathway	11	0.91	−0.266
Xylose, oxidoreductase pathway	10	0.45	−0.259
Other fructose/fucose/xylose pathways	0	0	n.a.

3.4.6. Complementary Outcomes, Sensitivity, and Species Confounding

Total functional-gene count was not significantly associated with lactic acid production or the proteolytic index (Table 12). Correlations between these phenotypes and individual genomic categories were also weak ($|r| < 0.34$).

Table 12. Mixed-effects models for complementary phenotypic outcomes.

Outcome	Predictor	<i>b</i>	<i>p</i>
Lactic acid production	Standardized total functional-gene count	−0.039	0.483
Proteolytic index	Standardized total functional-gene count	0.528	0.241

Excluding *E. durans* and *W. viridescens*, each represented by a single strain group, did not materially alter the direction or magnitude of the main coefficients and correlations (Table 13).

Table 13. Sensitivity analysis excluding species represented by a single strain.

Estimate	Complete (<i>n</i> = 22)	Sensitivity (<i>n</i> = 20)
Mixed-model coefficient for total genes	0.135	0.123
Correlation: carbon genes vs. ΔpH_{06}	0.675	0.677
Correlation: transport genes vs. ΔpH_{06}	0.633	0.582
Correlation: acid-stress genes vs. ΔpH_{06}	0.786	0.752

Species differed significantly in acid-stress, bile-stress, total functional-gene, energy-metabolism, carbon-metabolism, and transport-gene counts after FDR correction (Table 14). In contrast, ΔpH_{06} did not differ significantly among species ($p_{\text{FDR}} = 0.620$).

Table 14. Kruskal–Wallis tests of species differences with Benjamini–Hochberg FDR correction.

Variable	<i>H</i>	df	Raw <i>p</i>	FDR <i>p</i>
Acid-stress genes	18.86	7	0.009	0.031
Bile-stress genes	18.03	7	0.012	0.031
Total functional genes	17.73	7	0.013	0.031
Energy-metabolism genes	17.72	7	0.013	0.031
Carbon-metabolism genes	17.71	7	0.013	0.031
Transport genes	17.49	7	0.014	0.031
Nitrogen-metabolism genes	15.25	7	0.033	0.054
Total stress genes	14.84	7	0.038	0.054

Table 14. Cont.

Variable	H	df	Raw p	FDR p
Miscellaneous genes	13.56	7	0.060	0.076
Heat-stress genes	13.21	7	0.067	0.076
Oxidative-stress genes	11.11	7	0.134	0.139
Salt-stress genes	5.37	7	0.615	0.615
ΔpH_{06}	5.32	7	0.620	0.620

4. Discussion

The 59 LAB isolates recovered from artisanal alheira represented eight species across six genera and displayed a marked geographic structure. The dominance of *E. lactis* in Mirandela and the greater species richness in Valpaços may reflect differences in raw materials, processing environments, casings, smoking and ripening practices, and producer-specific microbiota [45]. However, the sampling design was not balanced among municipalities, and geographic patterns should therefore be interpreted as descriptive rather than as proof of local adaptation.

Bread is a defining component of alheira and provides starch-derived substrates for fermentation. The widespread occurrence of GH13 and the presence of multiple sugar-uptake and degradation systems are consistent with adaptation to this carbohydrate-rich matrix [46]. The results also illustrate why pathway annotation must be interpreted at the gene-component level. Although genes associated with the EMP, PPP, and ED routes were detected broadly, several strains lacked enzymes required for complete pathway operation. In *Ln. mesenteroides*, the absence of key EMP enzymes is consistent with reliance on heterofermentative metabolism, while the restricted distribution of transketolase, transaldolase, and the ED-specific aldolase indicates that pathway completeness varies among species [47–49].

The proteolytic and lipolytic repertoires reveal additional technological potential. Proteases and peptidases may release amino acids and peptides that contribute to flavor development, whereas esterases and lipases can generate fatty-acid-derived aroma precursors. However, genomic presence does not establish expression or activity. Regulatory repression, frameshifts, premature stop codons, and condition-dependent transcription can all weaken correspondence between genome annotation and laboratory phenotype [50–53]. This limitation was reflected in the absence of significant associations between total functional-gene count and either lactic acid production or the proteolytic index.

Additionally, safety screening is essential before indigenous LAB are proposed as starter cultures. Resistance-associated transporters were widespread in some strain groups, and genes associated with adhesion or biogenic-amine formation occurred in a smaller subset. These findings do not by themselves demonstrate transferable antimicrobial resistance, virulence, or amine production, but they identify strains that require phenotypic MIC testing, mobility assessment, and direct measurement of amine formation. In particular, *E. lactis*, *W. viridescens*, and *E. durans* require additional caution because they are not QPS-listed in the classification used in this study. Adhesion-related genes such as *acm* and *sgrA* should also be interpreted in genomic context because similar functions can contribute to host interaction without constituting classical virulence determinants [54].

On the combined evidence available, S11 (DC4755), S12 (DC4756), S17 (DC4757), and S19 (DC4777) are reasonable preliminary candidates for further evaluation. Their genomes lacked the resistance-associated, virulence-associated, and biogenic-amine genes summarized in Table 4; S11 and S12 additionally carried several acid- and oxidative-stress functions. Candidate status remains provisional and should be confirmed through

standardized safety testing, growth and acidification trials in an alheira model, sensory assessment, and challenge studies against relevant pathogens.

The corrected strain-level analysis did not support a simple one-to-one relationship between any of the five targeted metabolic pathways and acidification after multiple-testing correction (Table 7). Instead, broader functional categories, particularly transport, carbon and energy metabolism, and acid- and bile-stress responses, showed the strongest associations with total pH decline (Table 7).

This pattern is biologically plausible because acidification by LAB depends on coordinated carbohydrate uptake, central carbon metabolism, and energy transduction, whereas sustained metabolic activity at decreasing extracellular pH requires effective mechanisms for intracellular pH homeostasis and cellular stress adaptation [55–57].

The nearly identical AICc values and Akaike weights obtained for the carbon-plus-transport and stress models (Table 9) indicate that the data provided essentially equivalent support for these candidate models, rather than identifying a single clearly superior explanation [58]. From a biological perspective, this result is consistent with the interpretation that substrate uptake and metabolism, on the one hand, and stress tolerance, on the other, represent complementary components of the acidifying phenotype.

The exploratory methods converged qualitatively but also exposed the limitations of the dataset. At the penalty corresponding to the minimum cross-validated error ($\lambda_{\min}$), LASSO retained carbon- and transport-related predictors in the restricted analysis and acid- and bile-stress predictors in the full analysis. However, no predictors were retained at the more conservative one-standard-error penalty (λ_{1se}), which applies stronger regularization and generally produces a more parsimonious model [59,60].

Moreover, Ridge regression yielded slightly better cross-validated predictive performance than the sparse penalties, whereas partial least-squares regression concentrated much of the multivariate predictive information in the first latent component [61,62]. Random Forest analysis ranked carbon metabolism, energy metabolism, and acid-stress functions as the most influential predictors [41].

Taken together, these results are consistent with a distributed functional signal rather than a stable minimal genomic predictor set. Variable selection may be unstable when the number of candidate predictors is large relative to the number of observations, particularly when predictors are correlated [63]. Moreover, the modest cross-validated and out-of-bag predictive performance, combined with the small number of non-redundant strains, limits the expected generalization of the fitted models [64,65]. These exploratory analyses should therefore be used to prioritize genomic functions and strains for experimental validation rather than as stand-alone predictors or decision tools for starter-culture suitability.

The primary mixed-effects model showed a positive association between total functional-gene diversity and acidification, together with substantial species-level clustering (Table 8). The within-between decomposition showed that this association was detectable primarily between species, whereas no corresponding association was supported within species. This distinction is important because an association estimated from hierarchically structured data may combine within-group and between-group components that differ in magnitude or direction [66,67].

Moreover, the absence of a significant Kruskal–Wallis result for acidification among species (Table 13) was not inconsistent with the mixed-model findings, since the Kruskal–Wallis test evaluates whether the distributions of the response differ among groups [68], whereas the within-between model evaluates the covariance between functional-gene diversity and acidification at distinct levels of the hierarchy. Species may therefore structure the covariance between genomic and phenotypic traits without producing clearly separated marginal acidification distributions.

Future studies should prioritize larger within-species panels and independent collections to determine whether genomic variation predicts acidification within taxa. Finally, broader sampling, long-read or hybrid assemblies, transcriptomic validation, and replicated in-product fermentation trials will be needed to establish robust genotype–phenotype relationships and to distinguish ecological adaptation from sampling effects.

5. Conclusions

Whole-genome sequencing of 59 alheira-derived LAB isolates resolved 24 non-redundant strain groups spanning eight species and revealed substantial diversity in carbohydrate metabolism, transport, stress tolerance, proteolysis, lipolysis, and safety-relevant genes.

Acidification was associated more consistently with broad functions, especially carbon metabolism, transport, energy metabolism, and acid- and bile-stress responses, than with any single targeted pathway. The primary mixed model identified a positive association with total functional-gene count, but the within-between analysis indicated that the signal was mainly between species, and leave-one-species-out validation showed limited generalization to unobserved species. LASSO, Ridge, Random Forest, PLS, and AICc comparisons broadly supported a distributed multivariable signal while also showing that no stable sparse predictor set can be claimed from the present sample.

The candidate strains S11, S12, S17, and S19 represent preliminary candidates identified from the balance between their comparatively favorable genomic safety profiles and their technological characteristics. In contrast, some of the more strongly acidifying strains, including *E. lactis* presented genomic features requiring additional safety evaluation. This apparent trade-off highlights the need to consider technological performance and safety jointly rather than selecting starter cultures on acidification capacity alone.

Before application, these strains require phenotypic confirmation of antimicrobial susceptibility and biogenic amine production, replicated fermentation trials in an alheira matrix, sensory assessment, and pathogen challenge validation. The consolidated genomic, phenotypic, and statistical framework provides a defensible basis for those next-stage experiments.

Author Contributions: Conceptualization, V.C. and U.G.-B.; methodology, N.F., U.G.-B., A.D.C., V.I. and V.C.; software, N.F., U.G.-B., V.I. and V.C.; validation, U.G.-B., V.I., A.D.C. and V.C.; formal analysis, N.F., U.G.-B., A.D.C., V.I. and V.C.; investigation, N.F., A.D.C. and V.I.; resources, U.G.-B., A.D.C. and V.C.; data curation, N.F., U.G.-B., A.D.C., V.I. and V.C.; writing—original draft preparation, N.F.; writing—review and editing, N.F., U.G.-B., A.D.C., V.I. and V.C.; visualization, N.F., U.G.-B., A.D.C., V.I. and V.C.; supervision, U.G.-B., A.D.C. and V.C.; project administration, U.G.-B. and V.C.; funding acquisition, U.G.-B., A.D.C. and V.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Fundação para a Ciência e a Tecnologia (FCT/MCTES, PID-DAC) through CIMO (UID/00690/2025; DOI: 10.54499/UID/00690/2025) and SusTEC (LA/P/0007/2020; DOI: 10.54499/LA/P/0007/2020) and by PRIMA through project PAS-AGRO-PAS (PRIMA/0014/2022).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data supporting the findings are provided in the manuscript and Supplementary Appendix. Sequence reads are available from NCBI under BioProject PRJNA1463802.

Acknowledgments: Nathalia Fernandes acknowledges PhD grant 2025.02955.BD from Fundação para a Ciência e a Tecnologia (FCT).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Complete inventory of LAB isolates (DC4755–DC4813) from artisanal alheira sausages.

Isolate	Species	Origin	ANI (%)	AF	Ref.	Strain
DC4785	<i>Enterococcus lactis</i>	Valpaços	98.75	0.900	<i>a</i>	S1 *
DC4795	<i>Enterococcus lactis</i>	Mirandela (Cachão)	98.76	0.905	<i>a</i>	S1
DC4796	<i>Enterococcus lactis</i>	Mirandela (Cachão)	98.73	0.904	<i>a</i>	S1
DC4797	<i>Enterococcus lactis</i>	Mirandela (Cachão)	98.74	0.904	<i>a</i>	S1
DC4798	<i>Enterococcus lactis</i>	Mirandela (Cachão)	98.75	0.905	<i>a</i>	S1
DC4799	<i>Enterococcus lactis</i>	Mirandela (Cachão)	98.74	0.905	<i>a</i>	S1
DC4800	<i>Enterococcus lactis</i>	Mirandela (Cachão)	98.73	0.907	<i>a</i>	S1
DC4801	<i>Enterococcus lactis</i>	Mirandela (Cachão)	98.74	0.901	<i>a</i>	S1
DC4802	<i>Enterococcus lactis</i>	Mirandela (Cachão)	98.77	0.900	<i>a</i>	S1
DC4803	<i>Enterococcus lactis</i>	Mirandela (Cachão)	98.73	0.899	<i>a</i>	S1
DC4804	<i>Enterococcus lactis</i>	Mirandela (Cachão)	98.72	0.902	<i>a</i>	S1
DC4805	<i>Enterococcus lactis</i>	Mirandela (Cachão)	98.72	0.904	<i>a</i>	S1
DC4806	<i>Enterococcus lactis</i>	Mirandela (Cachão)	98.74	0.901	<i>a</i>	S1
DC4807	<i>Enterococcus lactis</i>	Mirandela (Cachão)	98.74	0.901	<i>a</i>	S1
DC4808	<i>Enterococcus lactis</i>	Mirandela (Cachão)	98.71	0.907	<i>a</i>	S1
DC4809	<i>Enterococcus lactis</i>	Mirandela (Cachão)	98.75	0.904	<i>a</i>	S1
DC4810	<i>Enterococcus lactis</i>	Mirandela (Cachão)	98.78	0.899	<i>a</i>	S1
DC4811	<i>Enterococcus lactis</i>	Mirandela (Cachão)	98.73	0.905	<i>a</i>	S1
DC4812	<i>Enterococcus lactis</i>	Mirandela (Cachão)	98.72	0.899	<i>a</i>	S1
DC4813	<i>Enterococcus lactis</i>	Bragança	98.28	0.903	<i>a</i>	S2 *
DC4765	<i>Leuconostoc mesenteroides</i>	Vinhais	99.17	0.931	<i>b</i>	S3 *
DC4772	<i>Leuconostoc mesenteroides</i>	Valpaços (Póvoa de Lila)	99.42	0.968	<i>b</i>	S4 *
DC4773	<i>Leuconostoc mesenteroides</i>	Valpaços (Póvoa de Lila)	99.23	0.923	<i>b</i>	S5 *
DC4774	<i>Leuconostoc mesenteroides</i>	Valpaços (Póvoa de Lila)	99.23	0.926	<i>b</i>	S5
DC4775	<i>Leuconostoc mesenteroides</i>	Valpaços (Póvoa de Lila)	99.23	0.925	<i>b</i>	S5
DC4776	<i>Leuconostoc mesenteroides</i>	Valpaços (Póvoa de Lila)	99.00	0.948	<i>b</i>	S6 *
DC4779	<i>Leuconostoc mesenteroides</i>	Valpaços (Póvoa de Lila)	98.95	0.952	<i>b</i>	S7 *
DC4780	<i>Leuconostoc mesenteroides</i>	Valpaços (Póvoa de Lila)	98.94	0.952	<i>b</i>	S7
DC4790	<i>Leuconostoc mesenteroides</i>	Vinhais (Espinheiro)	98.95	0.949	<i>b</i>	S7
DC4791	<i>Leuconostoc mesenteroides</i>	Vinhais (Espinheiro)	98.96	0.949	<i>b</i>	S7
DC4781	<i>Leuconostoc mesenteroides</i>	Valpaços (Póvoa de Lila)	99.57	0.947	<i>b</i>	S8 *
DC4782	<i>Leuconostoc mesenteroides</i>	Mirandela	99.31	0.953	<i>b</i>	S9 *
DC4789	<i>Leuconostoc mesenteroides</i>	Vinhais (Espinheiro)	98.88	0.918	<i>b</i>	S10 *
DC4755	<i>Lactiplantibacillus plantarum</i>	Vimioso	98.91	0.914	<i>c</i>	S11 *
DC4756	<i>Lactiplantibacillus plantarum</i>	Vimioso	99.09	0.916	<i>c</i>	S12 *
DC4758	<i>Lactiplantibacillus plantarum</i>	Vimioso	99.10	0.910	<i>c</i>	S12
DC4759	<i>Lactiplantibacillus plantarum</i>	Vimioso	99.13	0.912	<i>c</i>	S12
DC4760	<i>Lactiplantibacillus plantarum</i>	Vimioso	99.13	0.911	<i>c</i>	S12
DC4792	<i>Lactiplantibacillus plantarum</i>	Bragança	99.73	0.945	<i>c</i>	S13 *
DC4793 [†]	<i>Lactiplantibacillus plantarum</i>	Bragança	99.71	0.935	<i>c</i>	S13
DC4794 [†]	<i>Lactiplantibacillus plantarum</i>	Bragança	99.70	0.936	<i>c</i>	S13
DC4763	<i>Latilactobacillus sakei</i>	Vinhais	98.79	0.929	<i>d</i>	S14 *
DC4764	<i>Latilactobacillus sakei</i>	Vinhais	98.82	0.923	<i>d</i>	S14
DC4766	<i>Latilactobacillus sakei</i>	Vinhais	98.77	0.926	<i>d</i>	S14

Table A1. Cont.

Isolate	Species	Origin	ANI (%)	AF	Ref.	Strain
DC4767	<i>Latilactobacillus sakei</i>	Vinhais	98.75	0.924	^d	S14
DC4768	<i>Latilactobacillus sakei</i>	Vinhais	98.78	0.923	^d	S14
DC4769	<i>Latilactobacillus sakei</i>	Vinhais	98.74	0.927	^d	S14
DC4770	<i>Latilactobacillus sakei</i>	Vinhais	97.23	0.920	^d	S15 *
DC4771	<i>Latilactobacillus sakei</i>	Vinhais	97.96	0.906	^d	S16 *
DC4757	<i>Pediococcus pentosaceus</i>	Vimioso	98.74	0.912	^e	S17 *
DC4761 [†]	<i>Pediococcus pentosaceus</i>	Mirandela (Cachão)	98.31	0.883	^e	S18*
DC4777	<i>Pediococcus pentosaceus</i>	Valpaços (Póvoa de Lila)	98.52	0.920	^e	S19 *
DC4778	<i>Pediococcus pentosaceus</i>	Valpaços (Póvoa de Lila)	98.51	0.926	^e	S19
DC4783	<i>Lactocaseibacillus paracasei</i>	Valpaços	98.24	0.869	^f	S20 *
DC4784 [†]	<i>Lactocaseibacillus paracasei</i>	Valpaços	98.11	0.819	^f	S21
DC4786	<i>Lactocaseibacillus paracasei</i>	Valpaços	98.10	0.853	^f	S21 *
DC4787 [‡]	<i>Lactocaseibacillus paracasei</i>	Valpaços	97.92	0.853	^f	S22 *
DC4762	<i>Weissella viridescens</i>	Mogadouro	98.76	0.967	^g	S23 *
DC4788	<i>Enterococcus durans</i>	Valpaços	98.11	0.849	^h	S24 *

ANI: FastANI identity (%) to closest reference genome; AF: alignment fraction; Strain: strain group assigned by pairwise ANI $\geq 99.9\%$ within species (asterisk marks the representative retained for descriptive genomic summaries; DC4786 was used as the unflagged representative of S21 in inferential analyses). [†] Insufficient amino acids in MSA (<5%). [‡] >37.5% of markers with multiple hits. Reference accessions: ^a GCF_015751045.1; ^b GCF_000014445.1; ^c GCF_014131735.1; ^d GCF_002370355.1; ^e GCF_001437285.1; ^f GCF_000829035.1; ^g GCF_001437355.1; ^h GCF_001544215.1.

Table A2. Presence (+) and absence (−) of protease and proteinase genes across the 24 non-redundant LAB strain groups. Strains are grouped by species.

Gene	Function	E. Lac.			Ln. Mes.					L. Pla.				Lt. Sak.			P. Pen.			Lb. Par.			W. Vir.	E. Dur.	
		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24
ATP-dependent proteases																									
clpP	Clp protease, proteolytic subunit	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
clpX	Clp protease, ATPase subunit ClpX	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
clpE	Clp protease, ATPase subunit ClpE	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
clpL	ATP-dependent protease ClpL	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+
clpC	Clp protease, ATPase subunit ClpC	−	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	−	−	−	−
clpY	ATP-dependent protease, ATPase subunit	+	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+
hslV	ATP-dependent protease subunit HslV	+	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+	+	+	+	+	−	+	
Serine/membrane proteases																									
htrA	Serine protease HtrA (DegP)	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ftsH	ATP-dependent zinc metalloprotease FtsH	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
gluP	Rhomboid protease GluP	+	−	+	+	+	+	+	+	+	+	+	+	+	−	−	−	+	+	+	+	+	+	−	+
htpX	Protease HtpX	−	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	−	−	+	−
rasP	Sigma-W protease RasP	+	−	−	+	−	−	−	−	−	+	+	+	+	+	+	+	+	+	+	+	+	−	+	
prp	Ribosomal processing cysteine protease	+	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+	+	+	+	+	−	+	
ydeA	Putative protease YdeA	+	−	+	+	+	+	+	+	+	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−
mroQ	Membrane-embedded CAAX protease	+	−	−	−	−	−	−	−	−	+	+	+	−	−	−	−	+	+	−	+	+	−	+	
ctpA	Carboxy-terminal processing protease	−	−	−	−	−	−	−	−	−	+	+	+	−	−	−	−	+	+	+	+	+	−	−	
yraA	Putative cysteine protease	−	−	−	+	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
Proteinases																									
prtP	PII-type cell-envelope proteinase	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+	−	+
Ser. proteinase	Thermophilic serine proteinase	+	+	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	+	−	−	−

Table A3. Presence (+) and absence (−) of peptidase genes across the 24 non-redundant LAB strain groups, part 1 of 2. Strains are grouped by species.

Gene	Function	<i>E. Lac.</i>			<i>Ln. Mes.</i>						<i>L. Pla.</i>			<i>Lt. Sak.</i>			<i>P. Pen.</i>			<i>Lb. Par.</i>			<i>W. Vir.</i>	<i>E. Dur.</i>	
		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24
<i>Aminopeptidases</i>																									
<i>pepC</i>	Aminopeptidase C	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>map</i>	Methionine aminopeptidase	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>pepS</i>	Aminopeptidase PepS	+	+	+	+	+	+	+	+	+	+	−	−	+	+	+	+	+	+	+	+	+	+	+	+
<i>pepN</i>	Aminopeptidase N	−	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	−	−
<i>ypdF</i>	Aminopeptidase YpdF	+	−	+	+	+	+	+	+	+	+	+	+	+	−	−	−	−	−	−	+	+	+	+	+
<i>pepA</i>	Glutamyl aminopeptidase	+	−	+	+	+	+	+	+	+	+	−	−	+	−	−	−	+	−	−	−	+	−	−	+
<i>pepE</i>	Aminopeptidase E	−	−	−	−	−	−	+	−	−	+	+	+	+	+	+	−	+	−	−	+	+	+	−	−
<i>Dipeptidases/endopeptidases</i>																									
<i>pepT</i>	Peptidase T (tripeptidase)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>pepQ</i>	Xaa-Pro dipeptidase	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>pepO</i>	Neutral endopeptidase PepO	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>pepF</i>	Oligopeptidase F	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+
<i>pepF1</i>	Oligoendopeptidase F, plasmid	−	−	+	+	+	+	+	+	+	+	−	−	−	+	+	+	−	+	+	−	−	−	−	+
<i>pepV</i>	β-Ala-Xaa dipeptidase	+	+	−	−	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+	+	+	+	+
<i>pepDA</i>	Dipeptidase A	−	−	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+	+	+	+	+	+	−
<i>pepD</i>	Dipeptidase PepD	+	+	−	−	−	−	−	−	−	−	+	+	+	+	+	+	−	+	−	+	−	+	−	−

Table A3. Cont.

Gene	Function	<i>E. Lac.</i>			<i>Ln. Mes.</i>					<i>L. Pla.</i>					<i>Lt. Sak.</i>			<i>P. Pen.</i>			<i>Lb. Par.</i>			<i>W. Vir.</i>	<i>E. Dur.</i>
		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24
<i>X-Pro & signal peptidases</i>																									
<i>pepX</i>	Xaa-Pro dipeptidyl-peptidase	—	—	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	—
<i>spsB</i>	Signal peptidase IB	+	—	+	+	+	+	+	+	+	+	+	+	+	+	+	—	+	—	—	+	—	+	+	
<i>lspA</i>	Lipoprotein signal peptidase	+	+	—	—	—	—	—	—	—	—	+	+	+	+	+	+	+	+	+	+	+	—	+	
<i>fpaP</i>	Proline iminopeptidase	—	—	—	—	—	—	—	—	—	—	+	+	+	+	+	+	+	+	+	+	+	—	—	
<i>sipV</i>	Signal peptidase I V	+	—	—	—	—	—	—	—	—	—	—	—	+	+	+	—	+	—	—	+	—	—	+	
<i>pepIP</i>	Proline iminopeptidase, plasmid	—	—	—	—	—	—	—	—	—	—	+	+	+	—	—	—	—	—	+	+	+	—	—	

Table A4. Presence (+) and absence (–) of peptidase genes across the 24 non-redundant LAB strain groups, part 2 of 2. Strains are grouped by species.

Gene	Function	<i>E. Lac.</i>			<i>Ln. Mes.</i>						<i>L. Pla.</i>			<i>Lt. Sak.</i>			<i>P. Pen.</i>			<i>Lb. Par.</i>			<i>W. Vir.</i>	<i>E. Dur.</i>	
		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24
<i>Cell-wall peptidases</i>																									
<i>dacA</i>	D-Ala-D-Ala carboxypeptidase	+	—	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>ypwA</i>	Carboxypeptidase 1	+	—	+	+	+	+	+	+	+	+	—	—	+	—	—	—	—	—	—	—	+	—	+	+
<i>yfeW</i>	Putative D-Ala-D-Ala carboxypep.	—	—	—	—	—	—	—	—	—	—	+	+	+	—	—	+	+	+	+	+	+	+	+	+
<i>dap</i>	D-aminopeptidase	—	—	+	+	+	+	+	+	+	+	—	—	—	—	—	—	—	—	—	+	+	+	—	—
<i>yciB</i>	L,D-transpeptidase YciB	—	—	—	—	—	—	—	—	—	—	+	+	+	+	+	+	—	+	—	+	+	+	—	—
<i>comC</i>	Prepilin leader peptidase	+	—	—	—	—	—	—	—	—	—	+	+	+	+	+	+	—	+	—	—	+	—	—	—
<i>pcp</i>	Pyrrolidone-carboxylate peptidase	+	—	—	—	—	—	—	—	—	—	—	—	+	—	—	—	—	—	—	+	+	+	—	+
<i>ysdC</i>	Putative aminopeptidase YsdC	+	—	—	—	—	—	—	—	—	—	—	—	+	+	—	—	—	+	—	—	+	—	—	+
<i>sppA</i>	Signal peptide peptidase SppA	+	+	—	—	—	—	—	—	—	—	—	—	+	—	—	—	—	—	—	—	+	—	—	+
<i>iadA</i>	Isoaspartyl dipeptidase	+	—	—	—	—	—	—	—	—	—	—	—	+	—	—	—	—	+	—	—	+	—	—	+

Table A5. Presence (+) and absence (–) of stress tolerance genes across the 24 non-redundant LAB strains (DC4755–DC4813). Strains are grouped by species. Stress categories: heat shock, oxidative stress, acid tolerance, bile tolerance, and salt tolerance.

Gene	Function	<i>E. Lac.</i>			<i>Ln. Mes.</i>						<i>L. Pla.</i>				<i>Lt. Sak.</i>			<i>P. Pen.</i>			<i>Lb. Par.</i>			<i>W. Vir.</i>	<i>E. Dur.</i>
		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24
<i>Heat shock</i>																									
<i>clpP</i>	Clp protease proteolytic subunit	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>clpB</i>	Chaperone ClpB (disaggregase)	+	+	—	—	—	—	—	—	—	+	+	+	+	+	+	+	+	+	+	+	+	+	—	+
<i>dnaK</i>	Chaperone DnaK (Hsp70)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>dnaJ</i>	Co-chaperone DnaJ (Hsp40)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>grpE</i>	Nucleotide exchange factor GrpE	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>groEL</i>	Chaperonin GroEL (Hsp60)	+	—	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>groES</i>	Co-chaperonin GroES (Hsp10)	+	—	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Oxidative stress</i>																									
<i>trxB</i>	Thioredoxin/thioredoxin reductase	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>cydD/kefB</i>	Glutathione transport system	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>ahpC</i>	Alkyl hydroperoxide reductase C	+	+	+	+	+	+	+	+	+	+	—	—	+	—	—	—	—	—	—	+	+	+	—	+
<i>sodA</i>	Superoxide dismutase [Mn]	+	+	—	—	—	—	—	—	—	—	—	—	+	+	+	+	—	+	—	+	+	—	—	+
<i>katA</i>	Vegetative catalase	—	—	—	—	—	—	—	—	—	+	+	+	+	+	+	+	—	+	—	—	—	—	+	—

Table A5. Cont.

Gene	Function	E. Lac.			Ln. Mes.					L. Pla.					Lt. Sak.			P. Pen.			Lb. Par.			W. Vir.	E. Dur.
		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24
Acid tolerance																									
F ₀ F ₁ -ATPase	Proton-translocating ATPase	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
arcA	Arginine deiminase system	+	+	+	+	+	+	+	+	+	+	−	−	+	+	+	+	+	+	+	−	+	−	+	+
gadB	Glutamate decarboxylase	−	−	−	−	−	−	−	−	−	−	+	+	+	−	−	−	−	−	−	−	−	−	−	−
Bile tolerance																									
bsh	Bile salt hydrolase	+	−	−	−	−	−	−	−	−	−	+	+	+	−	−	−	−	−	−	−	+	−	−	+
cbh	Choloylglycine hydrolase	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
Salt tolerance																									
betL/opuC	Glycine betaine transporter	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
opuCD	Betaine/carnitine/choline transport	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
proV	Proline/betaine transport ATPase	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
opuA	Glycine betaine ABC transporter OpuA	+	−	+	+	+	+	+	+	+	+	−	−	+	−	−	−	−	+	−	−	+	−	−	+

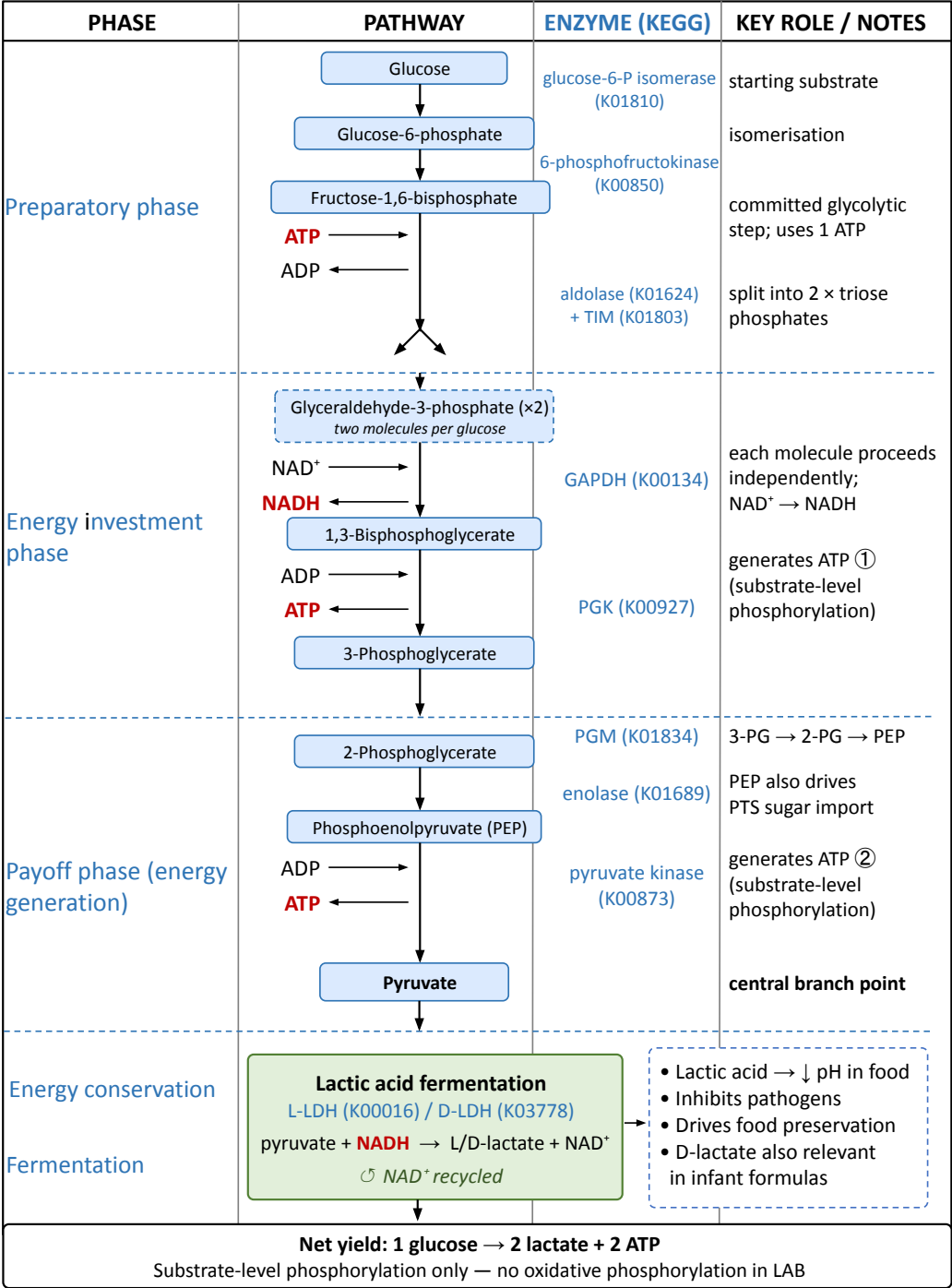
Table A6. Presence (+) and absence (−) of metabolic genes across the 24 non-redundant LAB strains. Strains are grouped by species.

Pathway	Gene	Function	E. Lac.			Ln. Mes.					L. Pla.					Lt. Sak.			P. Pen.			Lb. Par.			W. Vir.	E. Dur.
			S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24
Diacetyl	aldC	Acetolactate decarb.	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	alsS	Acetolactate synthase	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Citrate	citC	Citrate lyase ligase	−	−	+	−	−	−	−	−	−	−	+	+	+	+	−	−	−	+	−	+	+	+	+	+
	citD	Citrate lyase ACP	−	−	+	−	−	−	−	−	−	−	+	+	+	+	−	−	−	+	−	+	+	+	+	+
	citE	Citrate lyase β	−	−	+	−	−	−	−	−	−	−	+	+	+	+	−	−	−	+	−	+	+	+	+	+
	citF	Citrate lyase α	−	−	+	−	−	−	−	−	−	−	+	+	+	+	−	−	−	+	−	+	+	+	+	+
	citP	Citrate permease	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−
Acetaldehyde	aldB	Acetaldehyde reductase	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+	−	−
	aldehyde DH	Aldehyde dehydrogenase	+	+	+	+	+	+	+	+	+	+	+	+	+	−	−	−	+	+	+	+	+	+	+	+
EPS	eps	EPS biosynthesis	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	gtf	Glucosyltransferase	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	−	+	+	+	+	+
Lactose	galE	UDP-glucose 4-epimerase	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	galK	Galactokinase	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	galT	Gal-1-P uridylyltransferase	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+
	lacA	Galactoside O-acetyltransf.	−	−	−	−	−	−	−	−	−	−	+	+	+	−	−	−	−	−	−	+	+	+	−	+
	lacX	Lactose metabolism	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	lacY	Lactose permease	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
	lacZ	β-galactosidase	−	−	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−

Table A7. Proteolytic activity, lactic acid concentration and pH kinetics of the 24 non-redundant LAB strains (DC4755–DC4813). Values represent the range (min–max) across isolates within a strain group; single-isolate strains show one value. *n*: number of isolates in strain group. PI: proteolytic index, area of clearing zone on milk agar (mm). [LAC]: lactic acid concentration after 24 h (g L^{−1}). pH_{*t*}: pH at time *t* (h). ΔpH: pH drop between time points. *k*: Gompertz lag time (h); *P*: total acidification amplitude.

Strain	Species	<i>n</i>	Proteolytic			pH Kinetics							
			PI (mm)	[LAC] (g L ^{−1})	pH ₀	pH ₃	pH ₄	pH ₆	ΔpH _{0–3}	ΔpH _{3–6}	ΔpH _{0–6}	<i>k</i> (h)	<i>P</i>
S1	<i>E. lactis</i>	19	0.00–1.11	0.056–0.591	6.59–6.68	6.24–6.40	6.07–6.26	5.71–5.90	0.21–0.38	0.49–0.54	0.71–0.92	2 × 10 ¹ –2.6 × 10 ¹	1.31–1.80
S2	<i>E. lactis</i>	1	0.00	0.282	6.66	6.28	6.12	5.79	0.38	0.49	0.87	2.8 × 10 ¹	1.27
S3	<i>Ln. mesenteroides</i>	1	0.00	0.199	6.57	6.46	6.45	6.42	0.11	0.04	0.15	3.1 × 10 ⁴	0.44
S4	<i>Ln. mesenteroides</i>	1	4.84	0.028	6.60	6.42	6.39	6.34	0.18	0.09	0.26	1.4 × 10 ³	0.59
S5	<i>Ln. mesenteroides</i>	3	0.00–3.53	0.020–0.217	6.66–6.72	6.42–6.56	6.38–6.52	6.33–6.45	0.15–0.28	0.09–0.12	0.27–0.37	3 × 10 ² –7.9 × 10 ³	0.40–0.82
S6	<i>Ln. mesenteroides</i>	1	4.37	0.032	6.58	6.47	6.43	6.36	0.11	0.11	0.22	1.8 × 10 ²	0.99
S7	<i>Ln. mesenteroides</i>	4	0.00–8.26	0.011–0.031	6.66–6.72	6.41–6.57	6.28–6.53	6.00–6.45	0.11–0.30	0.09–0.42	0.23–0.72	3.2 × 10 ¹ –2.2 × 10 ³	0.51–1.30
S8	<i>Ln. mesenteroides</i>	1	5.37	0.216	6.58	6.45	6.42	6.37	0.14	0.08	0.22	9.2 × 10 ²	0.68
S9	<i>Ln. mesenteroides</i>	1	5.17	0.015	6.62	6.50	6.47	6.40	0.12	0.10	0.22	3 × 10 ²	0.88
S10	<i>Ln. mesenteroides</i>	1	3.45	0.019	6.71	6.53	6.49	6.43	0.18	0.10	0.28	7.9 × 10 ²	0.64
S11	<i>L. plantarum</i>	1	3.82	0.677	6.70	6.55	6.52	6.47	0.16	0.08	0.23	2 × 10 ³	0.58
S12	<i>L. plantarum</i>	4	0.00–4.47	0.248–0.586	6.64–6.71	6.53–6.56	6.48–6.51	6.27–6.48	0.11–0.17	0.06–0.28	0.16–0.40	2.8 × 10 ¹ –1 × 10 ⁴	0.46–1.79
S13	<i>L. plantarum</i>	3	1.78–2.20	0.361–0.515	6.58–6.62	6.29–6.37	6.14–6.22	5.80–5.86	0.26–0.33	0.49–0.50	0.75–0.82	2.2 × 10 ¹ –2.6 × 10 ¹	1.37–1.63
S14	<i>Lt. sakei</i>	6	0.00–4.29	0.016–0.620	6.66–6.72	6.42–6.59	6.40–6.56	6.37–6.52	0.08–0.30	0.04–0.08	0.15–0.34	2.4 × 10 ² –9.7 × 10 ⁷	0.18–1.02
S15	<i>Lt. sakei</i>	1	3.21	0.563	6.68	6.53	6.48	6.39	0.15	0.15	0.30	1.3 × 10 ²	1.00
S16	<i>Lt. sakei</i>	1	0.00	0.651	6.72	6.46	6.44	6.42	0.26	0.04	0.30	5.8 × 10 ⁶	0.22
S17	<i>P. pentosaceus</i>	1	0.00	0.560	6.70	6.55	6.52	6.47	0.15	0.07	0.23	2 × 10 ³	0.58
S18	<i>P. pentosaceus</i>	1	0.85	0.148	6.71	6.53	6.51	6.48	0.18	0.05	0.23	6.7 × 10 ⁴	0.36
S19	<i>P. pentosaceus</i>	2	4.74–5.48	0.029–0.663	6.59–6.69	6.45–6.46	6.43	6.38–6.41	0.14–0.23	0.04–0.08	0.18–0.31	4 × 10 ³ –4.7 × 10 ⁴	0.40–0.47
S20	<i>Lb. paracasei</i>	1	2.30	0.644	6.59	6.32	6.17	5.82	0.26	0.50	0.76	2.1 × 10 ¹	1.59
S21	<i>Lb. paracasei</i>	2	4.81–4.97	0.047–0.646	6.55	6.47–6.50	6.46–6.47	6.40–6.44	0.06–0.08	0.03–0.10	0.11–0.16	7.7 × 10 ¹ –5 × 10 ⁴	0.45–1.46
S22	<i>Lb. paracasei</i>	1	5.00	0.056	6.68	6.40	6.25	5.88	0.27	0.53	0.80	2.1 × 10 ¹	1.62
S23	<i>W. viridescens</i>	1	0.00	0.308	6.69	6.55	6.52	6.45	0.14	0.10	0.24	3.1 × 10 ²	0.84
S24	<i>E. durans</i>	1	3.70	0.068	6.59	6.29	6.14	5.79	0.30	0.49	0.79	2.4 × 10 ¹	1.47

Glycolysis → Fermentation: the core LAB energy engine



Abbreviations: ATP, adenosine triphosphate; PEP, phosphoenolpyruvate; PTS, phosphotransferase system; NAD⁺/NADH, nicotinamide adenine dinucleotide (oxidised/reduced); LDH, lactate dehydrogenase.

Figure A1. Overview of glycolytic and fermentative routes annotated in the LAB collection.

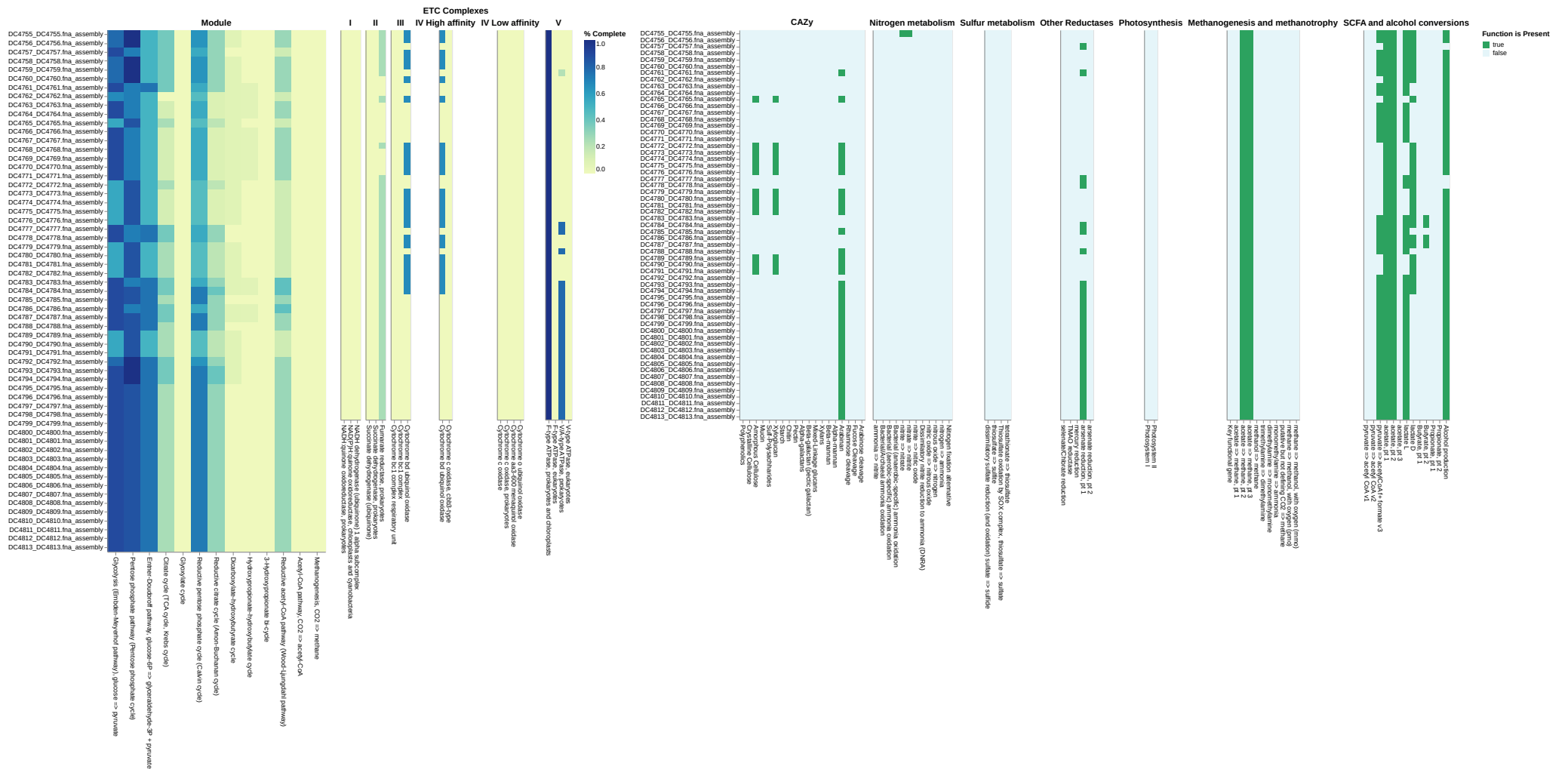


Figure A2. Summary heatmap of carbohydrate-active-enzyme annotations across the 59 LAB isolates.

References

1. Ferreira, V.; Barbosa, J.; Silva, J.; Felício, M.T.; Mena, C.; Hogg, T.; Gibbs, P.; Teixeira, P. Characterisation of Alheiras, Traditional Sausages Produced in the North of Portugal, with Respect to Their Microbiological Safety. *Food Control* **2007**, *18*, 436–440. <https://doi.org/10.1016/j.foodcont.2005.11.011>.
2. Patarata, L.; Judas, I.; Silva, J.A.; Esteves, A.; Martins, C. A Comparison of the Physicochemical and Sensory Characteristics of Alheira Samples from Different-Sized Producers. *Meat Sci.* **2008**, *79*, 131–138. <https://doi.org/10.1016/j.meatsci.2007.08.009>.
3. Albano, H.; Oliveira, M.; Aroso, R.; Cubero, N.; Hogg, T.; Teixeira, P. Antilisterial Activity of Lactic Acid Bacteria Isolated from “Alheiras” (Traditional Portuguese Fermented Sausages): In Situ Assays. *Meat Sci.* **2007**, *76*, 796–800. <https://doi.org/10.1016/j.meatsci.2007.01.019>.
4. Fernandes, N.; Faria, A.S.; Carvalho, L.; Choupina, A.; Rodrigues, C.; Gonzales-Barron, U.; Cadavez, V. Genetic Identification and Technological Potential of Indigenous Lactic Acid Bacteria Isolated from Alheira, a Traditional Portuguese Sausage. *Foods* **2024**, *13*, 598. <https://doi.org/10.3390/foods13040598>.
5. Albano, H.; Henriques, I.; Correia, A.; Hogg, T.; Teixeira, P. Characterization of Microbial Population of ‘Alheira’ (a Traditional Portuguese Fermented Sausage) by PCR-DGGE and Traditional Cultural Microbiological Methods. *J. Appl. Microbiol.* **2008**, *105*, 2187–2194. <https://doi.org/10.1111/j.1365-2672.2008.03947.x>.
6. Esteves, A.; Patarata, L.; Saraiva, C.; Martins, C. Assessment of the Microbiological Characteristics of Industrially Produced Alheira, with Particular Reference to Foodborne Pathogens. *J. Food Saf.* **2008**, *28*, 88–102. <https://doi.org/10.1111/j.1745-4565.2007.00097.x>.
7. Bouvard, V.; Loomis, D.; Guyton, K.Z.; Grosse, Y.; El Ghissassi, F.; Benbrahim-Tallaa, L.; Guha, N.; Mattock, H.; Straif, K. Carcinogenicity of Consumption of Red and Processed Meat. *Lancet Oncol.* **2015**, *16*, 1599–1600. [https://doi.org/10.1016/S1470-2045\(15\)00444-1](https://doi.org/10.1016/S1470-2045(15)00444-1).
8. Inguglia, E.S.; Song, Z.; Kerry, J.P.; O’Sullivan, M.G.; Hamill, R.M. Addressing Clean Label Trends in Commercial Meat Processing: Strategies, Challenges and Insights from Consumer Perspectives. *Foods* **2023**, *12*, 2062. <https://doi.org/10.3390/foods12102062>.
9. Doyle, M.P.; Steenson, L.R.; Meng, J. Bacteria in Food and Beverage Production. In *The Prokaryotes: Applied Bacteriology and Biotechnology*; Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., Thompson, F., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 241–256. https://doi.org/10.1007/978-3-642-31331-8_27.
10. Fernandes, N.; Achemchem, F.; Gonzales-Barron, U.; Cadavez, V. Biopreservation strategies using bacteriocins to control meat spoilage and foodborne outbreaks. *Ital. J. Food Saf.* **2024**, *13*, 12558. <https://doi.org/10.4081/IJFS.2024.12558>.
11. Dou, X.; Liu, Y.; Koutsoumanis, K.; Song, C.; Li, Z.; Zhang, H.; Yang, F.; Zhu, H.; Dong, Q. Employing genome-wide association studies to investigate acid adaptation mechanisms in *Listeria monocytogenes*. *Food Res. Int.* **2024**, *196*, 115106. <https://doi.org/10.1016/J.FOODRES.2024.115106>.
12. Buron-Moles, G.; Chailyan, A.; Dolejs, I.; Forster, J.; Mikš, M.H. Uncovering carbohydrate metabolism through a genotype-phenotype association study of 56 lactic acid bacteria genomes. *Appl. Microbiol. Biotechnol.* **2019**, *103*, 3135–3152. <https://doi.org/10.1007/S00253-019-09701-6>.
13. de Sena Brandine, G.; Smith, A.D. Falco: High-Speed FastQC Emulation for Quality Control of Sequencing Data. *F1000Research* **2019**, *8*, 1874. <https://doi.org/10.12688/f1000research.21142.2>.
14. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: An Ultra-Fast All-in-One FASTQ Preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
15. Seemann, T. Shovill: Fast De Novo Assembly Pipeline for Illumina Paired Reads, 2023. Available online: <https://github.com/tseemann/shovill> (accessed on 8 April 2026).
16. Wick, R.R.; Schultz, M.B.; Zobel, J.; Holt, K.E. Bandage: Interactive Visualization of De Novo Genome Assemblies. *Bioinformatics* **2015**, *31*, 3350–3352. <https://doi.org/10.1093/bioinformatics/btv383>.
17. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
18. Jain, C.; Rodriguez-R, L.M.; Phillippy, A.M.; Konstantinidis, K.T.; Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **2018**, *9*, 5114. <https://doi.org/10.1038/s41467-018-07641-9>.
19. Chaumeil, P.A.; Mussig, A.J.; Hugenholtz, P.; Parks, D.H. GTDB-Tk v2: Memory friendly classification with the genome taxonomy database. *Bioinformatics* **2022**, *38*, 5315–5316. <https://doi.org/10.1093/bioinformatics/btac672>.
20. Shaffer, M.; Borton, M.A.; Bolduc, B.; Faria, J.P.; Flynn, R.M.; Ghadermazi, P.; Edirisinghe, J.N.; Wood-Charlson, E.M.; Miller, C.S.; Chan, S.H.J.; et al. kbDRAM: Annotation and metabolic profiling of genomes with DRAM in KBase. *Bioinformatics* **2023**, *39*, btad110. <https://doi.org/10.1093/bioinformatics/btad110>.
21. Arkin, A.P.; Cottingham, R.W.; Henry, C.S.; Harris, N.L.; Stevens, R.L.; Maslov, S.; Dehal, P.; Ware, D.; Perez, F.; Canon, S.; et al. KBase: The United States department of energy systems biology knowledgebase. *Nat. Biotechnol.* **2018**, *36*, 566–569. <https://doi.org/10.1038/nbt.4163>.

22. Revell, L.J. phytools 2.0: An updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ* **2024**, *12*, e16505. <https://doi.org/10.7717/peerj.16505>.
23. Seemann, T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* **2014**, *30*, 2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
24. Feldgarden, M.; Brover, V.; Gonzalez-Escalona, N.; Frye, J.G.; Haendiges, J.; Haft, D.H.; Hoffmann, M.; Pettengill, J.B.; Prasad, A.B.; Tillman, G.E.; et al. AMRFinderPlus and the Reference Gene Catalog Facilitate Examination of the Genomic Links Among Antimicrobial Resistance, Stress Response, and Virulence. *Sci. Rep.* **2021**, *11*, 12728. <https://doi.org/10.1038/s41598-021-91456-0>.
25. Seemann, T. Abricate: Mass Screening of Contigs for Antimicrobial Resistance or Virulence Genes, 2023. Available online: <https://github.com/tseemann/abricate> (accessed on 8 April 2026).
26. pandas development team, T. *Pandas-Dev/Pandas: Pandas*; Zenodo: Geneva, Switzerland, 2020. <https://doi.org/10.5281/zenodo.3509134>.
27. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362. <https://doi.org/10.1038/s41586-020-2649-2>.
28. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>.
29. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. <https://doi.org/10.1109/MCSE.2007.55>.
30. Waskom, M.L. seaborn: Statistical data visualization. *J. Open Source Softw.* **2021**, *6*, 3021. <https://doi.org/10.21105/joss.03021>.
31. Molder, F.; Jablonski, K.; Letcher, B.; Hall, M.; Tomkins-Tinch, C.; Sochat, V.; Forster, J.; Lee, S.; Twardziok, S.; Kanitz, A.; et al. Sustainable data analysis with Snakemake [version 1; peer review: 1 approved, 1 approved with reservations]. *F1000Research* **2021**, *10*, 33. <https://doi.org/10.12688/f1000research.29032.1>.
32. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2025.
33. Wickham, H.; Bryan, J. *readxl: Read Excel Files*; R package. 2026. Available online: <https://readxl.tidyverse.org> (accessed on 26 August 2026).
34. Wickham, H.; François, R.; Henry, L.; Müller, K.; Vaughan, D. *dplyr: A Grammar of Data Manipulation*; R package. 2026. Available online: <https://cran.ma.imperial.ac.uk/web/packages/dplyr/dplyr.pdf> (accessed on 26 August 2026).
35. Wickham, H.; Vaughan, D.; Girlich, M. *tidyr: Tidy Messy Data*; R package. 2026. Available online: <https://cran.r-project.org/web/packages/tidyr/index.html> (accessed on 26 August 2026).
36. Wickham, H. *stringr: Simple, Consistent Wrappers for Common String Operations*; R package. 2025. Available online: <https://cran.r-project.org/web/packages/stringr/index.html> (accessed on 26 August 2026).
37. Bates, D.; Mächler, M.; Bolker, B.M.; Walker, S.C. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **2015**, *67*, 1–48. <https://doi.org/10.18637/jss.v067.i01>.
38. Kuznetsova, A.; Brockhoff, P.B.; Christensen, R.H.B. lmerTest Package: Tests in Linear Mixed Effects Models. *J. Stat. Softw.* **2017**, *82*, 1–26. <https://doi.org/10.18637/jss.v082.i13>.
39. Lüdtke, D.; Ben-Shachar, M.S.; Patil, I.; Waggoner, P.; Makowski, D. performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *J. Open Source Softw.* **2021**, *6*, 3139. <https://doi.org/10.21105/joss.03139>.
40. Lüdtke, D.ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *J. Open Source Softw.* **2018**, *3*, 772. <https://doi.org/10.21105/joss.00772>.
41. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. <https://doi.org/10.1023/A:1010933404324>.
42. Bartoň, K. *MuMIn: Multi-Model Inference*; R package. 2026. Available online: <https://cran.r-project.org/web/packages/MuMIn/index.html> (accessed on 26 August 2026).
43. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016. <https://doi.org/10.1007/978-3-319-24277-4>.
44. Pedersen, T.L. *patchwork: The Composer of Plots*; R package. 2025. Available online: <https://patchwork.data-imaginist.com> (accessed on 26 August 2026).
45. Coelho-Fernandes, S.; Zefanias, O.; Rodrigues, G.; Faria, A.S.; Fernandes, A.; Barros, L.; Cadavez, V.; Gonzales-Barron, U. Microbiological and Physicochemical Assessment of Artisanally Produced “Alheira” Fermented Sausages in Northern Portugal. *Proceedings* **2020**, *70*, 16. https://doi.org/10.3390/foods_2020-07627.
46. Plaza-Vinuesa, L.; Hernandez-Hernandez, O.; Moreno, F.J.; Rivas, B.D.L.; Muñoz, R. Unravelling the diversity of glycoside hydrolase family 13 α -amylases from *Lactobacillus plantarum* WCFS1. *Microb. Cell Factories* **2019**, *18*, 183. <https://doi.org/10.1186/S12934-019-1237-3>.
47. Wang, Q.; Ingram, L.O.; Shanmugam, K.T. Evolution of D-lactate dehydrogenase activity from glycerol dehydrogenase and its utility for D-lactate production from lignocellulose. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 18920–18925. <https://doi.org/10.1073/pnas.1111085108>.

48. Wilkinson, H.C.; Dalby, P.A. The Two-Species Model of transketolase explains donor substrate-binding, inhibition and heat-activation. *Sci. Rep.* **2020**, *10*, 4148. <https://doi.org/10.1038/s41598-020-61175-z>.
49. Widenmann, A.; Schiffer, C.; Ehrmann, M.; Vogel, R. Impact of different sugars and glycosyltransferases on the assertiveness of *Lactobacillus sakei* in raw sausage fermentations. *Int. J. Food Microbiol.* **2022**, *366*, 109575. <https://doi.org/10.1016/j.ijfoodmicro.2022.109575>.
50. Sierra, G. A simple method for the detection of lipolytic activity of micro-organisms and some observations on the influence of the contact between cells and fatty substrates. *Antonie Van Leeuwenhoek* **1957**, *23*, 15–22. <https://doi.org/10.1007/BF02545855>.
51. Yang, H.; Hao, L.; Yao, S.; Zhou, R.; Wu, C. Rational Engineering of Arginine Deiminase To Enhance the Acid Tolerance of Lactic Acid Bacteria. *J. Agric. Food Chem.* **2025**, *73*, 25539–25548. <https://doi.org/10.1021/ACS.JAFC.5C09676>.
52. Yu, P.; Zhao, Y.; Jiang, Y.; Yang, Y.; Liu, X.; Zhang, H.; Zhao, J.; Lee, Y.K.; Zhang, H.; Chen, W. Capacity of soybean carbohydrate metabolism in *Leuconostoc mesenteroides*, *Lactococcus lactis* and *Streptococcus thermophilus*. *Food Biosci.* **2021**, *44*, 101381. <https://doi.org/10.1016/j.fbio.2021.101381>.
53. Usman, S.; Xu, D.; Ma, J.; Sheoran, N.; Okoye, C.O.; Guo, X. Comparative Genomics Reveals the Molecular Mechanisms of a Newly Isolated *Pediococcus cellicola* zy165 Strain and Its Adaptation in Corn Silage. *Biochem. Genet.* **2026**, *64*, 1999–2019. <https://doi.org/10.1007/s10528-025-11114-2>.
54. Shu, H.; He, X.; Hong, Z.; Dong, K.; Zou, Y.; Cao, M.; Wang, R.; Xu, Y.; Liao, L.; Zuo, H.; et al. Screening and genome analysis of potential probiotic lactic acid bacteria with broad-spectrum antibacterial activity from Sichuan sun-dried vinegar grains (Cupei). *LWT* **2024**, *202*, 116288. <https://doi.org/10.1016/j.lwt.2024.116288>.
55. Hutkins, R.W.; Nannen, N.L. pH Homeostasis in Lactic Acid Bacteria. *J. Dairy Sci.* **1993**, *76*, 2354–2365. [https://doi.org/10.3168/jds.S0022-0302\(93\)77573-6](https://doi.org/10.3168/jds.S0022-0302(93)77573-6).
56. Konings, W.N.; Lolkema, J.S.; Bolhuis, H. The Role of Transport Processes in Survival of Lactic Acid Bacteria: Energy Transduction and Multidrug Resistance. *Antonie Van Leeuwenhoek* **1997**, *71*, 117–128. <https://doi.org/10.1023/A:1000143525601>.
57. Papadimitriou, K.; Alegría, Á.; Bron, P.A.; De Angelis, M.; Gobbetti, M.; Kleerebezem, M.; Lemos, J.A.; Linares, D.M.; Ross, R.P.; Stanton, C.; et al. Stress Physiology of Lactic Acid Bacteria. *Microbiol. Mol. Biol. Rev.* **2016**, *80*, 837–890. <https://doi.org/10.1128/MMBR.00076-15>.
58. Burnham, K.P.; Anderson, D.R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed.; Springer: New York, NY, USA, 2002. <https://doi.org/10.1007/b97636>.
59. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B (Methodol.)* **1996**, *58*, 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>.
60. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–22. <https://doi.org/10.18637/jss.v033.i01>.
61. Hoerl, A.E.; Kennard, R.W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* **1970**, *12*, 55–67. <https://doi.org/10.1080/00401706.1970.10488634>.
62. Wold, S.; Sjöström, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1).
63. Meinshausen, N.; Bühlmann, P. Stability Selection. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2010**, *72*, 417–473. <https://doi.org/10.1111/j.1467-9868.2010.00740.x>.
64. Varma, S.; Simon, R. Bias in Error Estimation When Using Cross-Validation for Model Selection. *BMC Bioinform.* **2006**, *7*, 91. <https://doi.org/10.1186/1471-2105-7-91>.
65. Vabalas, A.; Gowen, E.; Poliakoff, E.; Casson, A.J. Machine Learning Algorithm Validation with a Limited Sample Size. *PLoS ONE* **2019**, *14*, e0224365. <https://doi.org/10.1371/journal.pone.0224365>.
66. Mundlak, Y. On the Pooling of Time Series and Cross Section Data. *Econometrica* **1978**, *46*, 69–85. <https://doi.org/10.2307/1913646>.
67. Bell, A.; Fairbrother, M.; Jones, K. Fixed and Random Effects Models: Making an Informed Choice. *Qual. Quant.* **2019**, *53*, 1051–1074. <https://doi.org/10.1007/s11135-018-0802-x>.
68. Kruskal, W.H.; Wallis, W.A. Use of Ranks in One-Criterion Variance Analysis. *J. Am. Stat. Assoc.* **1952**, *47*, 583–621. <https://doi.org/10.1080/01621459.1952.10483441>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.