Food Bioscience 59 (2024) 104001

Contents lists available at ScienceDirect

Food Bioscience

journal homepage: www.elsevier.com/locate/fbio

Comparative genomics reveals the potential biotechnological applications of *Liquorilactobacillus nagelii* VUCC-R001, a strain isolated from kombucha tea

Ilaria Larini [a], Sarah Tintori [a], Veronica Gatto [a], Giovanna E. Felis [a,b], Elisa Salvetti [a,b,*], Sandra Torriani [a]

[a] *Department of Biotechnology, University of Verona, Strada Le Grazie 15, 37134 Verona, Italy*
[b] *Verona University Culture Collection – Dept. of Biotechnology (VUCC-DBT), Strada Le Grazie 15, 37134 Verona, Italy*

ARTICLE INFO

Keywords:
Liquorilactobacillus nagelii, comparative genomics
Core phylogenomics
Safety assessment
D-phenyllactic acid
Destransucrase

ABSTRACT

Liquorilactobacillus nagelii is a lactic acid bacterium frequently found in a variety of traditional fermented foods, where it contributes to their sensory properties and potential health benefits. However, research evaluating the genetic and functional features of *L. nagelii* is scarce in the literature. In this study, we sequenced and assembled the genome of *L. nagelii* VUCC-R001, a strain isolated from kombucha tea, assessing its safety and exploring its biotechnological potential, mainly in terms of D-phenyllactic acid and dextran production, through a comparative genomic approach with 35 *Liquorilactobacillus* genomes and related phenotypic validation. Bioinformatic analysis revealed a good-quality draft genome (~2.4 Mb) of VUCC-R001 with a completeness around 99.7% (N50 of 151,630 bp). Comparative genomic analyses showed the correct identification of the new strain, the absence of genes encoding transmissible antibiotic resistance, virulence factors, and biogenic amine production, underlining its safety, also confirmed by phenotypic tests. We identified genes putatively associated with D-phenyllactic acid (PLA) production and verified the capability of this strain to produce a high concentration (52 mg/L) of PLA *in vitro*. To date, this is the first study reporting a *Liquorilactibacillus* strain that produces D-phenyllactic acid. Genome analyses of *L. nagelii* also elicited the presence of a dextransucrase GH70 (EC 2.4.1.5), leading to the production of dextran from sucrose, an exopolysaccharide with applications in the food and biomedical industries. This investigation provides new insights into the genomic features and functional attributes of *L. nagelii,* opening new prospects for the biotechnological use of selected strains belonging to this species.

1. Introduction

Fermented foods represent promising solutions for the development of safer, more secure, and sustainable food systems; furthermore, the potential health-promoting effects of fermentation-associated beneficial compounds are being investigated (Ganzle et al., 2023; Mukherjee et al., 2023). Fermentation-associated microorganisms mainly belong to the functional groups of lactic acid bacteria (LAB), acetic acid bacteria (AAB), and yeasts. As regards LAB, despite the already known information on conventional species used for many years in the food industry (mainly belonging to the genera *Lactobacillus*, *Lactiplantibacillus*, *Lacticaseibacillus, Lactococcus* and *Leuconostoc*), strains of new underexplored species from traditional fermented foods and beverages are gaining attention thanks to their *reservoir* of enzymatic activities and production of bioactive peptides, amino acids, exopolysaccharides, and compounds with prebiotic potential (Mukherjee et al., 2023). Furthermore, non-conventional LAB strains connected to fermentation are of industrial interest as they can be employed for various biotechnological applications, such as functional starter cultures, probiotics, adjuncts, or components of simplified tailor-made microbial consortia to improve sensory quality, safety, and health-related attributes of fermented foods (Gaspar and Crespo, 2016, chap. 7; Savary et al., 2021).

Among these species, *Liquorilactobacillus nagelii*, formerly known as *Lactobacillus nagelii* (Zheng et al., 2020), holds an interesting place. First identified by Edwards and colleagues (Edwards et al., 2000), *L. nagelii* was originally isolated from a commercial grape wine undergoing sluggish alcoholic fermentation. Besides, it was also found in many other plant-based fermented foods and beverages, such as water kefir (Laureys and De Vuyst, 2014), coyol wine or tavern (the fermented sap of

* Corresponding author.Department of Biotechnology, University of Verona, Strada Le Grazie 15, 37134 Verona, Italy
E-mail address: elisa.salvetti@univr.it (E. Salvetti).

https://doi.org/10.1016/j.fbio.2024.104001
Received 6 February 2024; Received in revised form 26 March 2024; Accepted 29 March 2024
Available online 4 April 2024

Abbreviations

AAB	Acetic Acid Bacteria
ANI	Average Nucleotide Identity
ARG	Antimicrobial Resistance Gene
BLAST (p, n)	Basic Local Alignment Search Tool (protein, nucleotide)
BUSCO	Benchmarking Universal Single-Copy Orthologs
CARD	Comprehensive Antibiotic Resistance Database
CAZy	carbohydrate-active enzyme database
CFS	cell-free supernatant
dbCAN3	automated carbohydrate-active enzyme annotation database
dDDH	digital DNA-DNA Hybridization
Ddl	D-alanine-D-alanine ligase
EFSA	European Food Safety Authority
EPS	exopolysaccharides
GGDC	Genome to Genome Distance Calculator
HPLC(-UV)	High Performance Liquid Chromatography (-ultraviolet)
IDF	International Dairy Federation
KEGG	Kyoto Encyclopedia of Genes and Genomes
LAB	Lactic Acid Bacteria
MIC	Minimum Inhibitory Concentration
MRS	De Man-Rogosa-Sharpe medium
PLA	D-phenyllactic acid
QPS	Qualified Presumption of Safety
QUAST:	Quality Assessment tool for genome assemblies
RAST	Rapid Annotation using Subsystem Technology
RGI	Resistance Gene Identifier tool
SCOBY	Symbiotic Culture of Bacteria and Yeasts
SRA	Sequence Read Archive
VFDB	Virulence Factor Database
VUCC-DBT	Verona University Culture Collection - Department of Biotechnology

Acrocomia mexicana) (Rivera et al., 2020), ensiled fruit residues (Yang et al., 2016), beer (Umegatani et al., 2022), kombucha tea (Ferremi Leali et al., 2022), paocai (a traditional Chinese fermented vegetable food (Zhao et al., 2023);), and Shalgam (a traditional Turkish lactic acid fermented beverage (Yetiman and Ortakci, 2023)). Whilst considered to be a spoilage organism in wine and beer, in other well-known fermented beverages produced with natural complex microbial consortia, such as water kefir and kombucha, *L. nagelii* can positively interact with co-occurring yeasts and AAB, affecting the production of metabolites associated with potential health benefits, as well as sensory properties (Yetiman and Ortakci, 2023; Zhao et al., 2023). Among these metabolites, the production of dextran from sucrose *in situ* during fermentation is very promising: *Liquorilactobacillus* spp. frequently express extracellular glucansucrases, especially dextransucrases, which produce α−1 → 6 linked glucose polymers from sucrose (Ganzle et al., 2023) allowing these strains the ability to persist in sucrose-rich environments; furthermore, the extracellular production of dextran contributes to the physicochemical and nutritional properties of fermented beverages (Bechtner et al., 2019).

Additional benefits of *L. nagelii* strains can be linked to the production of antimicrobial compounds that contribute to improving food safety and shelf-life, such as the release of organic acids and bacteriocins, as observed by (Yetiman and Ortakci, 2023). Among organic acids, D-phenyllactic acid (PLA) is a natural microbial antagonist and it is frequently found in foods fermented by LAB (*e.g.* kimchi (Jung et al., 2019);). In addition to its application as a biological preservative, further investigation is warranted for this metabolite in non-conventional LAB also as a "fermentation-dependent" biomarker for lactic-fermented foods, indicating that it can act as a promising measure of fermented food intake and metabolism (Li et al., 2021).

In addition to the promising features described, the first step in selecting bacteria for the fermentation of new food is their safety assessment which is commonly based on taxonomic identification (usually through whole-genome sequencing), the absence of safety-related traits (*i.e.*, antibiotic resistance, virulence factors, and biogenic amines) and a consistent body of knowledge related to its lifestyle, ecology and history of safe use (Ganzle et al., 2023). From the regulatory perspective, *L. nagelii* is included in the International Dairy Federation (IDF) inventory of food cultures with proven beneficial association with a food matrix (IDF, 2022), but, to date, this species does not hold the Qualified Presumption of Safety (QPS) status by the European Food Safety Authority (EFSA), which is a prerequisite for its deliberate use in the food chain.

At the molecular level, the complete genome sequences of four *L. nagelii* strains isolated from various sources were recently subjected to a comparative genomic analysis with other 20 strains belonging to the other species of the genus *Liquorilactobacillus*, and their genetic relationships and evolutionary patterns were explored, identifying two clades that differed in specific genomic signatures (presence/absence of protophage infection and different gene selection pressure) (Liu et al., 2023).

Since research on the evaluation of safety-related and biotechnological attributes of *L. nagelii* remain scarce and scattered in the literature, in this study we carried out a thorough safety assessment and functional investigation of *L. nagelii* VUCC-R001, a strain previously isolated from an artisanal kombucha tea (Ferremi Leali et al., 2022), through genome sequencing and in-depth comparative genomic analysis with 35 other genomes of *Liquorilactobacillus* species.

To date, this is the first genome sequence of a *L. nagelii* strain isolated from kombucha; this study provides new insights on the genomic features and functional attributes in terms of dextran and PLA production of this non-conventional species through a genotype-phenotype correlation approach, opening a new perspective in the applicability of strains of this species in the fermented beverage industry.

2. Materials and methods

2.1. Bacterial strain and growth conditions

Liquorilactobacillus nagelii TLV-4R7, previously isolated from an artisanal kombucha tea (Ferremi Leali et al., 2022), was routinely grown in MRS broth medium (Oxoid, Milan, Italy) at 30 °C for 24 h in aerobic conditions. The strain is deposited in the Verona University Culture Collection - Department of Biotechnology (VUCC-DBT) and named VUCC-R001.

2.2. DNA extraction, whole genome sequencing, assembly, and annotation

Total genomic DNA of *L. nagelii* VUCC-R001 was extracted from 1.8 mL of broth culture using the *Wizard SV Genomic DNA Purification System kit* (Promega Italia, Milan, Italy), following the manufacturer's instructions with some modifications. Briefly, cells were harvested by centrifugation (986 G-force/10 min), washed with distilled water, and resuspended in 10 mg/mL of lysozyme solution (Merck, Darmstadt, Germany) at 37 °C for 60 min; then the sample was centrifuged, and the DNA extraction was carried out. The DNA quality was estimated by Qubit Fluorometer (Thermo Fischer Scientific, Waltham, MA, USA) and

NanoDrop One/OneC Microvolume UV–Vis Spectrophotometer (Thermo Fischer Scientific, Waltham, MA, US).

Library preparation and sequencing were performed by IGATech (Udine, Italy). In detail, libraries were obtained using CeleroTM DNA-Seq kit (NuGEN, San Carlos, CA, USA) according to the manufacturer's instructions; Illumina NovaSeq 6000 in paired-end 150 mode was then used as platform sequencing. The raw reads were demultiplexed and adapters were masked by IGATech tools.

The raw reads were trimmed with fastp (version 0.21.0; (Chen et al., 2018)) and *de novo* assembly was performed with SPAdes using default options (version 3.13.0; (Bankevich et al., 2012)). The quality of the genome assembly was assessed with the software QUAST (Quality Assessment Tool for Genome Assemblies) (version 5.0.2; (Gurevich et al., 2013)) and BUSCO (Benchmarking Universal Single-Copy Orthologs) (version 4.1.4; (Manni et al., 2021)) using 'lactobacillales_odb10' as lineage dataset. Genome annotation was performed with Prokka (version 1.14.6; (Seemann, 2014)). Proksee (https://proksee.ca/; (Grant et al., 2023)) was used to generate a circular graphical genome map of *L. nagelii* VUCC-R001 and RAST (Rapid Annotation using Subsystem Technology) (http://rast.nmpdr.org/rast.cgi) was used to obtain the subsystem category distribution for annotation (Overbeek et al., 2014).

2.3. Data availability

The sequencing reads of *L. nagelii* VUCC-R001 genome have been deposited at the NCBI Sequence Read Archive (SRA) under the BioProject ID PRJNA1044759 and BioSample ID SAMN38411806 with the accession number SRR26941028, whereas the corresponding assembly data set has been deposited at DDBJ/EMBL-ENA/GenBank under the accession number JAYMDD000000000. The version described in this paper is JAYMDD000000000.1.

2.4. Taxonomic identification and phylogenomic analysis

The whole-genome sequences of 35 *Liquorilactobacillus* strains were downloaded from the NCBI's RefSeq genome database on September 25, 2023 (Supplementary File S1), and their quality was assessed with BUSCO. Genomes with a completeness higher than 96% were selected and reannotated with Prokka. The Average Nucleotide Identity (ANI (Goris et al., 2007);), was calculated using pyani tool (https://github.com/widdowquinn/pyani), while digital DNA-DNA hybridization (dDDH (Auch et al., 2010);) was determined using GGDC (Genome to Genome Distance Calculator 3.0 (Meier-Kolthoff et al., 2022);).

Comparative genomic analysis was performed using Roary pipeline (Page et al., 2015) setting 70% as blast similarity and 99% as number of strains. The presence and absence genes table outputted from Roary pipeline was analyzed with the pg_power_law_fit function in R package pagoo (Ferres and Iraola, 2021) for the analysis of pangenome and a python script (https://github.com/MrTomRod/flower-plot) was used to produce the flower plot. After this, a BlastKOALA (Kanehisa et al., 2016) analysis was performed on the core-genome to obtain KEGG (Kyoto Encyclopedia of Genes and Genomes) orthologs and annotation. A phylogenomic tree of 130 concatenated core-genes shared with the outgroup *Companilactobacillus alimentarius* DSM 20249T was built with MEGA11 (version 11 (Tamura et al., 2021);) using Neighbor-joining as statistical method and 1000 as bootstrap value. The phylogenomic tree was visualized with ggtree R package (Xu et al., 2022).

2.5. Safety assessment

2.5.1. Genomic and phenotypic analysis of antibiotic resistance

The annotated sequences of the whole dataset were employed to query the Comprehensive Antibiotic Resistance Database (CARD, version 3.2.7; https://card.mcmaster.ca/) through the Resistance Gene Identifier tool (RGI, version 6.0.2 (Alcock et al., 2023);) selecting only "Perfect" and "Strict" hits and to query ResFinder 4.1 database (https://cge.food.dtu.dk/services/ResFinder/ (Florensa et al., 2022);) with 90% as percentage of identity and 60% as query coverage for acquired antimicrobial resistance genes.

Moreover, AAB18338.1 coding for D-alanine–D-alanine ligase (Ddl) of *Leuconostoc mesenteroides* ATCC 8293^T was used as query to perform a BLASTp search against all genomes of *Liquorilactobacillus* spp. All the amino acid sequences of Ddl retrieved in the annotated genomes were aligned with Clustal Omega online tool (https://www.ebi.ac.uk/Tools/msa/clustalo/ (Sievers et al., 2011);) and fully gap removed with Jalview (version 2.11.2.7 (Waterhouse et al., 2009),) to search the point mutation F261Y involved in vancomycin resistance (Campedelli et al., 2019).

The minimum inhibitory concentrations (MICs) were evaluated in VUCC-R001, as well as in *L. nagelii* DSM 13675^T in duplicate for nine antibiotics indicated by EFSA (ampicillin, clindamycin, chloramphenicol, erythromycin, gentamicin, kanamycin, streptomycin, tetracycline, and vancomycin) (Merck), following both homofermentative and facultative heterofermentative lactobacilli entries (EFSA, 2018). Briefly, growth was monitored using the Cary 60 UV-VIS Spectophotometer instrument (Agilent Technologies, Santa Clara, USA) at a wavelength of 625 nm, until an OD_{625} = 0,160–0,200 was reached, which corresponded to approximately 1×10^8 cells/mL. Subsequently, the culture was diluted 1:10 in LSM broth (90% ISO-SENSITEST Broth, Oxoid, and 10% MRS, Oxoid), then 100 μL were inoculated into tubes containing 9.9 mL LSM broth and three different concentrations of each antibiotic (half cut-off, cut-off, and two-fold cut-off) to reach a final concentration of 1×10^5 cells/mL. Tubes were incubated under aerobic conditions at 30 °C for 48 h, with a positive control (no antibiotics) and a negative one (no inoculation). MICs were read as the lowest concentration of an antimicrobial agent at which visible growth was inhibited.

2.5.2. Histamine and tyramine production

The presence of the genes encoding for histidine and tyrosine decarboxylases was investigated in the *Liquorilactobacillus* spp. genomes. BLASTn was performed with a threshold of nucleotide sequence identity >30%, e-value $<1 \times 10^{-5}$ and query coverage >70% using the following sequences (GenBank ID) as queries: AF446085.5 (*Levilactobacillus brevis* IOEB 9809); AF354231.1 (*Enterococcus faecalis* JH2-2); AB125629.1 (*Tetragenococcus muriaticus*); U58865.1 (*Oenococcus oeni* 9204); J02613.1 (*Ligilactobacillus saerimneri* 30A); AJ749838.1 (*Lentilactobacillus buchneri* B301); AY651779.1 (*Lentilactobacillus hilgardii* IOEB 0006).

To phenotypically confirm the absence of production of the two biogenic amines, *L. nagelii* VUCC-R001 and DSM13675^T were grown in the JNM medium (Merck) modified according to (Bover-Cid and Holzapfel, 1999), added with tyrosine or histidine in duplicate. Decarboxylation of amino acids on this medium is detected by the pH indicator color changing from yellow to purple, due to the pH increase for the amine production. *Ligilactobacillus saerimneri* 30a and *Enterococcus faecalis* EF37 were used as positive controls for histamine and tyramine production, respectively.

2.5.3. Virulence and pathogenicity factors analysis

VFDB (Virulence Factor DataBase; (Liu et al., 2022) was employed for virulence factor analysis. The search was focused on the enzymes and toxins of *Enterococcus* (the phylogenetically closest LAB) reported in VFDB. The protein sequences (GenBank ID) analyzed were gelatinase (WP_002369251); hyaluronidase (WP_002399773; WP_011109578); SprE (WP_002369252) and cytolysin (AAM75253; AAM75252; AAM75254; AAM75249; AAM75250; AAM75251; AAM75248; AAM75247). BLASTp was performed with a threshold of amino acid sequence identity >30%, e-value $< 1 \times 10^{-5}$ and query coverage >70%.

2.6. Identification of dextransucrase production

dbCAN3 (https://bcb.unl.edu/dbCAN2/index.php (Zheng et al., 2023);), a webserver for the automatic annotation of active carbohydrate enzymes which uses CAZy as a database (Carbohydrate-active enzymes database, http://www.cazy.org/ (Drula et al., 2022);) was employed to search for genes coding for dextransucrase in *L. nagelii* VUCC-R001. The amino acid sequence of the identified dextransucrase of VUCC-R001 was then aligned to the dextransucrase gene of *L. nagelii* TMW 1.1827 (GenBank ID: BSQ50_03510) and used as a query to search the *Liquorilactobacillus* genomes via BLASTp (identity >30%, e-value < 1×10^{-5} and query coverage >70%.) The retrieved sequences were aligned with Clustal Omega and edited with Jalview. Finally, MEGA11 (Molecular Evolutionary Genetics Analysis Version 11 (Tamura et al., 2021);) was also used to create a gene tree using the Maximum Likelihood statistical method.

To evaluate the capability to produce exopolysaccharides (EPS) by *L. nagelii* VUCC-R001 and DSM 13675^T, 10 μL of broth culture were spotted on MRS agar medium (Oxoid), supplemented with different sugars (glucose, sucrose, fructose, galactose, or lactose, 20 g/L) and with ruthenium red (0.08 g/L). EPS production was assessed by visual inspection of the phenotypic characteristics of the colonies (white/mucoid: EPS production; pink: no EPS production) (Kersani et al., 2017).

2.7. D-phenyllactic acid (PLA) production

2.7.1. Identification of PLA biosynthesis pathway genes

The protein accession numbers were retrieved using E.C. numbers reported in (Rajanikar et al., 2021). In details, a total of 12 query sequences were used: three queries were related to the core pathway for PLA production (E.C. 2.6.1.-AraT/aromatic amino acid aminotransferase; E.C. 1.1.1.28-D-LDH/D-lactate dehydrogenase; E.C. 1.1.1.27-L-LDH/L-lactate dehydrogenase); nine queries were associated to the *de novo* synthetic pathway (E.C. 2.5.1.54-AroA/3-deoxy-7-phosphoheptulonate synthase; E.C. 4.2.3.4-AroB/3-dehydroquinate synthase; E.C. 4.2.1.10- AroC/3-dehydroquinate dehydratase; E.C. 1.1.1.25-AroD/shikimate 5-dehydrogenase; E.C. 2.7.1.71-,AroI/shikimate kinase; E.C. 2.5.1.19-AroE/3-phosphoshikimate-1-carboxyvinyltransferase; E.C. 4.2.3.5- AroF/chorismate synthase; E.C. 5.4.99.5-AroH/chorismate mutase II; E.C. 4.2.1.51- PheA/prephenate dehydratase) Query sequences were employed to detect through BLASTp the PLA biosynthesis pathway gene sequences in *L. nagelii* VUCC-R001, considering sequence identity >30%, e-value < 1×10^{-5} and query coverage >70%. Moreover, an additional BLASTp was performed using the PLA biosynthesis pathway proteins retrieved for VUCC-R001 on the predicted proteins of the *Liquorilactobacillus* genomes.

2.7.2. HPLC determination of PLA

PLA production by *L. nagelii* VUCC-R001 and DSM 13675^T was determined by HPLC-UV analysis. In detail, a broth culture of each strain was centrifuged, and the supernatant was filtered through a 0.22 μm filter (Millipore, Milano, Italy) to obtain the cell-free supernatant (CFS). The CFS was directly injected in the HPLC instrument (Jasco Extrema LC-4000 - photo diode array detector Jasco MD-4010, Mary's Court Easton, USA) that was equipped with the ZORBAX Eclipse Plus C18 column (95 Å, 4.6 mm × 250 mm, 5 μm particles) (Agilent, Santa Clara, USA). The run was performed at 35 °C and 210 nm. Solvent A (water - 0.05% trifluoroacetic acid -TFA), and solvent B, (methanol - 0.05% TFA) were used with the following gradient elution: 10% B - 90% A at 5 min; 40% B - 60% A at 25 min; 100% B at 37 min and 10% B - 90% A B at 42 min (Valerio et al., 2016; Xu et al., 2020). The run lasted 50 min for each sample, with a flow rate of 1.00 mL/min at 40 MPa maximum pressure and injecting 10 μL of sample. Furthermore, different solutions of PLA (Sigma-Aldrich) in water were prepared to create the calibration curve: 3; 6; 15; 30; 60; 150; 300 g/L. The Jasco ChromNAV program (version 2.04.04) was used to process the chromatograms obtained from the HPLC-UV analysis. The amount of PLA produced by the two strains was determined by integrating calibration curves obtained from standards (Supplementary File S2). *Lacticaseibacillus casei* NRRL B-1922^T was used as positive control (Cortes-Zavaleta et al., 2014); the test was done in duplicate.

3. Results

3.1. General features of Liquorilactobacillus nagelii VUCC-R001 genome

A total of more than 8 million paired-end reads were *de-novo* assembled. A draft genome consisting of 37 scaffolds and a size of 2.4 Mb (scaffolds size ≥1000 bp) (Fig. 1a) with a high sequencing depth of 498 × (avg. coverage depth) and a coverage (≥1 ×) of 99.96% was obtained, highlighting a good quantity of sequencing. The GC content of *L. nagelii* VUCC-R001 was 36.68%, whereas the N50 value was 151,630 bp by QUAST statistics (Supplementary File S3). Based on the BUSCO tool, the genome resulted complete and in single copy at 99.7% (Supplementary File S4). A total of 2381 genes were predicted, showing a high level of completeness (99.7%) with BUSCO evaluation. These results confirmed that the assembly and annotation obtained are of high-quality, in accordance with the other *L. nagelii* genomes deposited in NCBI.

For the annotation performed by Prokka and RAST, 2531 coding sequences were predicted and most of the genes were related to Carbohydrates (326), followed by Amino Acids and Derivatives genes (239), Protein metabolism (171), Cell Wall and Capsule genes (124), Cofactors, Vitamins, Prosthetic Groups, Pigments (99), RNA Metabolism (99) and DNA Metabolism (98), as reported in Fig. 1b.

3.2. Core-genome and pan-genome comparison

The genome size of *Liquorilactobacillus* spp. ranged from 1.92 to 2.72 Mbp, with *L. satsumensis* TMW11829 showing the largest genome size (2.72 Mbp), and *L. uvarum* DSM 19971^T the highest number of genes (2,667); on the other hand, *L. cacaonum* DSM 21116^T showed both the minimum genome size (1.92 Mbp) and the minimum number of genes (1,885). The pan-genome of the 36 *Liquorilactobacillus* strains enclosed 14,593 genes, of which only 499 constitute the core genome (genes shared between 99% and 100% of strains), 59 are considered soft-core genes (genes shared between 95% and 99% of strains), 4477 are shell genes (genes shared between 15% and 95% of strains) and 9558 are cloud genes (genes shared between 0% and 15% of strains), as reported in Fig. 2a.

The pangenome of the genus *Liquorilactobacillus* was considered open, because it fit the Heap's law ($n = \kappa N^{\gamma}$) (Tettelin et al., 2008), where the value of γ was calculated to be 0.52 (<1.0) for the 36 genomes analyzed (Ferres and Iraola, 2021). Moreover, the number of accessory and unique genes for each strain varied between genomes as displayed in the flower plot in Fig. 2b ranging from zero (*L. mali* UCMA16447) to 2020 (*L. uvarum* DSM 19971^T) for accessory genes and spanning from 2 (*L. satsumensis* Kef-w11) to 509 (*L. ghanensis* DSM 18630^T) for unique genes. The genome of VUCC-R001 has 1817 accessory genes and 14 unique genes of which 10 are annotated as hypothetical proteins and the other coming from transposase families, transcriptional repressor and reductase.

About KEGG annotation of core-genes, the majority are related to genetic information processing (178) (in particular to ribosome related proteins and aminoacyl-tRNA biosynthesis), followed by carbohydrate metabolism (38) (*e.g.*, glycolysis/gluconeogenesis and pentose phosphate pathways), nucleotide metabolism (37), amino acids metabolisms (27) (especially to alanine, aspartate and glutamate metabolism), environmental information processing (24) (such as ABC transporters genes), signaling and cellular processes (19) (*e.g.*, cell cycle and flagellar assembly related genes), energy (17) and lipid metabolism (14) (*e.g.*, oxidative phosphorylation and fatty acid biosynthesis) and glycan

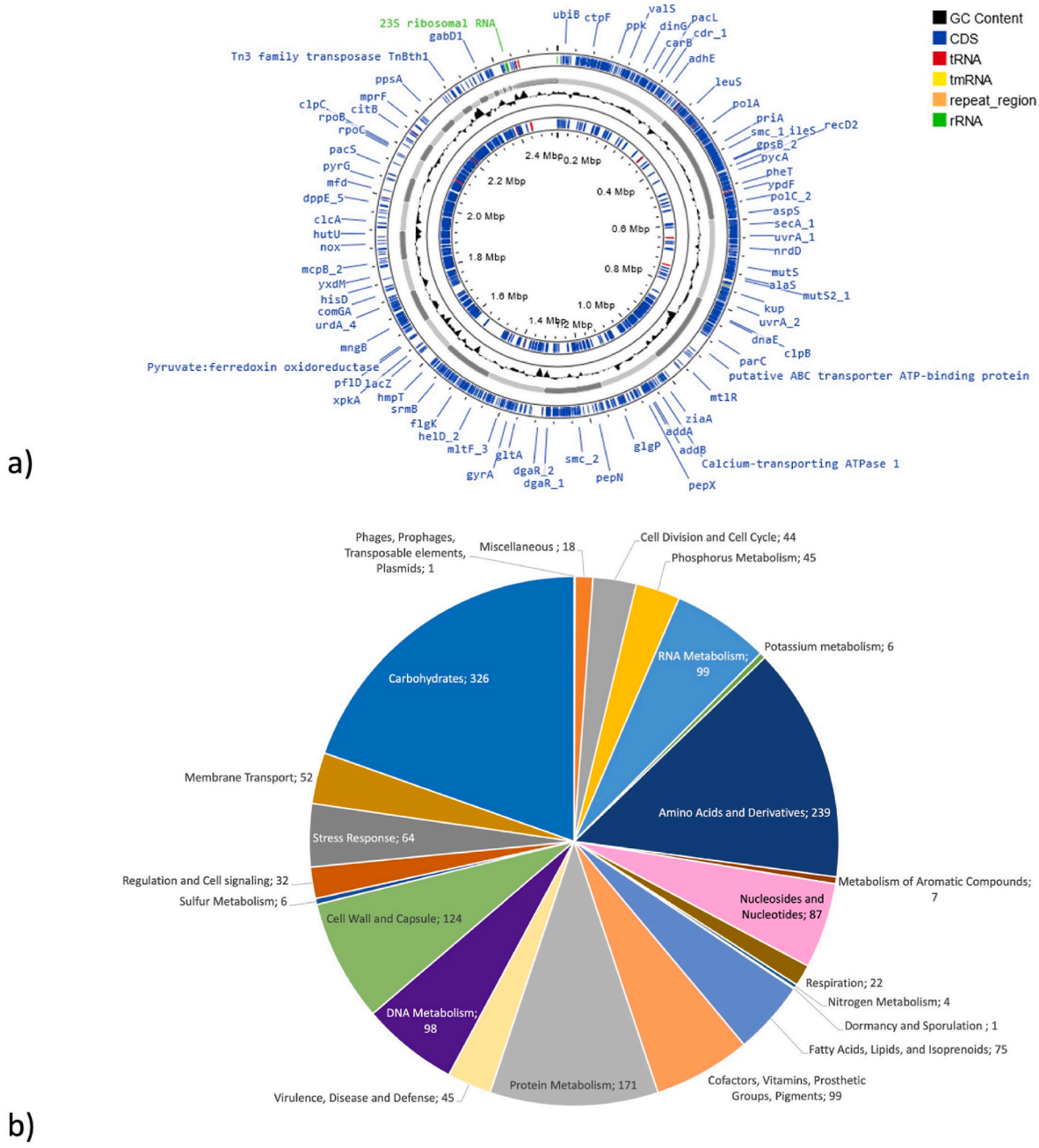

Fig. 1. Graphical representation of the genome and RAST (Rapid Annotation using Subsystem Technology) annotation of *Liquorilactobacillus nagelii* VUCC-R001. For panel (a) in the genome representation executed by Proksee, the coding regions were shown in blue, tRNA in red, rRNA in green, tmRNA (transfer-messenger RNA) in yellow, repeat regions in orange and the GC content in black. For panel (b) all the categories of RAST annotation were displayed in a pie chart with the numerical value for each. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

biosynthesis (10) (in particular for peptidoglycan biosynthesis), unlike 111 genes that remained unclassified as shown in Fig. 2c.

A phylogenomic analysis was performed, selecting and concatenating 130 core-genes shared with the outgroup *Companilactobacillus alimentarius* DSM 20249T which showed that, as expected, VUCC-R001 clustered with the other *L. nagelii* strains. Generally, two clades could be recognised: one clade including *L. nagelii* and it most closely related species (*L. ghanensis*, *L. sicerae* and *L. vini)*, while all the other *Liquorilactobacillus* species formed the other clade. Within the latter, three further subclusters emerged: the first one included the genomes of *L. satsumensis* and *L. oeni*, the second one enclosed the genomes of *L. cacaonum*, *L. hordei* and *L. mali*, while the last one grouped the genomes of *L. capillatus*, *L. sucicola*, *L. uvarum* and *L. aquaticus* strains (Fig. 3).

3.3. Taxonomic identification

Digital DNA-DNA hybridization (dDDH) and the Average Nucleotide Identity (ANI) were calculated to precisely determine the species assignation of VUCC-R001, previously defined by 16S rRNA gene sequencing and BLAST comparison (Ferremi Leali et al., 2022).

VUCC-R001 showed the highest dDDH and ANI values (89.40% and 98.80% respectively) with the genome of *L. nagelii* DSM 13675^T, as reported in Fig. 4 and Supplementary File S5. Since values obtained were higher than the proposed species cut-off boundary (70% and 95%, respectively) (Chun et al., 2018), the whole genomic data confirmed that

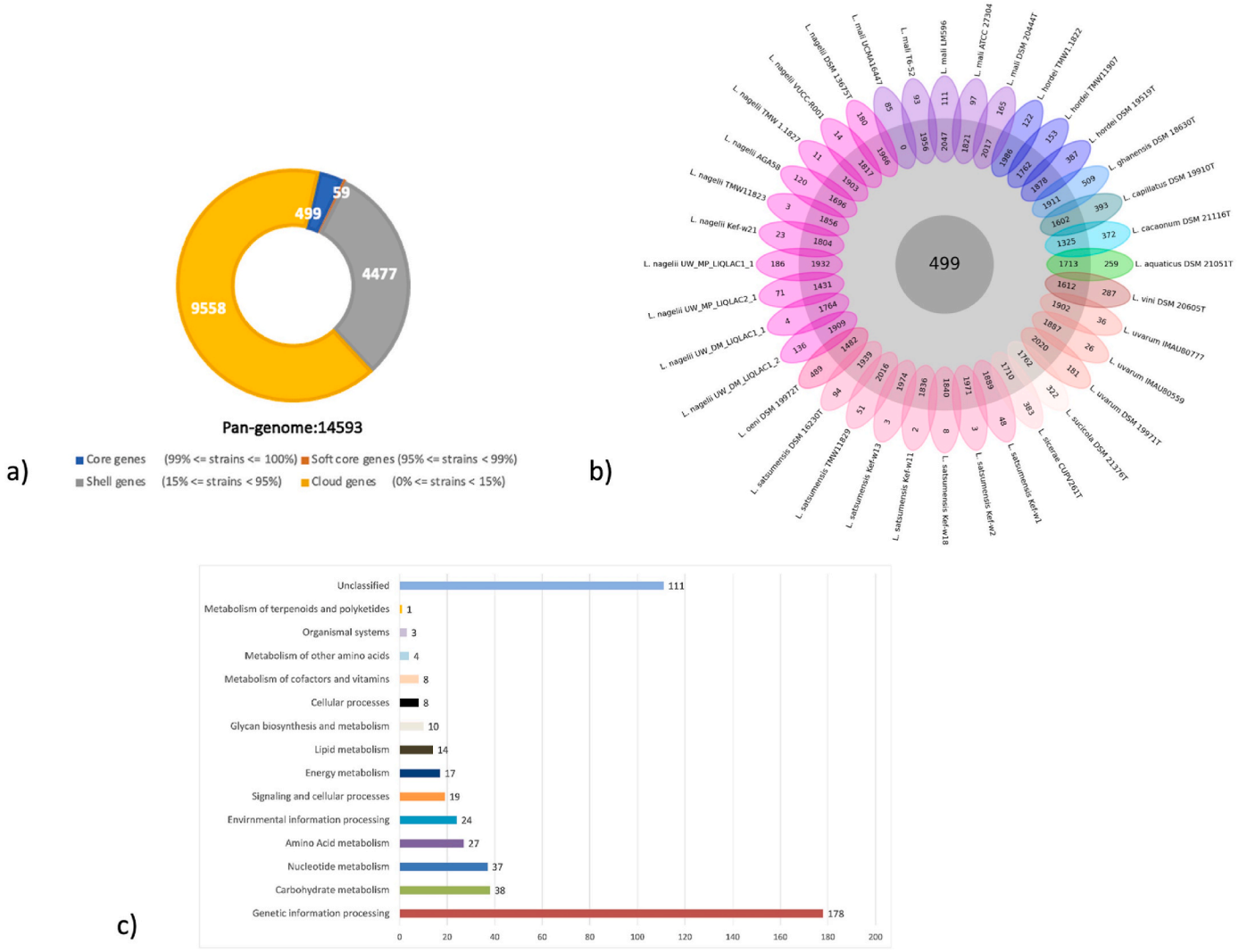

Fig. 2. Pan-genome and core-genome analysis of 36 *Liquorilactobacillus* strains (VUCC-R001 included) performed by Roary pipeline. (a) The ring-chart represents the distribution of core and accessory genes (split up in soft-core, shell and cloud genes) and (b) the flower plot depicts in the middle the number of core-genes and in the petals the number of accessory (closest to the middle), and unique genes for each genome; (c) BlastKOALA analysis performed on 499 core genes to obtain KEGG annotation.

VUCC-R001 indeed belongs to the species *L. nagelii,* also supported by the lower dDDH and ANI values obtained with the other species of the genus *Liquorilactobacillus* (Fig. 4, Supplementary File S5).

3.4. Safety assessment

3.4.1. Analysis of antibiotic resistance

To verify if the strain VUCC-R001 had any resistance towards antibiotics, both *in silico* and phenotypic analysis were carried out. The *in-silico* search of antibiotic resistance genes was performed using the databases CARD (Alcock et al., 2023) and ResFinder 4.0 (Florensa et al., 2022), as indicated in the EFSA guidance (2018).

ResFinder returned no hits either in VUCC-R001 or in the analyzed genomes of the genus *Liquorilactobacillus*, while CARD's RGI algorithm found antibiotic-resistance genes in all genomes displayed in red (Fig. 5). In particular, the gene *van*T of the *van*G cluster (accession no. ARO: 3002972, *Strict* cut-off) was found in VUCC-R001 (32.88% identity and 52.25% as query coverage), as well as in all genomes included in the dataset. This gene is involved in vancomycin resistance by altering the target of the antibiotic (Courvalin, 2006).

Furthermore, other *van* genes were detected with *Strict* cut-offs in other *Liquorilactobacillus* genomes: *van*H of the *van*A cluster (accession no. ARO: 3002942) in *L. hordei* DSM 19519^T, TMW11907 and TMW 1.1822; *van*H of the *van*B cluster (accession no. ARO: 3002943) in *L. ghanensis* DSM 18630^T, and in *L. nagelii* AGA58, UW_MP_LIQLAC1_1, UW_DM_LIQLAC1_1, and UW_DM_LIQLAC1_2; *van*H of the *van*P cluster (accession no. ARO: 3007188) in *L. oeni* DSM 19972^T; *van*H of the *van*O cluster (accession no. ARO: 3002948) in *L. uvarum* DSM 19971^T, IMAU80559, and IMAU80777.

As for vancomycin resistance, lactobacilli may show the Ddl ligase (D-alanine-D-alanine ligase) of type Y or of type F, which can either confer or not intrinsic resistance to vancomycin. For this reason, all the *Liquorilactobacillus* Ddls were aligned (including VUCC-R001, locus tag: SM360_00625) and showed the F-type (Supplementary File S6). These data indicated that all of them show intrinsic resistance to vancomycin (*i.e.,* it is less likely to be transferred to other microorganisms through horizontal gene transfer events) (Ammor et al., 2007).

Thanks to CARD, two other genes coding for efflux pumps associated with resistance to antibiotics, disinfectants and antiseptics were identified: the gene *mde*A (multidrug efflux A; accession no. ARO: 3007011) was found in *L. capillatus* DSM 19910^T and *L. hordei* TMW 1.1822; while the gene *qac*J (accession no. ARO: 3007014) was found in *L. hordei* TMW 1.1822 and TMW11907. *Mde*A confers resistance to fluoroquinolones, aminoglycoside, penicillins, tetracyclines, disinfectant and antiseptic

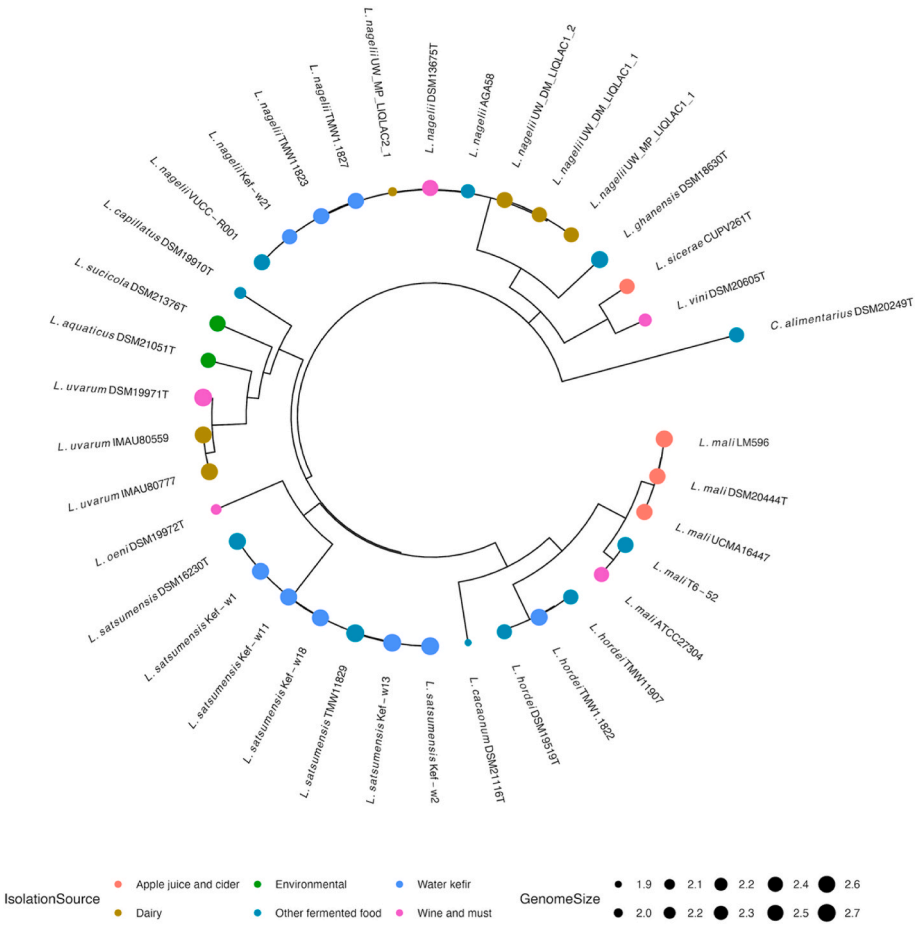

Fig. 3. The core phylogenomic tree of 36 *Liquorilactobacillus* strains inferred using the Neighbor-joining method and a bootstrap value of 1000. The tree was built based on the 130 core-genes shared with the outgroup *Companilactobacillus alimentarius* DSM 20249^T. The isolation sources are displayed in different colors (red: apple juice and cider, yellow: dairy products, light green: environmental isolates, light blue: other fermented foods, dark blue: water kefir, pink: wine and must) and the genome size is reported as the size of the node for each strain. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

agents (Kim et al., 2013), while *qacJ* is an efflux pump belonging to the Small Multidrug Resistance (SMR) family, that confers resistance to quaternary ammonium compounds (Bjorland et al., 2003).

The Minimum Inhibitory Concentration (MIC) analysis performed on VUCC-R001 and DSM 13675^T with the antibiotics indicated by EFSA (2018) revealed that the two *L. nagelii* strains were not able to grow beyond the cut-off values for eight antibiotics tested (ampicillin 7 mg/L; clindamycin 4 mg/L; chloramphenicol 4 mg/L; erythromycin 1 mg/L; gentamicin 16 mg/L, kanamycin 64 mg/L, streptomycin 64 mg/L, tetracycline 8 mg/L), but they are able to grow beyond vancomycin cut-off value (2 mg/L), confirming the resistance retrieved *in silico*.

3.4.2. Analysis of biogenic amine production

The genes encoding for tyrosine and histidine decarboxylases for tyramine and histamine production, respectively, were not detected in the *Liquorilactobacillus* genomes. These data were further confirmed at phenotypic level for VUCC-R001 and DSM 13675^T, where the two strains were not able to decarboxylate tyrosine nor histidine (Supplementary File S7).

3.4.3. Virulence and pathogenicity factors analysis

Analysis of the *Liquorilactobacillus* genomes with the Virulence Factor Database (VFDB) did not retrieve significant data, showing that these species do not harbor any known virulence and pathogenicity factors.

3.5. Identification of dextransucrase activity

The analysis with dbCAN3 of CAZy showed the presence of a protein belonging to the GH70 family (Glycoside Hydrolase), annotated in VUCC-R001 as a dextransucrase (EC. 2.4.1.5, NODE_1, position: 389-1188). This enzyme is responsible for dextran production, mainly from sucrose (Bechtner et al., 2019). The distribution of the amino acid sequence of the dextransucrase was also investigated in all the *Liquorilactobacillus* spp. genomes, which revealed that 27 out of 36 strains displayed this enzyme (Supplementary File S8).

L. cacaonum DSM 21116^T, *L. ghanensis* DSM 18630^T, *L. vini* DSM 20605^T, *L. sicerae* CECT 8227^T, and *L. oeni* DSM 19972^T did not display the gene encoding for dextransucrase. As for the species *L. nagelii* and *L. hordei*, an intraspecific diversity was highlighted: *L. nagelii* DSM 13675^T and *L. hordei* DSM 19519^T were the only strains among their species in which the dextransucrase was not detected. Interestingly the dextransucrase of VUCC-R001 formed a robust cluster with those of *L. nagelii* AGA58, TMW 1.1827, TMW11823 and Kef-21.

The presence of the gene encoding for a dextransucrase in VUCC-R001 was further validated phenotypically, as white/mucoid colonies (indicating exopolysaccharide production) were observed in presence of sucrose (Supplementary File S9). Conversely, the colonies of *L. nagelii* DSM 23675^T did not display the EPS production in the same conditions, confirming the prediction of the genomic data.

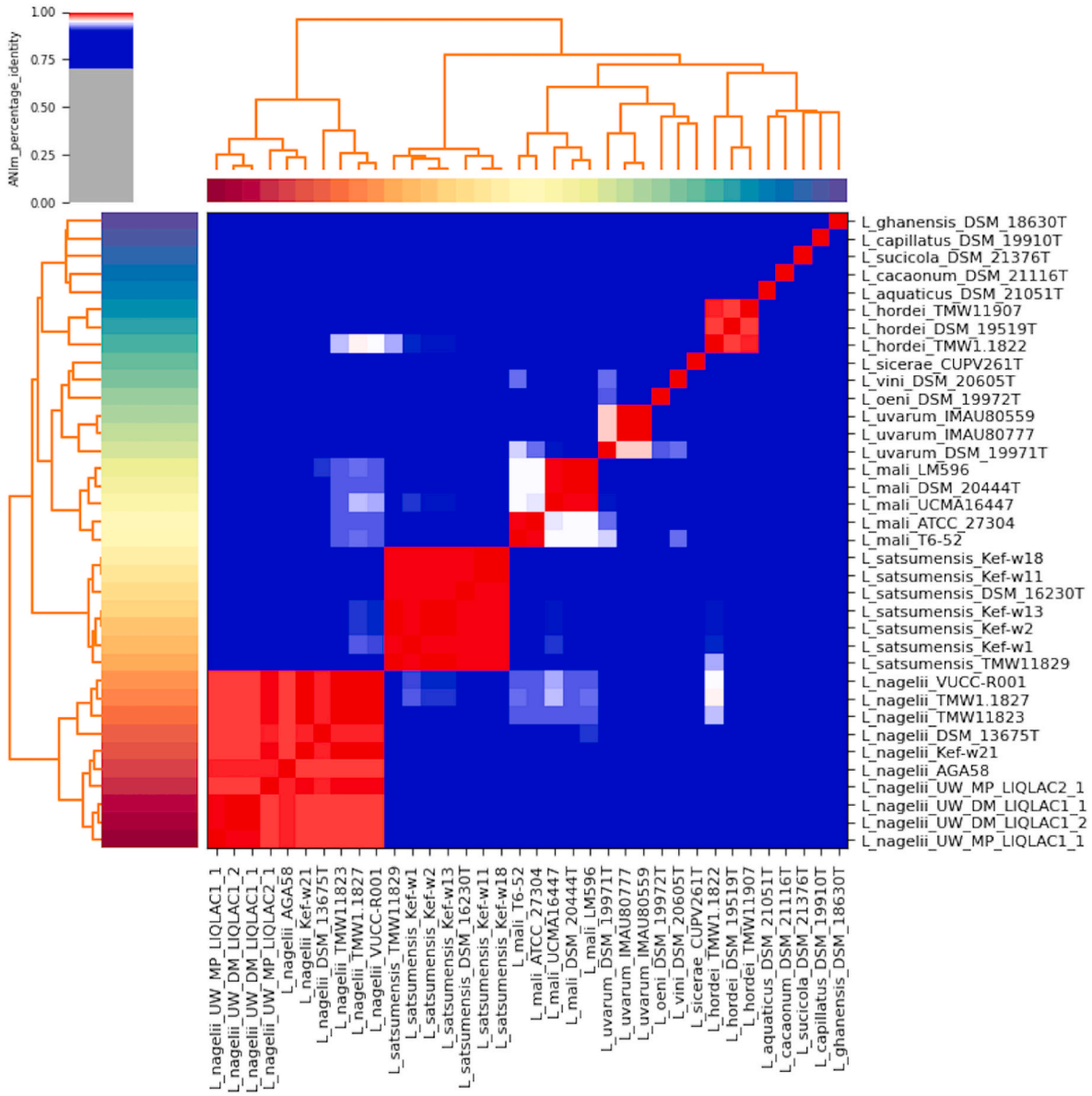

Fig. 4. Average Nucleotide Identity (ANI) calculation obtained using pyani tool v. (0.2.11) among the analyzed genomes of the genus *Liquorilactobacillus*. The values closest to 100% are in red, values closest to 90% are in white, values between 70 and 89% are in blue, and values in grey are lower than 70%. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.6. D-phenyllactic acid (PLA) production

All the genes related to the core pathway for PLA production were detected in the whole dataset (displayed in sugar-paper blue); conversely, the *de novo* biosynthesis pathway was retrieved in 21 out of 36 *Liquorilactobacillus* genomes (60%) except for prephenate dehydratase (PheA) which was not detected in any genomes (Fig. 6). At intraspecific level, all *L. hordei*, *L. satsumensis* and *L. uvarum* strains displayed the same pattern of genes, with *L. uvarum* and *L. satsumensis* characterized by a complete genetic background, while *L. hordei* strains missed five genes of the *de novo* biosynthesis pathway coding for 3-deoxy-7-phosphoheptulonate synthase, 3-dehydroquinate synthase, 3-dehydroquinate dehydratase, shikimate dehydrogenase and prephenate dehydratase (E.C. 2.5.1.54, E.C. 4.2.3.4, E.C. 4.2.1.10, E.C. 1.1.1.25 and E.C. 4.2.1.51). *L. nagelii* strains (including VUCC-R001; Supplementary file S10) displayed the PLA pathway apart from prephenate dehydratase (both core and *de novo* biosynthesis) except for AGA58 (missing E.C. 2.5.1.54, E.C. 4.2.3.4, E.C. 1.1.1.25 and E.C. 4.2.1.51) and UW_MP_LIQLAC2_1 (missing E.C. 2.5.1.54, EC 4.2.1.10, E.C. 1.1.1.25, and E.C. 4.2.1.51). As for L. *mali,* DSM 20444^T and LM596 did not display E.C. 2.5.1.54, E.C. 4.2.3.4 and E.C. 4.2.1.51, while the rest of the strains also missed E.C. 4.2.1.10 and E.C. 1.1.1.25. Interestingly, in *L. vini* DSM 20605^T genome, the gene coding for 3-phosphoshikimate 1-carboxyvinyltransferase (E.C. 2.5.1.19, belonging to the *de novo* biosynthesis pathway) was not detected while *L. oeni* DSM 19972^T was the only genome that displayed only the genes coding for the core pathway.

Given the presence of the complete PLA production pathways (both core and the *de novo* biosynthesis) except for prephenate dehydratase in *L. nagelli* VUCC-R001 and *L. nagelii* DSM 13675^T, HPLC analysis was conducted to assess the capability of these strains to produce PLA phenotypically. As reported in Supplementary Fig. S11, *L. nagelii* VUCC-R001 showed a PLA production up to 52.78 ± 0.42 mg/L after 24 h of growth in MRS medium, which was higher than those produced by *L. nagelii* DSM 13675^T (32.15 ± 2.09 mg/L) and by *Lacticaseibacillus casei* NRRL B-1922^T, used as reference (5.75 ± 0.64 mg/L)

Fig. 5. Heatmap with antimicrobial resistance genes (ARGs) obtained by RGI (Resistance Gene Identifier) tool in the dataset analyzed. The heatmap points out the presence (red) and absence (blue) of the antimicrobial resistance (AR) genes associated with the unrooted phylogenetic tree of the concatenated 499 core-genes. Moreover, in the tree the genome size as the size of each node was displayed and the isolation sources as the different color of the nodes were depicted. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

(Cortes-Zavaleta et al., 2014).

4. Discussion

In this study, we sequenced and assembled the genome of *L. nagelii* VUCC-R001, previously isolated from kombucha tea (Ferremi Leali

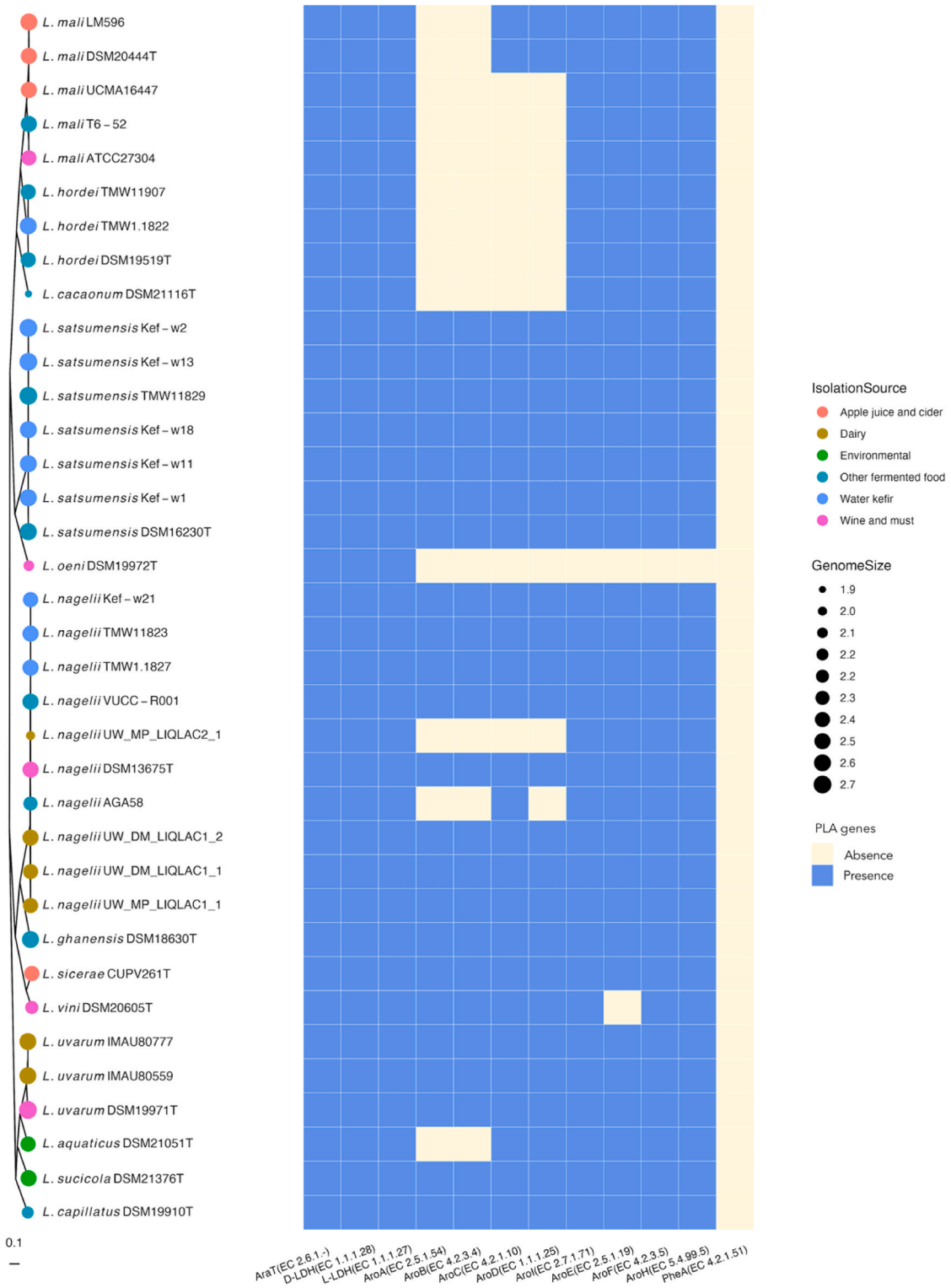

Fig. 6. Heatmap with D-phenyllactic acid (PLA) pathway genes obtained by BLAST analysis for the dataset. The heatmap points out the presence (sugar-paper blue) or absence (cream white) of PLA genes associated with the unrooted phylogenetic tree of the concatenated 499 core-genes for the 36 *Liquorilactobacillus* genomes analyzed. Moreover, in the tree the genome size as the size of each node was displayed and the isolation sources as the different color of the nodes were depicted. Gene acronyms and pathways are explained below. CORE PATHWAY (AraT: aromatic amino acid aminotransferase (EC 2.6.1.-), D-LDH: D-lactate dehydrogenase (EC 1.1.1.28),L-LDH: L-lactate dehydrogenase (EC 1.1.1.27)); DE NOVO BIO-SYNTHETIC PATHWAY (AroA: 3-deoxy-7-phosphoheptulonate synthase (EC 2.5.1.54), AroB: 3-dehydroquinate synthase (EC 4.2.3.4), AroC: 3-dehydroquinate dehydratase (EC 4.2.1.10), AroD: shikimate 5-dehydrogenase (EC 1.1.1.25), AroI: shikimate kinase (EC 2.7.1.71), AroE: 3-phosphoshikimate 1-carboxy-vinyltransferase (EC 2.5.1.19), AroF: chorismate synthase (EC 4.2.3.5), AroH: chorismate mutase II (EC 5.4.99.5), PheA: prephenate dehydratase (EC 4.2.1.51). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

et al., 2022), and explored its biotechnological potential, mainly in terms of D-phenyllactic acid and dextran production, through a comparative genomic approach with 35 other genomes of *Liquorilactobacillus* species (already available in NCBI) and related phenotypic validation. Furthermore, a thorough safety assessment was conducted as a prerequisite for its eventual deliberate use.

The sequencing of the *L. nagelii* VUCC-R001 genome provided high quality data both in terms of completeness and quality of the assembly: the amount of contigs obtained (37, Supplementary File S3) was less than 500, and the assembly size (~2.4 Mb, Supplementary File S1) was within +/− 20% of the expected genome size for the species (between 2.0 and 2.6 Mb) as recommended by EFSA (2021).

Besides confirming VUCC-R001 as allotted to *L. nagelii* species, genome comparison showed that the 36 *Liquorilactobacillus* strains formed a compact group with ANI values ranging from 70% to 78% (Zheng et al., 2020). The phylogenomic analysis based on 130 core genes highlighted the presence of the same clades reported in Liu and colleagues (clades I, II, III, IV) (Liu et al., 2023) based on 122 core genes and 24 genomes, thus proving the evolutionary relatedness within the *Liquorilactobacilllus* species.

The size of the pangenome (14,593 genes) pointed out a large heterogeneity among the species under analysis, which could be linked to the presence of two different clades (A and B) which emerged from the principal component analysis of the pangenome performed by (Liu et al., 2023). It was observed that clade A (enclosing *L. nagelii* and the majority of *Liquorilactobacillus* spp. except *L. cacaonum, L. hordei* and *L. mali*) had much greater genetic diversity than the clade B, most probably due to the number of genomes included in each clade (27 and 9, respectively) but also due to differences in motor capacity (clade A groups motile species while non-motile species fall into clade B) and clade-specific genomic signatures, such as prophage infection (clade B) and insertion sequences activity (clade A).

The phylogenomic analysis was used as a reference to better delineate the distribution of safety-related traits and functional characteristics, as discussed below.

In terms of safety, *L. nagelii* VUCC-R001 did not show any antibiotic resistance either at genotypic or phenotypic level. All the genomes of *Liquorilactobacillus* species showed the gene *van*T of the cluster *van*G which make their cell wall poorly affinitive to vancomycin and conferring an intrinsic resistance, as already observed for most lactobacilli (Campedelli et al., 2019). This gene encodes for the membrane protein serine racemase, which produces D-serine from L-serine. In turn, it is linked via the VanC ligase to a D-alanine, thus obtaining D-alanine-D-serine instead of D-alanine-D-alanine, reducing its affinity for vancomycin (Courvalin, 2006). It has been reported that the ability of some lactobacilli to have a different synthesis of peptidoglycan precursors seems to lead to a selective advantage in niches containing producers of glycopeptide antibiotic, especially for lactobacilli associated with plant fermentations, such as *Lactiplantibacillus plantarum*, *Levilactobacillus brevis*, *Lacticaseibacillus casei* and *Limosilactobacillus fermentum* (Kleerebezem et al., 2010).

Other strains of the genus *Liquorilactobacillus* have been found to possess the gene *van*H, which encodes a dehydrogenase that reduces pyruvate to D-lactate, providing the substrate for forming D-alanine-D-lactate instead of D-alanine- D-alanine (Courvalin, 2006). This gene is also related to vancomycin resistance, and it is one of the gene clusters of *van*A, which can be found associated with genetic mobile elements (Williams and Hergenrother, 2008), unlike *van*T which is usually present at the chromosomal level and, therefore, is less likely to be transferred.

The *in-silico* data were then confirmed by the determination of MICs in *L. nagelii* VUCC-R001 and *L. nagelii* DSM 13675^T, which revealed only phenotypic resistance towards vancomycin. Interestingly, data obtained on *L. nagelii* DSM 13675^T are different than those obtained by (Campedelli et al., 2019), where this strain was found to be resistant to kanamycin, chloramphenicol, and ampicillin through the application of vetMIC plates.

In this investigation, two cut-offs of the EFSA guidance (2018) were considered (for homofermentative species and facultative heterofermentative species) since *L. nagelii* is regarded as a homofermentative species according to the 2020 *Lactobacillus* reclassification (Zheng et al., 2020), but phylogenetically related to *Ligilactobacillus salivarius* (for which facultatively heterofermentative cut-offs are recommended). This interesting case study shows that after the description of novel genera from the former genus *Lactobacillus*, LAB metabolic groups in the EFSA guidance are obsolete, as they do not refer to the current homofermentative/heterofermentative distinction.

For the other safety aspects considered, such as the production of biogenic amines and virulence factors, no critical issues were detected.

Regarding the biotechnological application, our first analysis was aimed at the production of exopolysaccharides. VUCC-R001 was isolated from samples of SCOBY (Symbiotic Culture of Bacteria and Yeasts), a matrix composed by AAB, LAB and yeasts embedded in an extracellular cellulosic structure used for the artisanal kombucha production, a sucrose-rich environment (Ferremi Leali et al., 2022).

The presence of a GH70, corresponding to a dextransucrase (or 1,6-α-D-glucosyltransferase), in 75% of the genomes in the dataset, confirmed that the potential dextran production from sucrose is a general feature of *Liquorilactobacillus* spp. (Ganzle et al., 2023; Zheng et al., 2020). The lack of GH70 in *L. cacaonum*, *L. vini*, *L. ghanensis*, *L. sicerae* and *L. oeni* agreed with previously reported phenotypic data related to the absence of dextran or to the production of other EPS (De Bruyne et al., 2009; Manes-Lazaro et al., 2009; Nielsen et al., 2007; Puertas et al., 2023; Rodas et al., 2006). Among L. *nagelii,* all strains, including VUCC-R001, showed the presence of GH70, except *L. nagelii* DSM 13675^T. These predictions were further confirmed phenotypically since, among the two strains, only VUCC-R001 produced dextran from sucrose. Although the genotype-phenotype correlation for this trait in the two strains is met in this study, the absence of dextran production in DSM 13675^T is not consistent with the description of species *L. nagelii* by (Edwards et al., 2000). Interestingly, the dextransucrase of VUCC-R001 is 94% similar (Supplementary File S8) to that of *L. nagelii* TMW 1.1827, a strain isolated from water kefir, which harbors a C-terminal glucan-binding domain and a N-terminal signal motif that likely protect the enzyme from proteolytic digestion, leading to stable polysaccharide formation (Bechtner et al., 2019). Data obtained on GH70 in VUCC-R001 constitute the starting point for further investigations, mainly related to the characterization of the dextran produced and its possible industrial applications; an in-depth analysis could also open the prospects of the use of this polysaccharide as a prebiotic, as demonstrated for the EPS released by *L. satsumensis* Kef-w1 (Tan et al., 2022).

The production of compounds with antimicrobial activity is remarkable in characterizing food-related strains since their application as natural preservatives can limit the use of antibiotics or other chemical compounds in the food chain. Among natural preservatives, PLA is an organic acid produced by LAB that shows a broad spectrum of inhibition towards Gram-positive and Gram-negative bacteria and fungi (Wu et al., 2023). LAB primarily synthesize PLA from phenylalanine via a 2-step core pathway (transamination of phenylalanine to phenylpyruvic acid and reduction to PLA); more rarely, PLA can also be synthesized *de novo* from glucose (Wu et al., 2020). In this study, we reported for the first time the investigation of both core and *de novo* pathways for PLA synthesis in the genus *Liquorilactobacillus*, demonstrating that all strains included in the dataset had the core pathway and 60% of them (including *L. nagelii* strains) are also characterized by the genetic background for the *de novo* production of PLA except prephenate dehydratase.

Given the presence of a complete core pathway, PLA production was assessed phenotypically in both *L. nagelii* VUCC-R001 and DSM 13675^T, which produced 52.8 mg/L and 32.2 mg/L, respectively. These values were higher than those obtained by *Lacticaseibacillus casei* NRRL B-1922^T (5.8 mg/L), used as reference as reported by (Cortes-Zavaleta

et al., 2014). Although these values were not among the highest produced by LAB (reviewed in (Wu et al., 2023)), they could contribute to the antimicrobial action towards food spoilage molds (Cortes-Zavaleta et al., 2014). To date, this is the first study reporting PLA production by a strain isolated from kombucha tea and this feature, in combination with the release of other organic acids, large proteins and catechins, could further substantiate the antimicrobial activity already described for this fermented beverage (Sreeramulu et al., 2000). Furthermore, the amount of PLA produced by VUCC-R001 is comparable to that released by *Lactiplantibacillus plantarum* IDCC 3501, which showed a significant tyrosinase-inhibitory activity, opening the way for the potential application of PLA as an anti-melanogenesis agent in food and medicine (Shin et al., 2023).

5. Conclusions

The data obtained in this study confirmed the importance of genome mining and metabolite analyses to evaluate the functional and safety aspects of *Liquorilactobacillus* species, with a particular focus on *L. nagelii,* which could have an application in food systems. The genotype-phenotype correlation approach that we applied for the characterization of *L. nagelii* VUCC-R001 from kombucha tea allowed us to detect resistance to vancomycin, which can, however, be considered intrinsic, given its distribution throughout the genus. Further, genome analysis and related phenotypic investigation unravelled the production of high-value bioactive compounds, such as dextran and D-phenyllactic acid. Although further investigations are needed to understand better the mode of production and features of these metabolites, the data obtained are promising. They could be combined to form a robust body of knowledge for the inclusion of *L. nagelii* in the EFSA QPS list, kicking off its application in the fermented beverage industry.

Funding sources

This study was carried out within the Interconnected Nord-Est Innovation Ecosystem (iNEST) and received funding from the European Union - Next-Generation EU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.5 – D.D. 1058 June 23, 2022, ECS00000043). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them. Part of this work was granted by the European Commission – Next Generation EU, Project "Strengthening the MIRRI Italian Research Infrastructure for Sustainable Bioscience and Bioeconomy", code n. IR0000005. This work was partly developed within the research program "Dipartimento di Eccellenza 2023–2027" (Italian Ministry of University and Research).

CRediT authorship contribution statement

Ilaria Larini: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sarah Tintori:** Writing – original draft, Software, Investigation. **Veronica Gatto:** Validation, Methodology, Formal analysis. **Giovanna E. Felis:** Supervision, Conceptualization. **Elisa Salvetti:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Formal analysis, Conceptualization. **Sandra Torriani:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Conceptualization.

Declaration of competing interest

The authors confirm that they have no conflicts of interest with respect to the work described in this manuscript.

Data availability

Genome data included in the manuscript are publicly available and deposited in NCBI.

Acknowledgements

The authors would like to thank Federica Mainente for assistance in the HPLC analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fbio.2024.104001.

References

Alcock, B. P., Huynh, W., Chalil, R., Smith, K. W., Raphenya, A. R., Wlodarski, M. A., Edalatmand, A., Petkau, A., Syed, S. A., Tsang, K. K., Baker, S. J. C., Dave, M., McCarthy, M. C., Mukiri, K. M., Nasir, J. A., Golbon, B., Imtiaz, H., Jiang, X., Kaur, K., … McArthur, A. G. (2023). CARD 2023: Expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Research, 51*, D690–D699.

Ammor, M. S., Florez, A. B., & Mayo, B. (2007). Antibiotic resistance in non - enterococcal lactic acid bacteria and bifidobacteria. *Food Microbiology, 24*, 559–570.

Auch, A. F., von Jan, M., Klenk, H. P., & Goker, M. (2010). Digital DNA - DNA hybridization for microbial species delineation by means of genome - to - genome sequence comparison. *Standards in Genomic Sciences, 2*, 117–134.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single - cell sequencing. *Journal of Computational Biology, 19*, 455–477.

Bechtner, J., Wefers, D., Schmid, J., Vogel, R. F., & Jakob, F. (2019). Identification and comparison of two closely related dextransucrases released by water kefir borne *Lactobacillus hordei* TMW 1.1822 and *Lactobacillus nagelii* TMW 1.1827. *Microbiology (Reading), 165*, 956–966.

Bjorland, J., Steinum, T., Sunde, M., Waage, S., & Heir, E. (2003). Novel plasmid - borne gene *qacJ* mediates resistance to quaternary ammonium compounds in equine *Staphylococcus aureus, Staphylococcus simulans*, and *Staphylococcus intermedius*. *Antimicrobial Agents and Chemotherapy, 47*, 3046–3052.

Bover-Cid, S., & Holzapfel, W. H. (1999). Improved screening procedure for biogenic amine production by lactic acid bacteria. *International Journal of Food Microbiology, 53*, 33–41.

Campedelli, I., Mathur, H., Salvetti, E., Clarke, S., Rea, M. C., Torriani, S., Ross, R. P., Hill, C., & O'Toole, P. W. (2019). Genus - wide assessment of antibiotic resistance in *Lactobacillus* spp. *Applied and Environmental Microbiology, 85*(18), Article e01738.

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). Fastp: An ultra - fast all - in - one FASTQ preprocessor. *Bioinformatics, 34*, i884–i890.

Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D. R., da Costa, M. S., Rooney, A. P., Yi, H., Xu, X. W., De Meyer, S., & Trujillo, M. E. (2018). Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology, 68*, 461–466.

Cortes-Zavaleta, O., Lopez-Malo, A., Hernandez-Mendoza, A., & Garcia, H. S. (2014). Antifungal activity of lactobacilli and its relationship with 3 - phenyllactic acid production. *International Journal of Food Microbiology, 173*, 30–35.

Courvalin, P. (2006). Vancomycin resistance in Gram - positive cocci. *Clinical Infectious Diseases, 42*, S25–S34.

De Bruyne, K., Camu, N., De Vuyst, L., & Vandamme, P. (2009). *Lactobacillus fabifermentans* sp. nov. and *Lactobacillus cacaonum* sp. nov., isolated from Ghanaian cocoa fermentations. *International Journal of Systematic and Evolutionary Microbiology, 59*, 7–12.

Drula, E., Garron, M. L., Dogan, S., Lombard, V., Henrissat, B., & Terrapon, N. (2022). The carbohydrate - active enzyme database: Functions and literature. *Nucleic Acids Research, 50*, D571–D577.

Edwards, C. G., Collins, M. D., Lawson, P. A., & Rodriguez, A. V. (2000). *Lactobacillus nagelii* sp. nov., an organism isolated from a partially fermented wine. *International Journal of Systematic and Evolutionary Microbiology, 50*, 699–702.

European Food Safety Authority (EFSA). (2018). Guidance on the characterisation of microorganisms used as feed additives or as production organisms. *Annex to the EFSA Journal, 16*, 5206. Internet: https://efsa.onlinelibrary.wiley.com/doi/10.2903/j.efsa.2018.5206.

European Food Safety Authority (EFSA). (2021). EFSA statement on the requirements for whole genome sequence analysis of microorganisms intentionally used in the food chain. *Annex to the EFSA Journal, 19*, 6506. Internet: https://efsa.onlinelibrary.wiley.com/doi/10.2903/j.efsa.2021.6506.

Ferremi Leali, N., Binati, R. L., Martelli, F., Gatto, V., Luzzini, G., Salini, A., Slaghenaufi, D., Fusco, S., Ugliano, M., Torriani, S., & Salvetti, E. (2022). Reconstruction of simplified microbial consortia to modulate sensory quality of kombucha tea. *Foods, 11*, 3045.

Ferres, I., & Iraola, G. (2021). Protocol for post - processing of bacterial pangenome data using Pagoo pipeline. *STAR Protocols, 2*, Article 100802.
Florensa, A. F., Kaas, R. S., Clausen, P., Aytan-Aktug, D., & Aarestrup, F. M. (2022). ResFinder - an open online resource for identification of antimicrobial resistance genes in next - generation sequencing data and prediction of phenotypes from genotypes. *Microbial Genomics, 8*, 1–12.
Ganzle, M. G., Monnin, L., Zheng, J., Zhang, L., Coton, M., Sicard, D., & Walter, J. (2023). Starter culture development and innovation for novel fermented foods. *Annual Review of Food Science and Technology, 15*(5), 1–5.29.
Gaspar, F. B., & Crespo, M. T. (2016). Amylolytic lactic acid bacteria: Microbiology and technological interventions in food fermentations. In D. Montet, & C. R. Ray (Eds.), *Fermented foods, Part I* (pp. 148–165). Boca Raton: CRC Press.
Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., & Tiedje, J. M. (2007). Dna - DNA hybridization values and their relationship to whole - genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology, 57*, 81–91.
Grant, J. R., Enns, E., Marinier, E., Mandal, A., Herman, E. K., Chen, C. Y., Graham, M., Van Domselaar, G., & Stothard, P. (2023). Proksee: In - depth characterization and visualization of bacterial genomes. *Nucleic Acids Research, 51*, W484–W492.
Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). Quast: Quality assessment tool for genome assemblies. *Bioinformatics, 29*, 1072–1075.
International Dairy Federation (IDF). (2022). Inventory of microbial food cultures with safety demonstration in fermented food products. *Bulletin N°*, 514–2022. Internet: https://shop.fil-idf.org/products/bulletin-of-the-idf-n-514-2022-inventory-of-microbial-food-cultures-with-safety-demonstration-in-fermented-food-products.
Jung, J. Y., Kang, H. K., Jin, H. M., Han, S. S., Kwon, Y. C., Eun, J. J., Kim, S. C., Seo, M. J., Ryu, B. G., & Chung, E. J. (2019). *Companilactobacillus pabuli* sp. nov., a lactic acid bacterium isolated from animal feed. *International Journal of Systematic and Evolutionary Microbiology, 71*.
Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of Molecular Biology, 428*, 726–731.
Kersani, I., Zadi-Karam, H., & Karam, N.-E. (2017). Screening of exopolysaccharide - producing coccal lactic acid bacteria isolated from camel milk and red meat of Algeria. *African Journal of Biotechnology, 16*, 1078–1084.
Kim, D. K., Kim, K. H., Cho, E. J., Joo, S. J., Chung, J. M., Son, B. Y., Yum, J. H., Kim, Y. M., Kwon, H. J., Kim, B. W., Kim, T. H., & Lee, E. W. (2013). Gene cloning and characterization of MdeA, a novel multidrug efflux pump in *Streptococcus mutans*. *Journal of Microbiology and Biotechnology, 23*, 430–435.
Kleerebezem, M., Hols, P., Bernard, E., Rolain, T., Zhou, M., Siezen, R. J., & Bron, P. A. (2010). The extracellular biology of the lactobacilli. *FEMS Microbiology Reviews, 34*, 199–230.
Laureys, D., & De Vuyst, L. (2014). Microbial species diversity, community dynamics, and metabolite kinetics of water kefir fermentation. *Applied and Environmental Microbiology, 80*, 2564–2572.
Li, K. J., Brouwer-Brolsma, E. M., Burton-Pimentel, K. J., Vergeres, G., & Feskens, E. J. M. (2021). A systematic review to identify biomarkers of intake for fermented food products. *Genes & Nutrition, 16*, 5.
Liu, X., Li, W., Sun, Z., Zhong, Z., & Sun, T. (2023). Phylogenomics of the *Liquorilactobacillus* genus. *Current Microbiology, 80*, 274.
Liu, B., Zheng, D., Zhou, S., Chen, L., & Yang, J. (2022). VFDB 2022: A general classification scheme for bacterial virulence factors. *Nucleic Acids Research, 50*, D912–D917.
Manes-Lazaro, R., Ferrer, S., Rossello-Mora, R., & Pardo, I. (2009). *Lactobacillus oeni* sp. nov., from wine. *International Journal of Systematic and Evolutionary Microbiology, 59*, 2010–2014.
Manni, M., Berkeley, M. R., Seppey, M., & Zdobnov, E. M. (2021). BUSCO: Assessing genomic data quality and beyond. *Current Protocols in Cell Biology, 1*, e323.
Meier-Kolthoff, J. P., Carbasse, J. S., Peinado-Olarte, R. L., & Goker, M. (2022). TYGS and LPSN: A database tandem for fast and reliable genome - based classification and nomenclature of prokaryotes. *Nucleic Acids Research, 50*, D801–D807.
Mukherjee, A., Breselge, S., Dimidi, E., Marco, M. L., & Cotter, P. D. (2023). Fermented foods and gastrointestinal health: Underlying mechanisms. *Nature Reviews Gastroenterology & Hepatology, 21*, 248–266.
Nielsen, D. S., Schillinger, U., Franz, C., Bresciani, J., Amoa-Awua, W., Holzapfel, W. H., & Jakobsen, M. (2007). *Lactobacillus ghanensis* sp. nov., a motile lactic acid bacterium isolated from Ghanaian cocoa fermentations. *International Journal of Systematic and Evolutionary Microbiology, 57*, 1468–1472.
Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F., & Stevens, R. (2014). The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Research, 42*, D206–D214.
Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., Fookes, M., Falush, D., Keane, J. A., & Parkhill, J. (2015). Roary: Rapid large - scale prokaryote pan genome analysis. *Bioinformatics, 31*, 3691–3693.
Puertas, A. I., Llamas-Arriba, M. G., Etxebeste, O., Berregi, I., Pardo, M. A., Prieto, A., Lopez, P., & Duenas, M. T. (2023). Characterization of the heteropolysaccharides produced by *Liquorilactobacillus sicerae* CUPV261 and *Secundilactobacillus collinoides* CUPV237 isolated from cider. *International Journal of Food Microbiology, 397*, Article 110199.
Rajanikar, R. V. H. N. B., Naithani, H., Ali, S. A., Panjagari, N. R., & Behare, P. V. (2021). Phenyllactic acid: A green compound for food biopreservation. *Food Control, 128*, Article 108184.
Rivera, A., Becerra-Martínez, E.l., Pacheco-Hernández, Y., Landeta-Cortés, G., & Villa-Ruano, N. (2020). Synergistic hypolipidemic and hypoglycemic effects of mixtures of *Lactobacillus nagelii*/betanin in a mouse model. *Tropical Journal of Pharmaceutical Research, 19*(6), 1269–1276.
Rodas, A. M., Chenoll, E., Macian, M. C., Ferrer, S., Pardo, I., & Aznar, R. (2006). *Lactobacillus vini* sp. nov., a wine lactic acid bacterium homofermentative for pentoses. *International Journal of Systematic and Evolutionary Microbiology, 56*, 513–517.
Savary, O., Mounier, J., Thierry, A., Poirier, E., Jourdren, J., Maillard, M. B., Penland, M., Decamps, C., Coton, E., & Coton, M. (2021). Tailor - made microbial consortium for kombucha fermentation: Microbiota - induced biochemical changes and biofilm formation. *Food Research International, Int, 147*, Article 110549.
Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics, 30*, 2068–2069.
Shin, M., Truong, V. L., Lee, M., Kim, D., Kim, M. S., Cho, H., Jung, Y. H., Yang, J., Jeong, W. S., & Kim, Y. (2023). Investigation of phenyllactic acid as a potent tyrosinase inhibitor produced by probiotics. *Current Research in Food Science, 6*, Article 100413.
Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high - quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology, 7*, 539.
Sreeramulu, G., Zhu, Y., & Knol, W. (2000). Kombucha fermentation and its antimicrobial activity. *Journal of Agricultural and Food Chemistry, 48*, 2589–2594.
Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution, 38*, 3022–3027.
Tan, L. L., Ngiam, J. J., Sim, E. S. Z., Conway, P. L., & Loo, S. C. J. (2022). *Liquorilactobacillus satsumensis* from water kefir yields alpha - glucan polysaccharides with prebiotic and synbiotic qualities. *Carbohydrate Polymers, 290*, Article 119515.
Tettelin, H., Riley, D., Cattuto, C., & Medini, D. (2008). Comparative genomics: The bacterial pan - genome. *Current Opinion in Microbiology, 11*, 472–477.
Umegatani, M., Takesue, N., Asano, S., Tadami, H., & Uemura, K. (2022). Study of beer spoilage *Lactobacillus nagelii* harboring hop resistance gene *hor*A. *Journal of the American Society of Brewing Chemists, 80*(1), 92–98.
Valerio, F., Di Biase, M., Lattanzio, V. M., & Lavermicocca, P. (2016). Improvement of the antifungal activity of lactic acid bacteria by addition to the growth medium of phenylpyruvic acid, a precursor of phenyllactic acid. *International Journal of Food Microbiology, 222*, 1–7.
Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., & Barton, G. J. (2009). Jalview version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics, 25*, 1189–1191.
Williams, J. J., & Hergenrother, P. J. (2008). Exposing plasmids as the Achilles' heel of drug - resistant bacteria. *Current Opinion in Chemical Biology, 12*, 389–399.
Wu, W., Deng, G., Liu, C., Gong, X., Ma, G., Yuan, Q., Yang, E., Li, X., & Luo, Y. (2020). Optimization and multiomic basis of phenyllactic acid overproduction by *Lactobacillus plantarum*. *Journal of Agricultural and Food Chemistry, 68*, 1741–1749.
Wu, H., Guang, C., Zhang, W., & Mu, W. (2023). Recent development of phenyllactic acid: Physicochemical properties, biotechnological production strategies and applications. *Critical Reviews in Biotechnology, 43*, 293–308.
Xu, J. J., Fu, L. J., Si, K. L., Yue, T. L., & Guo, C. F. (2020). 3 - phenyllactic acid production by free -whole - cells of *Lactobacillus crustorum* in batch and continuous fermentation systems. *Journal of Applied Microbiology, 129*, 335–344.
Xu, S., Li, L., Luo, X., Chen, M., Tang, W., Zhan, L., Dai, Z., Lam, T., Guan, Y., & Yu, G. (2022). Ggtree: A serialized data object for visualization of a phylogenetic tree and annotation data. *iMeta, 1*, e56.
Yang, J., Tan, H., & Cai, Y. (2016). Characteristics of lactic acid bacteria isolates and their effect on silage fermentation of fruit residues. *Journal of Dairy Science, 99*, 5325–5334.
Yetiman, A. E., & Ortakci, F. (2023). Genomic, probiotic, and metabolic potentials of *Liquorilactobacillus nagelii* AGA58, a novel bacteriocinogenic motile strain isolated from lactic acid - fermented shalgam. *Journal of Bioscience and Bioengineering, 135*, 34–43.
Zhao, N., Huang, Y., Lai, H., Wang, Y., Mei, Y., Zeng, X., Yang, M., Li, H., Zhao, J. Z., Yong, Z. Y., & Ge, L. (2023). Illumination and reconstruction of keystone microbiota for reproduction of key flavor - active volatile compounds during paocai (a traditional fermented vegetable) fermentation. *Food Bioscience, 56*, Article 103148.
Zheng, J., Ge, Q., Yan, Y., Zhang, X., Huang, L., & Yin, Y. (2023). dbCAN3: Automated carbohydrate - active enzyme and substrate annotation. *Nucleic Acids Research, 51*, W115–W121.
Zheng, J., Wittouck, S., Salvetti, E., Franz, C., Harris, H. M. B., Mattarelli, P., O'Toole, P. W., Pot, B., Vandamme, P., Walter, J., Watanabe, K., Wuyts, S., Felis, G. E., Ganzle, M. G., & Lebeer, S. (2020). A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of Lactobacillaceae and Leuconostocaceae. *International Journal of Systematic and Evolutionary Microbiology, 70*, 2782–2858.