

Intra-individual genetic diversity of vaginal *Lactobacillus crispatus* revealed through citizen science-driven isolation and pangenome analysis

Freek de Kreek¹, Rosanne Hertzberger^{1,2}, Fleur van Eeden¹, Shardelice Illidge¹, Evert-Jan Teunis³, Mark Hanemaaijer³, Elke Lievens³, Fanny Rienstra⁴, Deniz Er Wiedhaup⁴, Paola Lisotto⁵, Derek Butler⁵, Douwe Molenaar¹, Remco Kort^{1,*}

¹Amsterdam Institute for Life and Environment, Vrije Universiteit Amsterdam, De Boelelaan 1108, 1081 HZ Amsterdam, The Netherlands

²Stichting Crispatus, Rembrandtlaan 22, 2665 XG Bleiswijk, The Netherlands

³Winclove B.V., tt. Vasumweg 221, 1033 SJ Amsterdam, The Netherlands

⁴Yoni B.V., Pand Noord, Meeuwenlaan 98–100, 1021 JL Amsterdam, The Netherlands

⁵BaseClear B.V., Sylviusweg 74, 2333 BE Leiden, The Netherlands

*Corresponding author. Amsterdam Institute for Life and Environment, Vrije Universiteit Amsterdam, De Boelelaan 1108, 1081 HZ Amsterdam, The Netherlands. E-mail: r.kort@vu.nl

Abstract

Aims A vaginal microbiome dominated by *Lactobacillus crispatus* is associated with positive reproductive and sexual health outcomes, yet intra-individual genetic diversity within this species remains largely unexplored. This study characterized inter- and intra-individual genomic variation in *L. crispatus* strains isolated through a citizen science initiative and assessed implications for multi-strain probiotic development.

Methods and results Fifty-three women participated in this citizen science project. Self-sampling resulted in 48 shotgun metagenomes. Twenty-two participants isolated their own *L. crispatus* strains using selective enrichment and LAMP-based species confirmation, resulting in 53 whole-genome-sequenced isolates. *Lactobacillus crispatus* dominated 20 of 48 metagenomes (50.1%–99.6% relative abundance). Pangenome analysis revealed 3456 gene families, of which 43.7% were core and 56.3% accessory. A 14-kb plasmid harbouring a Fic-domain toxin–antitoxin protein, but devoid of antimicrobial resistance genes, was present in 44 of 53 strains. Strains from the same individual clustered closely together yet harboured 1–123 gene differences. Intra-individual variation was observed in the pullulanase type I gene required for glycogen degradation: 40 strains were predicted to grow on glycogen, six showed genetic disruptions with unknown consequences, and seven were predicted to lack this ability entirely. Variation within individuals was also found for bacteriocin classes and CRISPR-Cas genes.

Conclusions Substantial functional diversity exists within *L. crispatus*, even among strains from the same individual, supporting the rationale for multi-strain vaginal probiotics. This citizen science approach enabled discovery of host-specific adaptations while ensuring participant ownership of their strains.

Impact statement

This study is the first to systematically characterize the genetic diversity of multiple *Lactobacillus crispatus* isolates from within the same individual, revealing that functionally distinct strains co-exist in a single vaginal microbiome. These findings challenge the assumption that a single probiotic strain can recapitulate the diversity of an individual's native *L. crispatus* population and support the development of personalized, multi-strain vaginal probiotics. The citizen science design of this study offers a framework for participant-driven microbiome research in which women retain ownership of their bacterial strains.

Keywords *Lactobacillus crispatus*, vaginal microbiome, pangenome analysis, intra-individual diversity, citizen science, probiotics

Introduction

The human vaginal microbiome is unusual among body-site-associated microbial communities because it is frequently dominated by a single bacterial species (Ravel *et al.* 2011). Five com-

munity state types (CSTs) have been described: four are dominated by lactobacilli, while a fifth type is characterized by a diverse, non-*Lactobacillus*-dominated community often associated with dysbiosis. Among these, *Lactobacillus crispatus* is considered the species most strongly associated with favourable reproductive

and sexual health outcomes in women of reproductive age (France *et al.* 2022).

Lactobacillus crispatus is associated with vaginal health and is thought to contribute to it through several generic and strain-variable mechanisms: it acidifies the vaginal environment by producing lactic acid, which suppresses a broad range of unrelated taxa such as *Gardnerella* spp., it produces bacteriocins, which act on closely related Gram-positive bacteria, it adheres to vaginal epithelial cells, and it degrades glycogen, a primary carbon source in the vaginal tract (Jenkins *et al.* 2023). Recent work has further demonstrated that *L. crispatus* exopolysaccharides modulate host immune responses and mediate interactions with the vaginal mucosa (Croatti *et al.* 2026). These properties have made *L. crispatus* the leading candidate for vaginal probiotic development as a complement to antibiotic treatment of vaginal dysbiosis (Cohen *et al.* 2020).

Lactobacillus crispatus is a genetically diverse species, with genome sizes ranging from 2.0 to 2.5 Mb and an open pangenome (Pan *et al.* 2020, Fontana *et al.* 2021). Comparative genomic studies have revealed inter-individual variation in glycogen metabolism, bacteriocin repertoires, and surface-associated proteins (van der Veer *et al.* 2019, Mancabelli *et al.* 2021), while core genome analyses have identified single-nucleotide variation in conserved genes as a driving force for niche competition (Tarracchini *et al.* 2023). However, the extent of genetic variation *within* a single individual remains poorly resolved. Metagenomic approaches have provided indirect evidence for intra-individual strain co-existence: the VIRGO gene catalogue showed that individual vaginal metagenomes contain on average 1.6 times more *L. crispatus* genes than a single genome, indicating the routine presence of multiple strains (Ma *et al.* 2020). Yet these approaches cannot link specific variable genes to specific co-existing strains. Our previous analysis of pullulanase variants in over 1500 vaginal metagenomes in the VIRGO database detected only a single variant per sample (Hertzberger *et al.* 2022), and a recent metagenome-assembled genome (MAG)-based study of 354 metagenomes was unable to recover multiple *L. crispatus* genomes from any single sample (Nori *et al.* 2025). The latter study showed strain-level variation in cell wall biogenesis, carbohydrate metabolism, and adhesion-related genes between individuals using MAGs, but intra-individual differences in these traits remain unresolved. Isolate-based approaches, in which multiple strains are cultured from the same individual, can overcome these limitations but have not been applied to the vaginal microbiome.

Citizen science approaches have recently proven powerful for large-scale vaginal microbiome mapping, as demonstrated by the Isala project (Lebeer *et al.* 2023), which characterized the vaginal microbiomes of 3345 women through self-sampling. In parallel, we developed a citizen science protocol in which women collected their own vaginal swabs, performed selective enrichment, and identified *L. crispatus* isolates using loop-mediated isothermal amplification (LAMP) during a supervised practicum (Illidge *et al.* 2024). This protocol was carried out by 53 women, organized by the *crispatus* foundation, which ensures that participants retain ownership of their strains and collectively govern future licensing decisions. Twenty-two participants successfully isolated their own strains, yielding 53 whole-genome-sequenced *L. crispatus* isolates alongside 48 matched shotgun metagenomes.

Here, we combine metagenomic profiling and pangenome analysis to characterize the inter- and intra-individual genetic diversity of *L. crispatus*. We systematically assess traits relevant to probiotic function, including glycogen degradation capacity, bacteriocin production, adhesion-related proteins, and glycosyltransferase (GT) fragments involved in surface glycosylation, as well as safety markers such as antimicrobial resistance genes, virulence factors, and a previously undescribed plasmid. By comparing multiple isolates from the same individual, we aim to evaluate the functional consequences of intra-individual diversity and to inform the rational design of multi-strain vaginal probiotics. Multi-strain *L. crispatus* live biotherapeutic products have recently entered clinical testing as well (Potloane *et al.* 2026), underscoring the translational relevance of characterizing strain-level diversity within and between individuals.

Materials and methods

Ethics statement

The Medical Ethics Committee (METC) of the Vrije Universiteit Amsterdam reviewed the research protocol (reference 2022.0791) and concluded that the Dutch Medical Research Involving Human Subjects Act (WMO) does not apply to this study (Illidge *et al.* 2024). All participants provided written informed consent prior to participation. To protect participant identity, each citizen scientist chose a pseudonym unrelated to their real name; participants also named the bacterial strains they isolated. The study was conducted in accordance with the Declaration of Helsinki.

Metagenomic profiling of the vaginal microbiome

Vaginal swabs were self-collected by 53 women (aged 18–45 years) who participated as citizen scientists between 30 April and 1 September 2023, coordinated by the *crispatus* foundation (Dutch: Stichting *crispatus*). Sample collection and the citizen science practicum have been described previously (Illidge *et al.* 2024). Metagenomic DNA was extracted from swabs using the ZymoBIOMICS-96 MagBead DNA Kit (Zymo Research, D4302) on the KingFisher Flex robot. The extracted DNA was subjected to quality control using a 16S rRNA quantitative polymerase chain reaction (qPCR) assay that specifically targets bacterial biomass (Suzuki *et al.* 2000) with a cut-off of 2.75×10^3 gene copies μL^{-1} for successful metagenomic sequencing, and to short-read shotgun sequencing on an Illumina (San Diego, CA, USA) platform, based on a Nextera XT sequencing library prepared and sequenced on a NovaSeq 6000 instrument with a paired-end 150 nt protocol. Raw sequencing data were demultiplexed and converted to FASTQ files with bcl2fastq version 2.20 (Illumina), and an initial quality assessment was based on reads passing the Illumina Chastity filter. Reads containing PhiX control signal were removed with an in-house filtering protocol, and reads containing (partial) adapters were clipped up to a minimum read length of 50 bp. A second quality assessment of the remaining reads was performed with FastQC version 0.11.8. Human host reads were depleted by aligning the reads to the human reference genome GRCh38 with Bowtie 2 version 2.4.5 (Langmead and Salzberg 2012) at default settings; the resulting alignments were processed with SAMtools

version 1.9 (Danecek *et al.* 2021) and BBMap version 38.79 (Bushnell 2014) to recover the unaligned paired and singleton reads. No subsampling or rarefaction was applied, so that all host-depleted reads entered the taxonomic classification. Classification was performed with Kraken2 version 2.1.1 (Wood *et al.* 2019) against a reference database of RefSeq genomes of bacteria, fungi, archaea, and viruses downloaded in July 2021, at the default confidence threshold of 0 and the default minimum of two hit groups per classification. Kraken2 assigns each read to the lowest common ancestor of its matching k-mers and reports read counts that are not corrected for genome length or read length. Species- and genus-level relative abundances were therefore estimated from the Kraken2 output with Bracken version 2.6.0 (Lu *et al.* 2017) at default settings for a read length of 150 bp. Only bacterial classifications were retained, and abundances were renormalized to the bacterial fraction; the 10 most abundant species across samples were plotted individually, with all remaining taxa grouped as 'other'. A graphical overview of the complete bioinformatic workflow, including all tools, version numbers, and reference databases, is provided in [Supplementary Fig. S1](#).

Strain isolation and whole-genome sequencing

Of the 53 participants, 22 successfully isolated their own *L. crispatus* strains during the practicum, following the selective enrichment and LAMP-based identification protocol described previously (Illidge *et al.* 2024). Briefly, vaginal swabs were cultured in tryptic soy broth supplemented with 10% horse serum, 1 g l⁻¹ Tween 80, and 10 g l⁻¹ glucose at pH 5.0, adjusted with 10% acetic acid. Species identification was performed by participants using LAMP and microscopy. Microscopy was carried out under continuous supervision by trained members of the research team; participants placed their slide under the microscope, after which a team member adjusted the image, which was then viewed on a large screen or through the eyepiece and photographed for documentation. In total, 53 *L. crispatus* isolates were obtained from 22 individuals ([Supplementary Table S1](#)).

Genomic DNA was extracted from bacterial pellets using the KingFisher 96 system (Thermo Fisher Scientific, Waltham, MA, USA) with the Chemagic DNA Bacteria Kit (Revvity chemagen, Baesweiler, Germany). Sequencing libraries were prepared with the Nextera XT DNA Library Prep Kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. Libraries were pooled and sequenced on an Illumina MiSeq platform using a v3 cartridge generating 2 × 300 bp paired-end reads. For four isolates from a single participant, additional long-read sequencing was performed on an Oxford Nanopore Technologies (Oxford, UK) MinION using the Rapid PCR Barcoding Kit according to the manufacturer's protocol.

Genome assembly and quality control

Short reads were trimmed by removing adapters and PhiX reads and filtered on quality ($Q > 30$) and length (> 100 bp). Quality control before and after trimming was assessed with FastQC version 0.12.1 (parameters: '—noextract' and '—nogroup') and reports were aggregated with MultiQC (Ewels *et al.* 2016). *De novo* assembly was performed with SPAdes version 4.0.0 (Bankevich *et al.*

2012) using the '—isolate' parameter; contigs shorter than 1000 bp were removed. Assembly quality was evaluated with QUAST version 5.2.0, applying thresholds of $N50 > 14\,000$ bp and fewer than 300 contigs.

For the four long-read-sequenced samples, reads were filtered on quality ($Q > 10$) and length (> 1000 bp) using Filtlong version 0.2.1 (Wick 2024). Hybrid *de novo* assembly combining corresponding short and long reads was performed with Unicycler version 0.5.1 (Wick *et al.* 2017) at default parameters. Assembly quality was assessed as above. Coding sequence prediction and functional annotation of all 53 assemblies were performed with Bakta version 1.9.4 (Schwengers *et al.* 2021) using the full database (version 5.1), with the parameters `-skip-pseudo`, `-genus 'Lactobacillus'`, `species 'crispatus'`.

Pangenome construction and analysis

Pangenome analysis of the 53 annotated genomes was performed with PIRATE version 1.0.5 (Bayliss *et al.* 2019) at default parameters. A multi-sequence alignment and a phylogenetic tree based on gene presence-absence were generated with PIRATE and visualized with Phandango (Hadfield *et al.* 2018). Additional pangenome visualizations were produced with the custom Python script PANDAVIS.py. A separate alignment and phylogenetic tree was constructed for isolates from the same individual to assess intra-individual variation.

Gene families identified by PIRATE were assigned to Clusters of Orthologous Groups (COG) categories based on gene descriptions, using a custom Python script (functional_categorizer.py). Protein products were matched against the EggNOG version 5.0.0 annotations database (Huerta-Cepas *et al.* 2019) by exact matching, followed by fuzzy matching (partial ratio threshold of 80) and keyword-based assignment ([Supplementary Table S2](#)). Phage-associated genes, hypothetical proteins, and proteins annotated as domain of unknown function were assigned to separate functional categories.

Analysis of potential probiotic traits and safety assessment

Potential probiotic traits and safety-relevant features were assessed *in silico* for each of the 53 strains. Plasmid presence was determined using PlasmidFinder version 2.1.6 (Carattoli and Hasmann 2020); in addition, trimmed reads were mapped against a plasmid identified in the complete assemblies from long-read sequencing using bwa-mem2 version 2.2.1 (Vasimuddin *et al.* 2019). The origin of replication was located with Ori-Finder 2022 (Dong *et al.* 2022), and read coverage was assessed with Qualimap version 2.3 (Okonechnikov *et al.* 2016) to confirm plasmid presence or absence. Virulence genes were identified with ABRicate version 1.0.1 (Seemann 2020) against the VFDB database version 6.0 (Liu *et al.* 2022), and antimicrobial resistance genes were detected with StarAMR version 0.11.0 (Zankari *et al.* 2012, 2017, Bharat *et al.* 2022); both analyses were performed on chromosomal genomes and the plasmid. The capacity for glycogen utilisation was predicted from deletions in the N-terminus of the pullulanase type I gene (glycoside hydrolase family 13), following our previously reported classification (van der Veer *et al.* 2019): no deletions indicating growth on glycogen, specific partial deletions with un-

known consequences for growth on glycogen, and extensive deletions or gene absence predicting inability to grow on glycogen. Deletions were identified from the multi-sequence alignment of the pullulanase type I gene generated during pangenome construction.

Bacteriocin-encoding genes were identified with BAGEL4 version 1.2 (van Heel *et al.* 2018) and classified into classes I, II (subclasses IIa and IIb), and III according to established nomenclature (Kumariya *et al.* 2019). The pangenome gene collection was searched for adhesion-associated proteins using the keywords: LPxTG domain-containing protein, MucBP domain-containing protein, pilus-like protein, and putative adhesin domain-containing protein. CRISPR-Cas loci in each strain were identified and classified with CRISPRcasIdentifier version 1.1.0 (Padilha *et al.* 2020). Finally, three GT fragments and associated genes (flippase, UDP-galactopyranose mutase, and GT superfamily 1 and 2 genes) were characterized by mapping reads against a 22 kb reference fragment from strain CTV-05 (van der Veer *et al.* 2019) using bwa-mem2 version 2.2.1.

Statistical analysis of genomic trait variation

To quantify intra- versus inter-individual variation, a trait vector of 21 characters was compiled for each of the 53 strains, encompassing glycogen growth prediction, plasmid presence, absence of antimicrobial resistance and virulence genes, bacteriocin classes (I, IIa, IIb, III), CRISPR-Cas subtypes (I-E, II-A, II-C, V-A, V-F, VI-B, VI-C), adhesion-related gene types (LPxTG, MucBP, type IV pilus-like, putative adhesin domain-containing proteins), glycosyltransferase fragments (GT1/GT2/GTB, GTA), and mobile genetic elements. Of these, 12 were variable across the dataset; the remaining nine were invariant (present or absent in all 53 strains) and therefore contributed zero to all pairwise distances. Glycogen growth prediction was coded as a three-state variable (1 = full growth, 0.5 = unknown, 0 = absent); all other traits were binary. For Hamming distance computation, any difference between states was counted as one.

Pairwise Hamming distances (number of differing traits) were calculated for all 1378 strain pairs. Differences between intra-individual and inter-individual distances were assessed using a one-sided Mann-Whitney U test (testing whether intra-individual distances were smaller than inter-individual distances), with rank-biserial correlation as effect size measure. Variance partitioning was performed using permutational multivariate analysis of variance (PERMANOVA; 9999 permutations, Hamming distance) with individual identity as the grouping factor. For the five individuals with multiple strains, trait-level concordance was assessed by calculating the proportion of individuals showing intra-individual variation for each trait.

Results

Lactobacillus crispatus isolation coincides with metagenomic dominance

Metagenomic profiling of 48 vaginal samples revealed considerable inter-individual variation in microbiome composition (Fig. 1).

Lactobacillus crispatus was the predominant species (>50% relative abundance) in 20 of 48 metagenomes, with relative abundances ranging from 50.1% to 99.6%. In several participants (e.g. Tiziano, Iris, Steff22, and Simon), *L. crispatus* exceeded 95% relative abundance, indicative of highly stable, *Lactobacillus*-dominated communities. Among the remaining samples, *Lactobacillus iners* dominated seven metagenomes and *Gardnerella vaginalis* was the most abundant species in 12, while some harboured *L. crispatus* at abundances as low as 0.2%. Consistent with the selective enrichment protocol, 20 of 22 participants from whom isolates were obtained had *L. crispatus*-dominated metagenomes.

The pangenome is open with a 56.3% accessory fraction

The 53 *L. crispatus* draft genomes had a mean size of 2.17 Mbp (range: 2.01–2.36 Mbp) and a mean GC content of 36.8% (range: 36.5–37.0%). The mean coding ratio was 85.6%, with an average of 2238 predicted genes per genome (range: 2036–2466). One high-quality assembly generated from long-read sequencing (isolate Bella) had a slightly larger genome (2.4 Mbp), 2427 predicted CDSs, and a coding density of 89.2% (Fig. 2). Pangenome analysis of all 53 genomes identified 3456 gene families, classified as core (1597 genes; 43.7%), soft-core (57 genes; 1.6%), shell (821 genes; 22.5%), and cloud (981 genes; 26.8%), applying the default PIRATE thresholds in which gene families were classified as core when present in at least 99% of genomes, soft-core at 95%–99%, shell at 15%–95%, and cloud below 15%. The accessory genome (soft-core, shell, and cloud combined) thus comprised 56.3% of the pangenome, consistent with the open pangenome previously described for this species (Pan *et al.* 2020).

Metabolism is conserved while phage and antimicrobial defence genes vary in the pangenome

Approximately 62.1% of predicted proteins could be assigned to functional categories (COG) based on the EggNOG database (Fig. 3; Supplementary Table S3). Transcription-related genes were the most abundant category and were predominantly core genes, indicating strong conservation of transcriptional machinery across strains. Carbohydrate and lipid transport and metabolism were likewise highly conserved, reflecting the essential capacity of *L. crispatus* to utilize glycogen and maintain membrane integrity. Shell and cloud genes in these categories may reflect strain-specific adaptations to alternative carbohydrate sources. In contrast, genes associated with defence mechanisms, prophages, and transposases were predominantly found in the shell and cloud fractions, indicating high variability across strains.

Intra-individual strains cluster together yet differ by up to 123 genes

Hierarchical clustering based on gene presence-absence of functionally annotated genes revealed that strains from the same individual generally clustered together, indicating high intra-

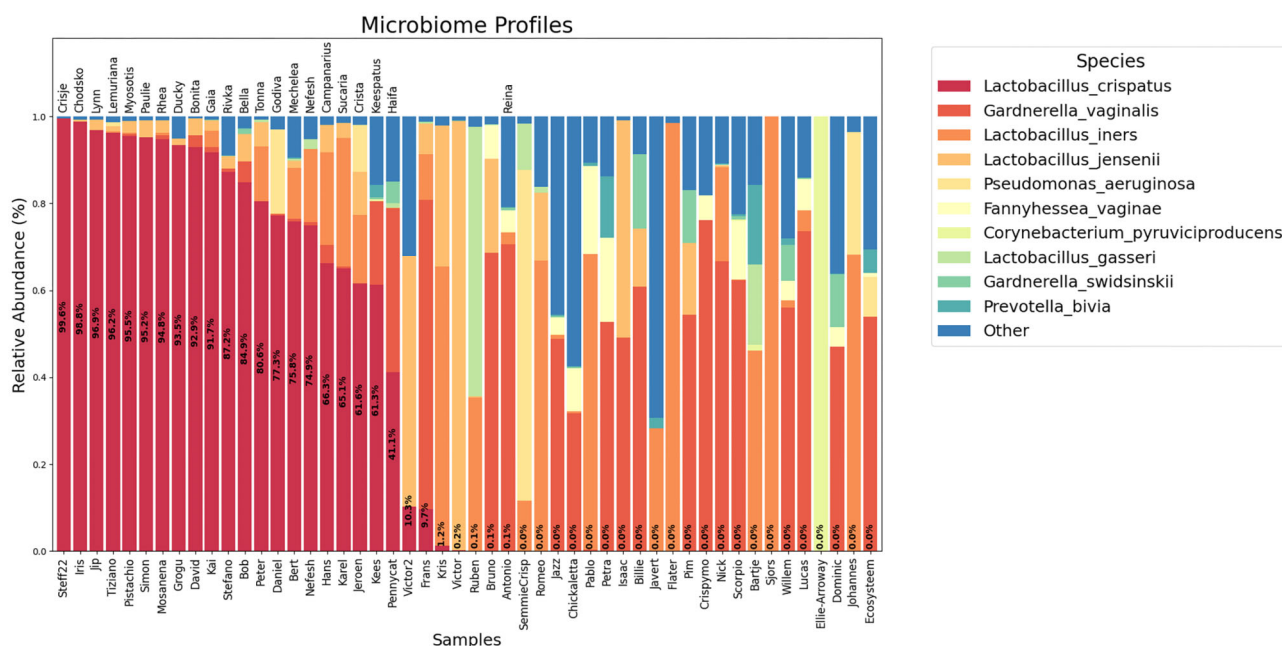


Figure 1 Stacked bar chart of the bacterial microbiome profiles from 48 metagenomic samples. The top 10 most frequently occurring species are colour-coded; remaining species are grouped as 'Other'. *Lactobacillus crispatus* relative abundance is annotated for each sample. Samples are ordered by decreasing *L. crispatus* abundance; the first 22 samples (left) originate from participants from whom *L. crispatus* strains were isolated.

individual genetic similarity (Fig. 4; [Supplementary Table S4](#)). Nevertheless, intra-individual variation ranged from 1 to 123 gene differences.

Inter-individual differences were generally larger. The strain Haifa exhibited the greatest genetic divergence from all other isolates. Conversely, some inter-individual distances were small: the Bella strains clustered closely with Nefesh, Rivka, Lynn, and Tonna (67–193 gene differences), demonstrating that intra-individual variation can in some cases exceed inter-individual variation. Among the variable genes shared across all intra-individual comparisons, two transposase genes were consistently present. Strains from closely related clusters (Bella, Nefesh, and Rivka) shared additional variable genes, including four transposases, a SagB-type dehydrogenase domain protein, a GAF domain-containing protein, and an ABC-2 type transporter.

Functional traits vary within individuals

The presence and variation of potential probiotic traits and safety markers across all 53 strains are summarised in Fig. 5. The capacity for glycogen utilisation was predicted from the pullulanase type I gene, a close homologue of *glgX* that is essential for utilising glycogen, the main carbon source in the human vaginal tract. Of the 53 strains, 40 did not contain any deletions in the N-terminus and were predicted to grow on glycogen, 6 harboured partial deletions with unknown consequences for growth on glycogen, and 7 were predicted to lack glycogen-degrading capacity due to extensive deletions or absence of the gene entirely ([Supplementary Table S5](#)). Notably, intra-individual variation in this gene was observed: among the Bella strains, one harboured a full-length (functional) pullulanase while the remainder carried N-terminal deletions predicting loss of capacity for extracellular

glycogen-degrading capacity. Among the Rhea strains, three were predicted to grow on glycogen, while four were not. The three remaining strains predicted to lack glycogen growth presumably depend on alternative carbon sources or on cross-feeding by co-occurring glycogen-degrading strains within the same microbiome.

Four adhesin-associated gene types were identified: LPxTG motif-containing proteins, MucBP domain-containing proteins, type IV pilus-like proteins, and putative adhesins. All genomes harboured at least three of these types; MucBP domain-containing proteins were detected in only four strains.

All 53 strains carried bacteriocin-encoding genes. The class III bacteriocin helveticin J was universally present, consistent with previous reports (Mancabelli *et al.* 2021). Class IIa bacteriocins were identified in 41 strains, class IIb in 38, and class I in fewer than half of the strains ($n = 18$). Intra-individual variation was observed primarily in class IIa and IIb bacteriocins, with one strain predicted negative and others positive within the same individual; variation in class I was limited to a single case. This variation may affect the breadth of pathogen species that can be inhibited by bacteriocin production, with individuals harbouring multiple strains potentially having a broader antimicrobial repertoire than those with a single strain.

CRISPR-Cas analysis revealed that all strains carried the Cas3 helicase, but only the strain Haifa contained the complete type I-E system. The remaining 52 strains were classified as type II-A based on the presence of Cas9 and Csn2 and the absence of Cas4. The subtypes CAS-V-A and CAS-VI-B were conserved across strains, while CAS-II-C was the only subtype that varied both between and within individuals.

Three GT fragments, hypothesized to modulate cell-surface glycosylation, were characterized. The first and longest fragment

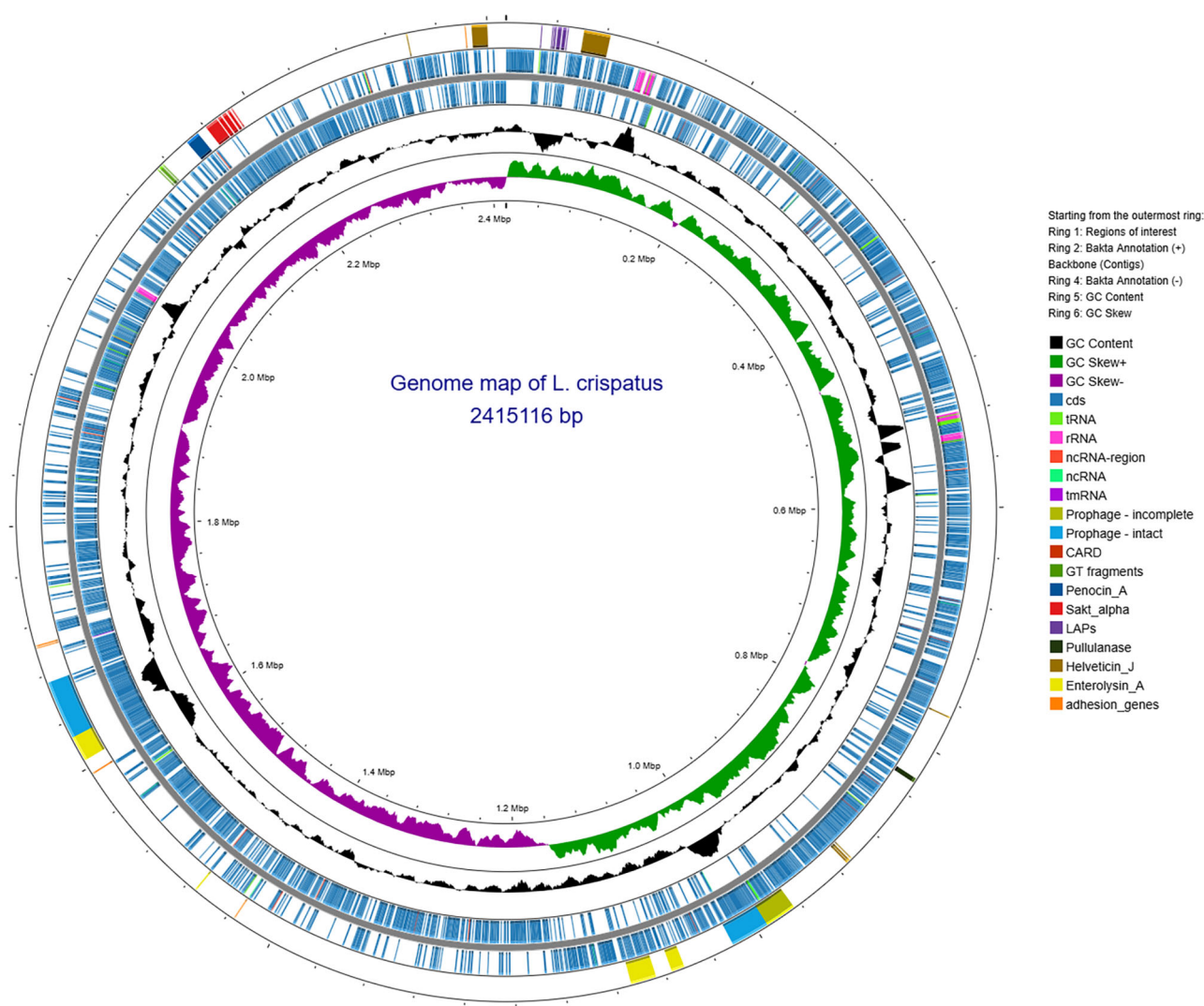


Figure 2 Genome map of *L. crispatus* constructed from the long-read-based assembly of isolate Bella (2,415,116 bp). From the outermost ring inward: regions of interest (potential probiotic traits analysed in this study), Bakta annotation of the forward (+) strand, backbone contigs, Bakta annotation of the reverse (–) strand, GC content, and GC skew. Colour coding of the outermost ring indicates the genomic positions of prophage regions, GT fragments, bacteriocin-encoding genes, the pullulanase locus, and adhesion-related genes.

(GT1) shares close homology with a known GTA and is suspected to contain the catalytic active site. Twenty-nine strains harboured GT fragments 1 and 2 together with GTA and GTB (one strain lacked GTB), while the remaining 24 strains lacked all GT-associated genes. No intra-individual variation in GT fragment presence was observed. A 14 kb plasmid was identified in 44 of 53 strains (Supplementary Fig. S2). The plasmid encoded a replication initiator protein, a plasmid partition protein (ParB), and an ATPase, indicating the capacity for autonomous replication and stable maintenance through vertical transfer. It also harboured ribonucleoside-diphosphate reductase subunits alpha and beta, multiple transposases, and a filamentation induced by cAMP (Fic) domain protein, which is widely associated with toxin-antitoxin systems in bacteria. Plasmid presence was concordant within individuals: all strains from the same participant either carried or lacked the plasmid, consistent with stable vertical maintenance potentially mediated by a toxin-antitoxin system (Fig. 5). Neither the plasmid nor the chromosomal genomes contained antimicrobial resistance or virulence genes.

Individual identity explains the majority of trait variation

Among the five individuals with multiple strains (Bella, $n = 7$; Campanarius, $n = 5$; Nefesh, $n = 9$; Rhea, $n = 7$; Rivka, $n = 8$), CAS-II-C varied in three individuals (60%), glycogen growth prediction, bacteriocin class I, and bacteriocin class IIb each varied in two individuals (40%), and bacteriocin class IIa varied in one individual (20%). In contrast, plasmid presence, GT fragment composition, and adhesion-related genes showed perfect concordance within individuals (0/5 varying), consistent with stable vertical inheritance. Intra-individual Hamming distances ($n = 116$ pairs) averaged 0.68 trait differences (median 1.0; range 0–3), whereas inter-individual distances ($n = 1262$ pairs) averaged 3.88 trait differences (median 4.0; range 0–11). This difference was highly significant (one-sided Mann-Whitney $U = 9\,194$, $P = 2.67 \times 10^{-56}$). PERMANOVA confirmed that individual identity explained the majority of trait variation (pseudo- $F = 38.9$; $P = 0.0001$). The effect size rank-biserial correlation $r = 0.87$.

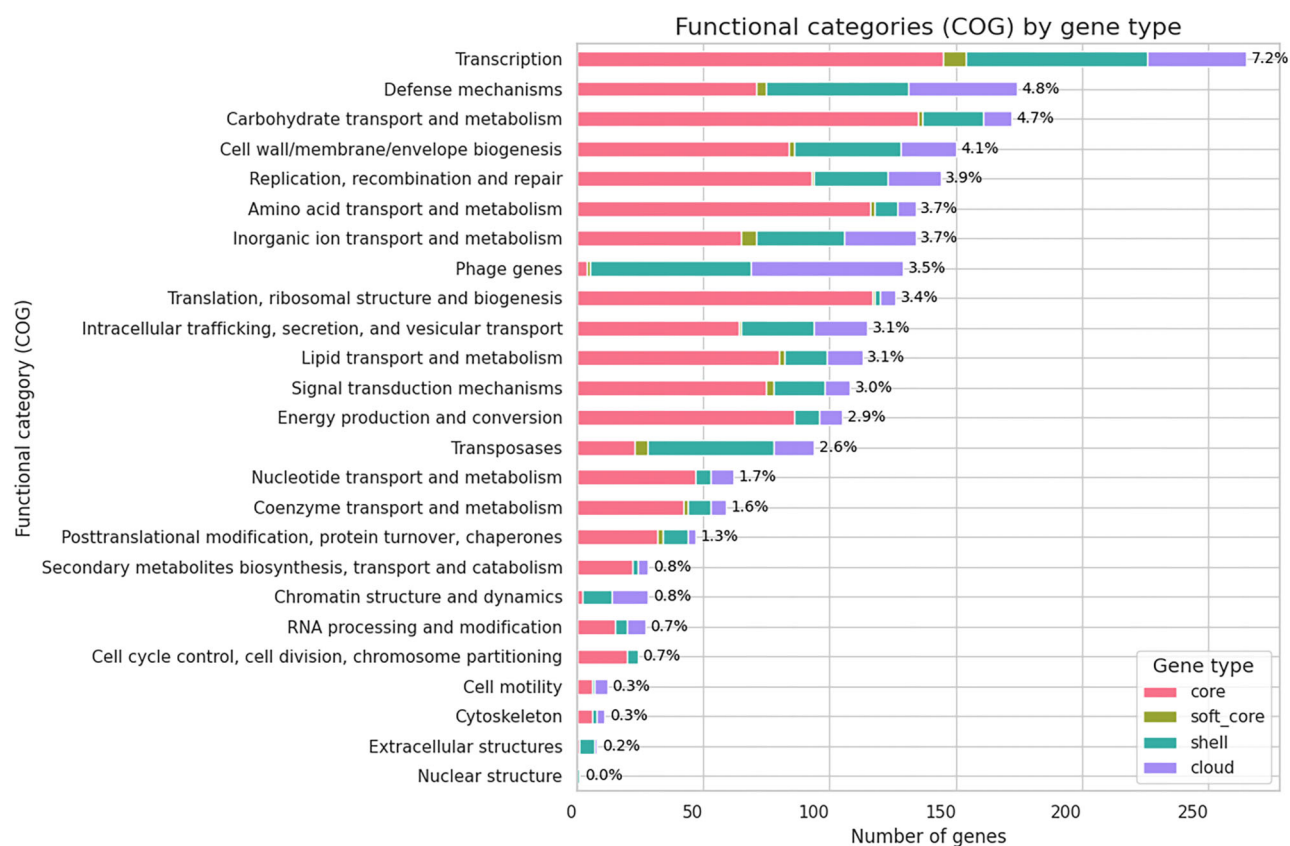


Figure 3 Functional categories (COG) of the *L. crispatus* pangenome by gene type. Approximately 62.1% of predicted proteins were assigned to functional categories based on the EggNOG database and keyword matching (Supplementary Table S2). Functional categories are listed in descending frequency. Bars indicate core, soft-core, shell, and cloud genes. Prophage and transposase categories were added based on keyword matching.

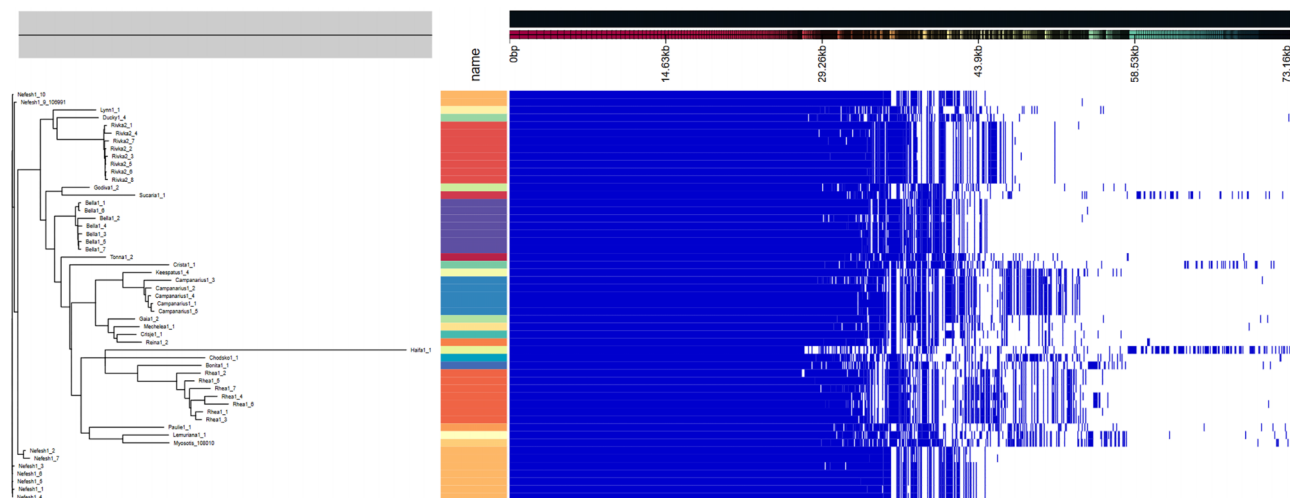


Figure 4 Phylogenetic tree and block alignment of genetic inter- and intra-individual differences among 53 *L. crispatus* strains. The phylogenetic tree (left) was constructed using the neighbour-joining algorithm from pangenome gene presence–absence data and visualized with Phandango. In the block alignment (right), gene presence is indicated in blue and absence in white. Genes are sorted from most commonly shared (left) to least commonly shared (right). Strains are colour-coded by donor (middle column).

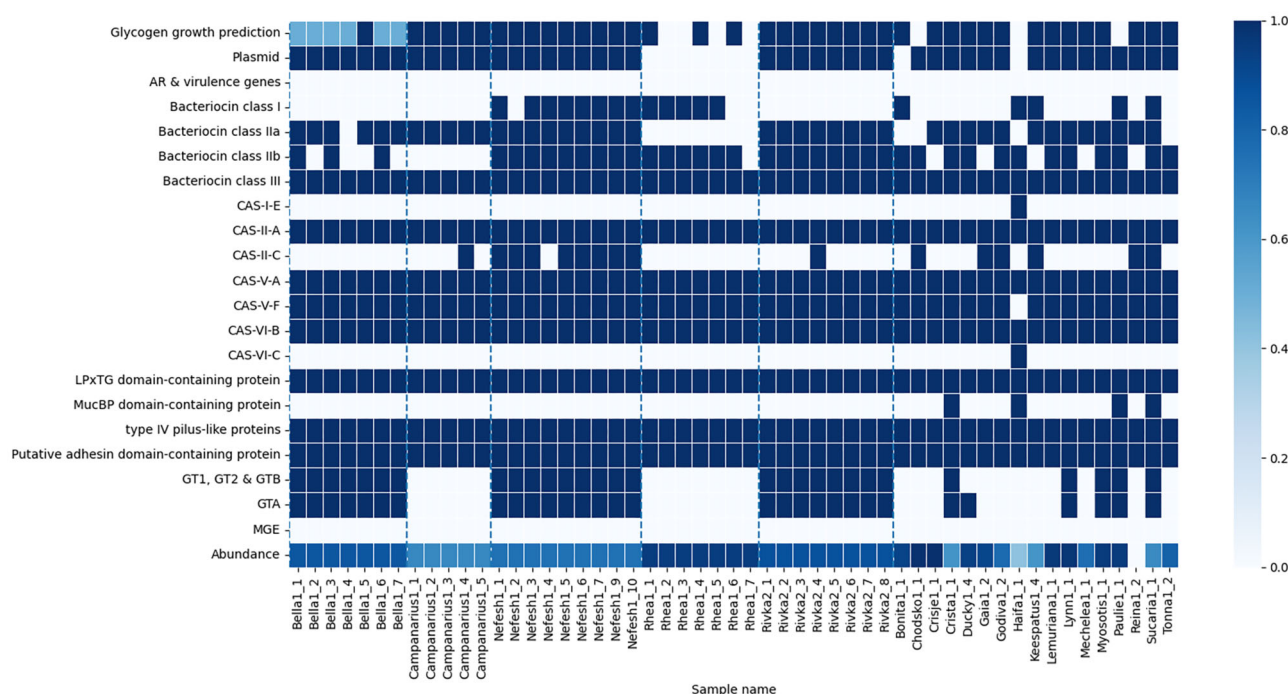


Figure 5 Heatmap of potential probiotic traits and safety assessment of the 53 *L. crispatus* strains. Traits are listed on the y-axis; each column represents an individual strain on the x-axis. Dashed vertical lines separate isolates from different individuals. From top to bottom: glycogen growth prediction (dark fill = full growth, light fill = unknown growth, unfilled = no growth); plasmid presence; antimicrobial resistance (AR) and virulence gene absence; bacteriocin classes I, IIa, IIb, and III; CRISPR-Cas subtypes (CAS-I-E through CAS-VI-C); adhesion-related genes (LPxTG, MucBP, type IV pilus-like, and putative adhesin domain-containing proteins); glycosyltransferase fragments (GT1, GT2, GTB, GTA); mobile genetic elements (MGE); and *L. crispatus* relative abundance in the corresponding metagenome.

Discussion

This study presents the first systematic characterization of intra-individual genetic diversity in *L. crispatus* using multiple isolates from the same vaginal swab, rather than relying on metagenome-assembled genomes. The pangenome of 53 strains from 22 individuals confirmed that *L. crispatus* possesses an open pangenome with a substantial accessory genome (56.3%), in agreement with previous observations (Pan *et al.* 2020, Fontana *et al.* 2021). Critically, strains from the same individual consistently clustered together yet harboured functionally relevant differences in traits such as glycogen degradation and bacteriocin repertoires.

The observation of intra-individual variation in the pullulanase type I gene is particularly noteworthy. The coexistence of glycogen-degrading and non-degrading strains within a single host is consistent with metabolic niche partitioning through cross-feeding, a mechanism well-documented in other bacterial systems (Rosenzweig *et al.* 1994, D'Souza *et al.* 2018). In this model, primary degraders hydrolyse vaginal glycogen into maltose and maltodextrins, while secondary consumers utilize these breakdown products. Such producer-consumer relationships can stabilize strain coexistence by reducing direct competition for the same substrate. The statistical observation that intra-individual distances were significantly smaller than inter-individual distances supports a model in which strains diversify from a common founder while occupying complementary metabolic niches. Notably, this intra-individual pullulanase variation was not detected in our previous metagenome-based analysis of over 1500 vaginal communities, where only a single pullulanase variant was

observed per sample (Hertzberger *et al.* 2022), illustrating how culture-based isolation of minority strains can reveal diversity that is masked in metagenomic data dominated by the most abundant genotype.

The intra-individual variation in bacteriocin classes I, IIa, and IIb, contrasted with the universal presence of class III (helveticin J), suggests a division of labour in antimicrobial defence. Different bacteriocin classes target distinct pathogen species, and individuals harboring strains with complementary repertoires may have broader collective pathogen coverage than those with a single strain type. This pattern suggests functional complementarity among co-resident strains, where genetic diversity in antimicrobial repertoires may collectively broaden pathogen coverage (Ackermann 2015).

The observation that CAS-II-C was the only CRISPR-Cas subtype showing intra-individual variation reflects its intermediate population frequency (30.2%) rather than unique biological properties. Subtypes at or near fixation (CAS-II-A, CAS-V-A, CAS-V-F, CAS-VI-B) or near absence (CAS-I-E, CAS-VI-C) lack the population-level polymorphism necessary to manifest within-host variation. The type II-A system, universally present across all 53 strains, likely represents the core adaptive immune machinery, while CAS-II-C may represent a more dynamic element subject to ongoing acquisition or loss in response to local phage pressure. CRISPR spacer diversity at the population level is predicted to enhance collective immunity by preventing phage escape mutants from sweeping through the entire community (Childs *et al.* 2012, Watson *et al.* 2021).

The identification of a 14 kb plasmid in 44 of 53 strains, devoid of antimicrobial resistance and virulence genes, provides im-

portant safety data for probiotic development. The universal absence of such determinants across both chromosomal genomes and the plasmid is consistent with the non-pathogenic profile of *L. crispatus* (van der Veer *et al.* 2019). The plasmid-encoded Fic domain protein is of particular interest, as Fic proteins in bacteria frequently function as the toxin component of type II toxin-antitoxin modules that contribute to plasmid stability through post-segregational killing (Roy and Cherfils 2015). Whether this Fic protein similarly promotes stable plasmid maintenance in *L. crispatus*, and whether it contributes to competitive fitness in the vaginal niche, warrants functional investigation.

The inter-strain variation in GT fragments is of relevance. The GT gene cluster, encoding components of a cell-wall anchored polysaccharide pellicle, has been proposed to mediate both bacteriophage sensitivity and host immune evasion (van der Veer *et al.* 2019). The observation that strains naturally differ in GT fragment composition, with 29 strains harbouring GT1-2 and 24 lacking all GT-associated genes, indicates that strains within this species naturally differ in surface polysaccharide architecture.

The perfect concordance of plasmid presence and GT fragment composition within individuals (0/5 multi-strain individuals showing variation) contrasts sharply with the variable traits and points to distinct evolutionary dynamics. For the plasmid, the Fic-domain-mediated toxin-antitoxin system discussed above likely ensures stable vertical inheritance. For GT fragments, the lack of intra-individual variation despite substantial inter-individual polymorphism (54.7% prevalence) suggests strong purifying selection within hosts, possibly driven by the functional importance of cell-surface glycosylation in host-microbe interactions or phage defense.

The substantial inter-individual variation (pseudo- $F = 38.9$) indicates that each woman's *L. crispatus* population represents a unique genomic constellation, while the functionally significant intra-individual variation documented above points to metabolic complementarity among co-resident strains. Together, these patterns support a multi-strain probiotic strategy on two grounds: strain diversity increases host compatibility, and complementary strains may cooperate through metabolic exchange and collective pathogen defence, as occurs naturally within individuals. This rationale aligns with recent clinical evidence that multi-strain *L. crispatus*-based products can durably reshape the vaginal microbiome. A multi-strain synbiotic converted vaginal communities to CST-I (Ravel *et al.* 2025), and a phase 1 randomized trial of two multi-strain *L. crispatus* live biotherapeutics achieved metagenomic colonization in 66% of women with bacterial vaginosis after only 3–7 days of dosing, with engraftment persisting to twelve weeks in approximately half of early colonizers (Potloane *et al.* 2026). *In vitro* work further supports host-independent synergism between vaginal lactobacilli (Vander Donck *et al.* 2025).

An important consideration is that the goal of a vaginal probiotic is to restore a protective community and displace pathogens, not merely adding more of a species that may already be present (Cohen *et al.* 2020, Argentini *et al.* 2022). Supplementation with well-characterized exogenous strains can nonetheless be valuable even in women who already carry *L. crispatus*: endogenous populations are frequently lost after antibiotic treatment or other disturbances, and a defined strain with a known trait profile may improve the resilience and reproducibility of recolonisation (Ojala *et al.* 2014, Hemmerling *et al.* 2024).

In a microbiome dominated by *L. iners* or in dysbiosis, an introduced *L. crispatus* strain must compete for adhesion sites and nutrients and displace an established community, which is more demanding than colonising an empty niche; the adhesion, bacteriocin, and glycogen-utilization traits catalogued here are relevant to that capacity. Our analysis was genomic and did not include functional displacement or competition assays, which are an important direction for future work. Likewise, the ability to revive rapidly from a dried or lyophilized state, is a relevant property that we did not measure (Juárez Tomás *et al.* 2015).

Finally, women with an *L. iners*-dominated microbiome carrying subdominant *L. crispatus* were not separately sampled, so we cannot directly address why a resident strain may fail to predominate in such individuals (Argentini *et al.* 2022).

A distinctive feature of this study is the citizen science framework through which the strains were collected. Participants not only performed their own sampling, enrichment, and LAMP-based identification (Illidge *et al.* 2024), but also retained ownership of their strains through the *crispatus* foundation, which collectively governs future licensing decisions. The genomic results reinforce this ownership model by demonstrating that each participant's strains represent a unique microbial fingerprint; they are not generic or interchangeable but intimately connected to the host. The rights to these strains are centralized in the foundation, which represents the participants in negotiating future licence agreements. Financial proceeds from such agreements are reinvested in citizen science research into sexual and reproductive health and shared directly with the contributing women. The stakeholder hackathon extended this participatory model to product development, yielding insights into end-user preferences that will guide the formulation and delivery of future probiotic products.

Several limitations should be acknowledged. Taxonomic profiling relied on k-mer-based classification of 150 bp short reads, an approach with well-documented constraints. Short reads carry limited phylogenetic information, so reads originating from conserved or horizontally acquired regions can be assigned to the wrong species, and false-positive assignments to taxa absent from the sample are common, in particular among closely related congeners such as *L. crispatus*, *L. jensenii*, and *L. gasseri* (Ye *et al.* 2019). Classification accuracy further depends on the composition and completeness of the reference database and on the confidence threshold applied (Wright *et al.* 2023), and several vaginal taxa remain under-represented in public reference collections. Although Bracken was used to redistribute reads and to correct for genome- and read-length effects, the relative abundances reported here should be interpreted as estimates rather than exact proportions, and no strain-level inference was drawn from the short-read profiles; all strain-level analyses were based on the isolate genomes. Reassuringly, the metagenomic assignments were consistent with culture, as *L. crispatus* was isolated from 20 of the 22 participants whose metagenomes were classified as *L. crispatus*-dominated.

The study relied on draft genome assemblies from short-read sequencing for most strains, which may result in fragmented assemblies and incomplete resolution of repetitive regions. Long-read sequencing was performed for only four strains from a single individual; expanding this to additional participants would improve genome completeness. The correlation analyses between traits and *L. crispatus* abundance were constrained by the fact that nearly all isolates originated from *L. crispatus*-dominated

metagenomes, limiting the statistical power to detect associations. Moreover, the observed diversity may underrepresent divergent strains present at low abundance in non-*L. crispatus*-dominated communities. The pangenome and trait distributions reported here should therefore be interpreted as representative of dominant *L. crispatus* populations rather than the full strain-level diversity within the vaginal niche. Notably, GT3, which has been associated with low-abundant *L. crispatus* strains in a dysbiotic environment (van der Veer *et al.* 2019), was not observed at all in the present study reflecting the isolation bias towards health-associated, dominant *L. crispatus* populations. More broadly, the *in-silico* nature of the trait predictions presented here remains a limitation; experimental validation of glycogen degradation predictions, e.g., through growth assays on glycogen versus glucose for strains with contrasting pullulanase genotypes, would provide essential phenotypic confirmation of the current classifications. However, it should be noted that previous work established the presence of the *pulA* gene as a condition for the ability to grow on glycogen (van der Veer *et al.* 2019, Hertzberger *et al.* 2022).

Two further points deserve mention. First, *L. crispatus* was isolated almost exclusively from women of 45 years or younger whose microbiome was dominated by this species. Because oestrogen promotes glycogen accumulation in the vaginal epithelium (Farage and Maibach 2006, Mirmonsef *et al.* 2015), glycogen availability and *L. crispatus* colonization tend to decline with age, so strains that are less dependent on glycogen may be more relevant for peri- and post-menopausal women; such strains were not specifically represented in our sampling. Second, some *L. crispatus* strains can produce biogenic amines (Puebla-Barragan *et al.* 2021), which we did not assess here; screening for amino-acid decarboxylase genes combined with phenotypic assays would be a useful addition in future safety evaluations.

Future work should explore the interactions between *L. crispatus* strains and other species in the vaginal microbiome, as well as the relative abundance of co-existing *L. crispatus* strains within individuals using strain-resolved metagenomic approaches. Phenotypic characterisation of the predicted traits, coupled with *in vitro* competition assays, will be essential to guide strain selection for multi-strain probiotic products targeting vaginal health.

In conclusion, this study combined metagenomic and pangenomic approaches to reveal substantial functional diversity within *L. crispatus*, even among strains from the same individual. These findings contribute to the rational design of multi-strain vaginal probiotics and demonstrate the power of citizen science in advancing microbiome research while ensuring that participants retain ownership of their microbial strains.

Acknowledgements

We gratefully acknowledge the 53 citizen scientists represented by the *crispatus* foundation whose participation in self-sampling, culturing, and isolating their own *Lactobacillus crispatus* strains made this research possible. Artificial intelligence (AI) tools were used during the preparation of this manuscript. Specifically, Claude (Anthropic) was used to assist with language editing and manuscript review. The authors reviewed and edited all AI-assisted output and take full responsibility for the content of the published article.

Author contributions

Freek de Kreek (Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft), Rosanne Hertzberger (Conceptualization, Funding acquisition, Investigation, Supervision, Writing – review & editing), Fleur van Eeden (Project administration, Supervision, Visualization), Shadelice Illidge (Investigation, Methodology), Evert-Jan Teunis (Investigation, Methodology), Mark Hanemaaijer (Data curation, Formal Analysis), Elke Lievens (Project administration), Fanny Rienstra (Project administration), Deniz Er Wiedhaup (Project administration), Paola Lisotto (Data curation, Formal.docx Analysis), Derek Butler (Project administration), Douwe Molenaar (Supervision), Remco Kort (Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing)

Supplementary material

Supplementary material is available at *Journal of Applied Microbiology* online.

Conflicts of interest

R. Hertzberger is chair of the non-profit organisation Stichting *crispatus*. R. Kort is an inventor on a patent application related to multi-strain *Lactobacillus crispatus* formulations (WO 2025/242814). F. van Eeden is an employee of MyMicroZoo (Leiden, The Netherlands). E.-J. Teunis, M. Hanemaaijer, and E. Lievens are employees of Winclove B.V. (Amsterdam, The Netherlands), a probiotics company. F. Rienstra and D. Er Wiedhaup are employees of Yoni (Amsterdam, The Netherlands), a femcare product company. P. Lisotto and D. Butler are employees of BaseClear B.V. (Leiden, The Netherlands), a company providing sequencing and bioinformatics services. F. de Kreek, S. Illidge, and D. Molenaar declare no competing interests.

Funding

This project received support from the ZonMw (The Netherlands Organisation for Health Research and Development) and by Health~Holland, Top Sector Life Sciences and Health project Citizen Science for Health Care (project number 404270598028). The funders had no role in study design, data collection, analysis, interpretation, or the decision to submit for publication.

Data availability

Metagenomic and whole-genome sequencing data generated in this study have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive under BioProject accession PRJNA1267825. The data are released upon publication. Custom scripts used for pangenome visualisation (PAN-DAVIS.py), functional categorisation (functional_categorizer.py), and downstream analyses are available at <https://github.com/FreekdeKreek/LCPA>.

References

- Ackermann M. A functional perspective on phenotypic heterogeneity in microorganisms. *Nat Rev Micro* 2015;**13**:497–508. <https://doi.org/10.1038/nrmicro3491>
- Argentini C, Fontana F, Alessandri G *et al.* Evaluation of Modulatory Activities of *Lactobacillus crispatus* Strains in the Context of the Vaginal Microbiota. *Microbiol Spectr* 2022;**10**:e0273321.
- Bankevich A, Nurk S, Antipov D *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;**19**:455–77. <https://doi.org/10.1089/cmb.2012.0021>
- Bayliss SC, Thorpe HA, Coyle NM *et al.* PIRATE: a fast and scalable pangenomics toolbox for clustering diverged orthologues in bacteria. *Gigascience* 2019;**8**:giz119. <https://doi.org/10.1093/gigascience/giz119>
- Bharat A, Petkau A, Avery BP *et al.* Correlation between phenotypic and in silico detection of antimicrobial resistance in *Salmonella enterica* in Canada using Staramr. *Microorganisms* 2022;**10**:292. <https://doi.org/10.3390/microorganisms10020292>
- Bushnell B. BBMap: a fast, accurate, splice-aware aligner. 2014. <https://sourceforge.net/projects/bbmap/> (26 July 2026, date last accessed).
- Carattoli A, Hasman H. PlasmidFinder and in Silico pMLST: identification and typing of plasmid replicons in whole-genome sequencing (WGS). *Methods Mol Biol* 2020;**2075**:285–94. https://doi.org/10.1007/978-1-4939-9877-7_20
- Childs LM, Held NL, Young MJ *et al.* Multiscale model of CRISPR-induced coevolutionary dynamics: diversification at the interface of Lamarck and Darwin. *Evolution* 2012;**66**:2015–29. <https://doi.org/10.1111/j.1558-5646.2012.01595.x>
- Cohen CR, Wierzbicki MR, French AL *et al.* Randomized trial of lactin-V to prevent recurrence of bacterial vaginosis. *N Engl J Med* 2020;**382**:1906–15. <https://doi.org/10.1056/NEJMoa1915254>
- Croatti V, Dricot C, Eilers T *et al.* Exopolysaccharides of *Lactobacillus crispatus* mediate key balancing interactions with the vaginal mucosa. *npj Biofilms Microbiomes* 2026;**12**:70. <https://doi.org/10.1038/s41522-026-00937-5>
- Danecek P, Bonfield JK, Liddle J *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* 2021;**10**:giab008. <https://doi.org/10.1093/gigascience/giab008>
- Dong MJ, Luo H, Gao F. Ori-Finder 2022: a comprehensive web server for prediction and analysis of bacterial replication origins. *Genom Proteom Bioinform* 2022;**20**:1207–13. <https://doi.org/10.1016/j.gpb.2022.10.002>
- D'Souza G, Shitut S, Preussger D *et al.* Ecology and evolution of metabolic cross-feeding interactions in bacteria. *Nat Prod Rep* 2018;**35**:455–88. <https://doi.org/10.1039/C8NP00009C>
- Ewels P, Magnusson M, Lundin S *et al.* MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016;**32**:3047–8. <https://doi.org/10.1093/bioinformatics/btw354>
- Farage M, Maibach H. Lifetime changes in the vulva and vagina. *Arch Gynecol Obstet* 2006;**273**:195–202. <https://doi.org/10.1007/s00404-005-0079-x>
- Fontana F, Alessandri G, Lugli GA *et al.* Probiogenomics analysis of 97 *Lactobacillus crispatus* strains as a tool for the identification of promising next-generation probiotics. *Microorganisms* 2021;**9**:73. <https://doi.org/10.3390/microorganisms9010073>
- France M, Alizadeh M, Brown S *et al.* Towards a deeper understanding of the vaginal microbiota. *Nat Microbiol* 2022;**7**:367–78. <https://doi.org/10.1038/s41564-022-01083-2>
- Hadfield J, Croucher NJ, Goater RJ *et al.* Phandango: an interactive viewer for bacterial population genomics. *Bioinformatics* 2018;**34**:292–3. <https://doi.org/10.1093/bioinformatics/btx610>
- Hemmerling A, Wierzbicki MR, Armstrong E *et al.* Response to antibiotic treatment of bacterial vaginosis predicts the effectiveness of LACTIN-V (*Lactobacillus crispatus* CTV-05) in the prevention of recurrent disease. *Sexual Trans Dis* 2024;**51**:437–40. <https://doi.org/10.1097/OLQ.0000000000001962>
- Hertzberger R, May A, Kramer G *et al.* Genetic elements orchestrating *Lactobacillus crispatus* glycogen metabolism in the vagina. *IJMS* 2022;**23**:5590. <https://doi.org/10.3390/ijms23105590>
- Huerta-Cepas J, Szklarczyk D, Heller D *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 2019;**47**:D309–14. <https://doi.org/10.1093/nar/gky1085>
- Illidge S, Kort R, Hertzberger R *et al.* 'From women for women': a citizen science approach engaging women in the isolation and application of the vaginal health-associated bacterium *Lactobacillus crispatus*. *PLoS One* 2024;**19**:e0308526. <https://doi.org/10.1371/journal.pone.0308526>
- Jenkins DJ, Woolston BM, Hood-Pishchany MI *et al.* Bacterial amy-lases enable glycogen degradation by the vaginal microbiome. *Nat Microbiol* 2023;**8**:1641–52. <https://doi.org/10.1038/s41564-023-01447-2>
- Juárez Tomás MS, De Gregorio PR, Leccese Terraf MC *et al.* Encapsulation and subsequent freeze-drying of *Lactobacillus reuteri* CRL 1324 for its potential inclusion in vaginal probiotic formulations. *Eur J Pharm Sci* 2015;**79**:87–95. <https://doi.org/10.1016/j.ejps.2015.08.010>
- Kumariya R, Garsa AK, Rajput YS *et al.* Bacteriocins: classification, synthesis, mechanism of action and resistance development in food spoilage causing bacteria. *Microb Pathog* 2019;**128**:171–7. <https://doi.org/10.1016/j.micpath.2019.01.005>
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;**9**:357–59. <https://doi.org/10.1038/nmeth.1923>
- Lebeer S, Ahannach S, Gehrmann T *et al.* A citizen-science-enabled catalogue of the vaginal microbiome and associated factors. *Nat Microbiol* 2023;**8**:2183–95. <https://doi.org/10.1038/s41564-023-01500-0>
- Liu B, Zheng D, Zhou S *et al.* VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res* 2022;**50**:D912–7. <https://doi.org/10.1093/nar/gkab1107>
- Lu J, Breitwieser FP, Thielen P *et al.* Bracken: estimating species abundance in metagenomics data. *PeerJ Comput Sci* 2017;**3**:e104. <https://doi.org/10.7717/peerj-cs.104>
- Ma B, France MT, Crabtree J *et al.* A comprehensive non-redundant gene catalog reveals extensive within-community intraspecies diversity in the human vagina. *Nat Commun* 2020;**11**:940. <https://doi.org/10.1038/s41467-020-14677-3>
- Mancabelli L, Mancino W, Lugli GA *et al.* Comparative genome analyses of *Lactobacillus crispatus* isolated from different ecological niches reveal an environmental adaptation of this species to the human vaginal environment. *Appl Environ Microb* 2021;**87**:e02899–20. <https://doi.org/10.1128/AEM.02899-20>

- Mirmonsef P, Modur S, Burgad D *et al.* Exploratory comparison of vaginal glycogen and *Lactobacillus* levels in premenopausal and postmenopausal women. *Menopause* 2015;**22**:702–9. <https://doi.org/10.1097/GME.0000000000000397>
- Nori SRC, Walsh CJ, McAuliffe FM *et al.* Strain-level variation among vaginal *Lactobacillus crispatus* and *Lactobacillus iners* as identified by comparative metagenomics. *npj Biofilms Microbiomes* 2025;**11**:49. <https://doi.org/10.1038/s41522-025-00682-1>
- Ojala T, Kankainen M, Castro J *et al.* Comparative genomics of *Lactobacillus crispatus* suggests novel mechanisms for the competitive exclusion of *Gardnerella vaginalis*. *Bmc Genom [Electronic Resource]* 2014;**15**:1070. <https://doi.org/10.1186/1471-2164-15-1070>
- Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 2016;**32**:292–4. <https://doi.org/10.1093/bioinformatics/btv566>
- Padilha VA, Alkhnbashi OS, Shah SA *et al.* CRISPRcasIdentifier: machine learning for accurate identification and classification of CRISPR-Cas systems. *Gigascience* 2020;**9**:giaa062. <https://doi.org/10.1093/gigascience/giaa062>
- Pan M, Hidalgo-Cantabrana C, Barrangou R. Host and body site-specific adaptation of *Lactobacillus crispatus* genomes. *NAR Genom Bioinform* 2020;**2**:lqaa001. <https://doi.org/10.1093/nargab/lqaa001>
- Potloane D, Symul L, Ngcapu S *et al.* VIBRANT: a phase 1 randomized trial of multi-strain vaginal *L. crispatus* live biotherapeutic products in people with bacterial vaginosis. *Cell Host & Microbe* 2026;**34**:S1931312826000843. <https://doi.org/10.1016/j.chom.2026.02.016>
- Puebla-Barragan S, Watson E, van der Veer C *et al.* Interstrain variability of human vaginal *Lactobacillus crispatus* for metabolism of biogenic amines and antimicrobial activity against urogenital pathogens. *Molecules* 2021;**26**:4538. <https://doi.org/10.3390/molecules26154538>
- Ravel J, Gajer P, Abdo Z *et al.* Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci USA* 2011;**108**:4680–7. <https://doi.org/10.1073/pnas.1002611107>
- Ravel J, Simmons S, Jaswa EG *et al.* Impact of a multi-strain *L. crispatus*-based vaginal synbiotic on the vaginal microbiome: a randomized placebo-controlled trial. *npj Biofilms Microbiomes* 2025;**11**:158. <https://doi.org/10.1038/s41522-025-00788-6>
- Rosenzweig RF, Sharp RR, Treves DS *et al.* Microbial evolution in a simple unstructured environment: genetic differentiation in *Escherichia coli*. *Genetics* 1994;**137**:903–17. <https://doi.org/10.1093/genetics/137.4.903>
- Roy CR, Cherfils J. Structure and function of Fic proteins. *Nat Rev Micro* 2015;**13**:631–40. <https://doi.org/10.1038/nrmicro3520>
- Schwengers O, Jelonek L, Dieckmann MA *et al.* Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb Genom* 2021;**7**:000685. <https://doi.org/10.1099/mgen.0.000685>
- Seemann T. Abricate: Mass Screening of Contigs for Antimicrobial and Virulence Genes. San Francisco (CA): GitHub, Inc, 2020. <https://github.com/tseemann/abicate> (18 February 2025, date last accessed).
- Suzuki MT, Taylor LT, DeLong EF. Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5'-Nuclease Assays. *Appl Environ Microb* 2000;**66**:4605–14. <https://doi.org/10.1128/AEM.66.11.4605-4614.2000>
- Tarracchini C, Argentin C, Alessandri G *et al.* The core genome evolution of *Lactobacillus crispatus* as a driving force for niche competition in the human vaginal tract. *Microb Biotechnol* 2023;**16**:1774–89. <https://doi.org/10.1111/1751-7915.14305>
- Vander Donck L, Victor M, Van Beeck W *et al.* Host-independent synergism between *Lactobacillus crispatus* and other vaginal lactobacilli. *Cell Reports* 2025;**44**:116171. <https://doi.org/10.1016/j.celrep.2025.116171>
- van der Veer C, Hertzberger RY, Bruisten SM *et al.* Comparative genomics of human *Lactobacillus crispatus* isolates reveals genes for glycosylation and glycogen degradation: implications for *in vivo* dominance of the vaginal microbiota. *Microbiome* 2019;**7**:49. <https://doi.org/10.1186/s40168-019-0667-9>
- van Heel AJ, de JA, Song C *et al.* BAGEL4: a user-friendly web server to thoroughly mine RPPs and bacteriocins. *Nucleic Acids Res* 2018;**46**:W278–81. <https://doi.org/10.1093/nar/gky383>
- Vasimuddin M, Misra S, Li H *et al.* Efficient architecture-aware acceleration of BWA-MEM for multicore systems. 2019:314–24. <https://doi.org/10.1109/IPDPS.2019.00041>
- Watson BNJ, Steens JA, Staals RHJ *et al.* Coevolution between bacterial CRISPR-Cas systems and their bacteriophages. *Cell Host Microbe* 2021;**29**:715–25. <https://doi.org/10.1016/j.chom.2021.03.018>
- Wick RR. Filtlong: Quality Filtering Tool for Long Reads. 2024. <https://github.com/rrwick/Filtlong> (8 October 2024, date last accessed).
- Wick RR, Judd LM, Gorrie CL *et al.* Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;**13**:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>
- Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol* 2019;**20**:257. <https://doi.org/10.1186/s13059-019-1891-0>
- Wright RJ, Comeau AM, Langille MGI. From defaults to databases: parameter and database choice dramatically impact the performance of metagenomic taxonomic classification tools. *Microb Genom* 2023;**9**:mgen000949. <https://doi.org/10.1099/mgen.0.000949>
- Ye SH, Siddle KJ, Park DJ *et al.* Benchmarking metagenomics tools for taxonomic classification. *Cell* 2019;**178**:779–94. <https://doi.org/10.1016/j.cell.2019.07.010>
- Zankari E, Allesøe R, Joensen KG *et al.* PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. *J Antimicrob Chemother* 2017;**72**:2764–8. <https://doi.org/10.1093/jac/dkx217>
- Zankari E, Hasman H, Cosentino S *et al.* Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 2012;**67**:2640–4. <https://doi.org/10.1093/jac/dks261>