scientific reports

Check for updates

OPEN

Comparative genomic analysis of bacteriocin genes in *Lactobacillus crispatus* strains

Giulia Santarelli[1], Maurizio Sanguinetti[1,2], Giovanni Delogu[1,3,4] & Flavio De Maio[2,4]✉

***Lactobacillus crispatus*, a key member of the vaginal microbiota, exhibits strain-specific diversity with important implications for host–microbe interactions and probiotic potential. This study investigates the genomic relationships and bacteriocin diversity across 95 *L. crispatus* strains to elucidate the mechanisms underlying strain-specific adaptation and microbial competitiveness. Using publicly available data from NCBI, we applied BAGEL4 and core genome multilocus sequence typing (cgMLST) approaches to identify and characterize bacteriocin genes, revealing a high level of conservation of major bacteriocins among strains. Analysis of nucleotide polymorphisms within bacteriocin-encoding sequences showed that these genes are generally conserved, likely due to strong purifying selection. These findings suggest that bacteriocins in *L. crispatus* are functionally conserved, underscoring their potential role in maintaining vaginal health through microbial competition and pathogen exclusion.**

Keywords *Lactobacillus crispatus*, Bacteriocins, Antimicrobial peptides, Vaginal microbiota

The human vaginal microbial community (VMC) includes a community of microorganisms that establish dynamic and complex interactions with the host. In healthy reproductive-aged women, the VMC is mainly dominated by bacteria, belonging to the *Lactobacillus* genus, that preserve the vaginal homeostasis by protecting the host mucosa from colonization and overgrowth of several pathobionts[1]. These functions are related to distinct mechanisms and including competitive exclusion[2], pH lowering[3], direct antimicrobial activity[4] and immunomodulation[5,6].

VMC appears highly variable and influenced by numerous factors: hormone levels, pregnancy, urogenital infections and drug treatments[7,8]. A decrease in lactobacilli abundance is classically associated with an increase in alpha diversity and a microbial instability that favours the onset of different pathologies[9]. The importance in maintaining a *Lactobacillus*-dominated environment is supported by numerous clinical observations and by the results obtained in clinical trials where eubiotic VMC were restored following vaginal microbiota transplantation harvested from healthy donors[10,11]. A central point of this innovative approach relies on the transfer of both microorganisms and associated compounds[12].

Recent studies have shown the presence of numerous Community State Types (CSTs) strictly associated to the prevalence of one *Lactobacillus* species[13]. *L. crispatus*, *L. gasseri*, *L. iners* and *L. jensenii* are the most prevalent *Lactobacillus* species that define the CST-I, CST-II, CST-III and CST-V, respectively[14]. These findings further highlighted that not all lactobacilli are able to perform the same protective functions or to maintain VMC homeostasis[15,16]. CST-III community, dominated by *L. iners* is mainly associated with an increased risk of acquiring vaginal infection, and also to other dysfunctions as infertility[17]. On the other hand, *L. crispatus* has been extensively described among *Lactobacillus* species to have health promoting properties for its peculiar antimicrobial activities. Dominance of *L. crispatus* species (producing D- and L-lactic acid, S-layer proteins, hydrogen peroxide and bacteriocins) prevents pathobionts overgrowth lowering the risk of upper genital tract infections[18]. For this reason, *L. crispatus* has emerged as a key component of several probiotics for both gut and vaginal environments. An example of this approach is the *L. crispatus* M247 (LcM247), isolated from infant faeces in 1989, that has revealed significant clinical potential with genetic and probiotic properties that make it suitable for restoring and maintaining a healthy VMC, especially when administrated in women with CST-IV[19]. LcM247 can adhere to vaginal epithelium, outcompete pathogens like *Gardnerella vaginalis*, and produce bacteriocins and hydrogen peroxide, enhancing its protective effects. Clinically, it can be used orally to combat

[1]Dipartimento di Scienze Biotecnologiche di Base, Cliniche Intensivologiche e Perioperatorie-Sezione di Microbiologia, Università Cattolica del Sacro Cuore, Rome 00168, Italy. [2]Dipartimento Scienze di Laboratorio ed Ematologiche, Fondazione Policlinico Universitario "A. Gemelli", IRCCS, Rome 00168, Italy. [3]Mater Olbia Hospital, Olbia 07026, Italy. [4]Giovanni Delogu and Flavio De Maio have contributed equally to this work. ✉email: flavio.demaio@unicatt.it

dysbiosis, reduce inflammation, and prevent conditions like vaginosis, preterm birth and recurrent urinary infections[20]. The mechanism by which lactobacilli are transferred from the gut to the vagina remains unclear, requiring administration of lactobacilli for 4–5 weeks or longer before observing any effect on the vaginal community[21]. LcM247 showed promise as a potential therapeutic agent against Human Papillomavirus (HPV) infections, enhancing HPV clearance and reducing the progression of related cervical lesions after long-term oral administration[22].

LcM247 properties, not matched in other *L. crispatus* isolates, suggest that different *L. crispatus* strains found in CST-I dominated communities may have differential protective properties in preserving the healthy VMC. On the other hand, the VMCs appear strictly associated with women's ethnicity, highlighting that lactobacillus dominated CST, and in particularly CST-I, cannot be universally recognized as healthy VMC and host factors may significantly contribute to maintain the vaginal homeostasis[13,23,24]. LcM247 genome, organized in one circular chromosome of 2.3 Mbp like other *L. crispatus* strains[25,26], contains a large mobilome, including the 14,105 bp Tn*7088* IME, 42,510 bp ΦM247 prophage, the CRISPR-Cas locus and 226 ISs, which accounts for 14% of the whole genome. The key genetic element has been identified in the Tn*7088* sequence, an integrative and mobilizable elements carrying a biosynthetic gene cluster for a class I bacteriocin[27]. In this study, we aim to investigate whether the genomic diversity and the polymorphisms and asset of the bacteriocin-encoding genes observed within the *L. crispatus* isolates support or not this possibility.

Materials and methods

Sequence description and bioinformatic analysis

High-throughput sequencing data were retrieved from the Sequence Read Archive (SRA), managed by the National Center for Biotechnology Information (NCBI), by searching for *L. crispatus*. The SRA Toolkit was used within the R environment (RStudio, R version version 4.3.1) to download and process the data, following the workflow described at https://rpubs.com/snijesh/sratool-kit-data-retrieval. The SRA Toolkit provides command-line utilities to download, manipulate, and convert raw sequencing data from the SRA repository into compressed FASTQ files, which were subsequently used in downstream analyses.

Paired-end sequencing data from 95 *L. crispatus* isolates were retrieved. Quality assessment of the raw reads was performed using FastQC version 0.12.0 (https://github.com/s-andrews/FastQC). Adapter trimming and quality filtering were carried out using fastp[28]. Taxonomic identification of the trimmed reads was performed prior to assembly using CLC Genomics Workbench version 25.0 (Qiagen) with the k-mer-based "best match" plugin.

De novo assembly of the quality-filtered reads was performed using Unicycler version[29] in normal mode with a minimum FASTA length threshold of 500 bp.

In addition to the assembled sequences from the 95 isolates, 24 publicly available *L. crispatus* complete genome assemblies were downloaded from the NCBI database and used to generate a core genome multilocus sequence typing (cgMLST) scheme using chewBBACA version 3.3.10 with default parameters[30]. Briefly, the chewBBACA tool was deployed using the NCBI reference sequence *L. crispatus* ST1 (reference number: GCA_000091765.1) as a training file. After the removal of 814 loci (3 containing ambiguous or invalid characters and 811 shorter than 201 bp), a total of 4193 loci were identified. Of these loci, 999 corresponded to the genes of the cgMLST (identified as present in every isolate). This cgMLST scheme was subsequently applied to assess the phylogenetic relationships among the 95 assembled genomes by constructing a minimum spanning tree.

Putative bacteriocin-encoding genes were identified in the assembled genomes using BAGEL4[31]. Each identified gene was aligned to detect single nucleotide variants (SNVs), including synonymous (dS) and non-synonymous (dNS) substitutions, as well as insertions and deletions (indels). The ratio of synonymous to non-synonymous substitutions (dS/dNS) was then calculated to assess selective pressure on these genes. All sequences used for the study were reported in Supplementary Table 1.

Results

Lactobacillus crispatus genomic differences

The emergence of next-generation sequencing technologies and the development of bioinformatic tools have significantly contributed to increase knowledge on *L. crispatus* genomics providing important information on gene content, metabolomic capabilities and antibacterial properties[32].

Table 1 resumes the major findings highlighted by whole genome sequencing of *L. crispatus* strains, which have been isolated mainly from human vagina or chicken gut and urine. The common thread of these studies is focused on the genes involved in its metabolism. Van der Veer et al. studied 33 strains isolated from the human vagina, revealing that strains associated with a healthy VMC were *L. crispatus*, showing specific deletions in gene encoding for an enzyme involved in the carbohydrate metabolism, which implied a direct glycogen utilization[32]. On the contrary, strains isolated from patients with a vaginal dysbiosis carried a fragmentated glycosyltransferase gene which appeared impairing interactions between the vaginal microbiota and the host. Zhang and colleagues analysed 37 *L. crispatus* strains highlighting a first significant phylogenetic difference among isolates derived from vagina or human gut. These strains were not only differentially clustered, but they were functionally divergent[33,34]. Genome annotation revealed that vaginal strains were enriched in genes mediating acid tolerance and carbohydrate metabolism[33]. The presence of these genes led to support the idea of an important adaptation of vaginal strains to vaginal environment that may be involved in carbohydrate utilization and host immune system interactions[35,36]. Importantly, whole genome analysis further demonstrated that some *L. crispatus* isolates may be responsible for urinary tract infections[37].

Author	Country	Method	Number of strains	Host source	Biological niche	Genome features	Major findings
van der Veer et al, Microbiome 2019	Netherlands	WGS	33	homo sapiens	vagina; 16 (3) LVM, 17 (1) DVM)	average lenght: 2.31Mbp; GC content: 36.8%; gene number: 4384; CDS: NA; core genome: 1429 genes	Strains showing less-efficient or no growth on glycogen carried N-terminal deletions (respectively, 29 and 37 amino acid deletions) in a putative pullulanase type I protein suggesting that L. crispatus can directly utilize glycogen. L. crispatus strains isolated from LVM were not phenotypically distinct from L. crispatus strains isolated from DVM. DVM associated strains were more likely to carry a 3-fragmented glycosyltransferase gene indicating a role for cell surface glycoconjugates, which may impair vaginal microbiota-host interactions.
McComb et al	USA	WGS	1	homo sapiens	vagina	average lenght: ~2.3Mbp; GC content: 37.1%; gene number: NA; CDS: 2307; core genome: 1429 genes	First complete genome of a strain isolated from a human vagina
zhang et al, Genes 2020	China	WGS	37	homo sapiens; chicken	human: vagina (17), gut (11); chicken gut (9)	average lenght: 2.09 Mbp; GC content: 36.76%; gene number: 5014; average CDS: 2110 core genome: 1250 genes	No significant difference was found in the three sources about their genomic features. Core genome based phylogenetic tree highlighted that vagina-derived L. crispatus and feces-derived strains were each clustered separately suggesting that the environmental niche exerts an important impact on L. crispatus evolution. According to gene annotation, the L. crispatus derived from the vagina possess a high abundance of genes related to acid tolerance, redox reactions, pullulanase and carbohydrate-binding modules. These genes helped L. crispatus to better adapt to the acidic environment of the vagina, to obtain more nutrients and to maintain its dominance in respect with other strains. In feces-derived bacteria, more genes encoding CRISPR/Cas system, glycoside hydrolases family and tetracycline/lincomycin resistance genes are found probably to adapt these strains to the complex intestinal environment.
Pan et al, 2020	USA* (Isolati non tutti USA)	WGS	105	homo sapiens; chicken	human: vagina (50), gut (4), urine/bladder (10), other (3); chicken gut (13), Turkey Ileum (25)	average lenght: 2.21 Mbp; GC content: 36.95%; gene number: 12114 (human isolates and poultry isolates pan genome were constituited by 10198 (8402 genes with only 38 strains) and 5250 genes, respectively); average CDS: NA; core genome: 465 genes	Human and poultry isolates evolved to encompass different genomic features (carbohydrate usage, CRISPR-Cas immune systems, prophage occurrence) in order to thrive in different environmental niches. Chicken and turkey L. crispatus isolates can be also differentiated based on their genomic information, suggesting significant differences may exist between these two poultry gut niches. These results provide insights into host and niche-specific adaptation patterns.
Mancabelli et al, 2021	Italy	WGS	16	homo sapiens; chicken	human: vagina (8), other (1); chicken gut (7)	average lenght: 2.10 Mbp (2.1 Mbp and 2.01 Mbp in humans and chicken, respectively); GC content: NA; gene number: 2496-1912; average CDS: NA; core genome: 452 genes	A genetic adaptation of strains to their respective ecological niche is demonstrated. In vitro growth experiments involving all isolates, together with a simulated vaginal microbiota, revealed how this species are able to modulate the composition of the vaginal microbial community.
Hulyalkar et al, 2021	USA	WGS	3	homo sapiens	urine	average lenght: 2.43 Mbp; GC content: 37.1%; gene number: NA; average CDS: 2400 ; core genome: NA	Complete genome of a strains isolated from UTI patients
Oliveira de Almeida et al 2021	Brazil	WGS	4	homo sapiens	vagina	average lenght: 2.95 Mbp; GC content: NA; gene number: NA; average CDS: 2410 ; core genome: NA	Characterization of 4 isolates in Brazil
Fontana et al, 2021	Italy	WGS	16	homo sapiens, other host	human: vagina (8), other host (8)	average lenght: NA; GC content: NA; gene number: NA; average CDS: 2118; core genome: NA	Comparative genomic analysis to find out information on the use these strains as probiotics
Costantini et al, 2021	Italy	WGS	1	homo sapiens	vagina	average lenght: 2.10; GC content: 36.7; gene number: NA; average CDS: 2015; core genome: NA	L. crispatus strain BC5 efficiently utilizes different carbon sources, although some specific carbon sources that can be associated to the human diet were not metabolized (uridine, amygdalin, tagatose).While L. crispatus strain BC5 showed ability to use the nitrogen sources, it did not show tolerance/resistance towards twice the number of stressors (i.e. antibiotics, toxic metals etc.).

Table 1. Most significant publications on genomic characterization of *Lactobacillus crispatus* with the details of the key biological proprieties of corresponding strains.

These differences suggested that not all *L. crispatus* strains may be protective against vaginal pathobiont overgrowth, prompting to genotypical and phenotypical characterization to explore their functional activity and then their potential use as probiotics[34,38].

Lactobacillus crispatus Bacteriocins

As phylogenetic divergence has been described from comparative genomic studies, we aim to elucidate how bacteriocins encoding genes were conserved among some *L. crispatus* strains. While *L. crispatus* diversity has been deeply investigated based on host isolation, biological niche and metabolism, few data are available

regarding antimicrobial activity of these peptides. For this aim, we retrieved a total of 95 *L. crispatus* sequences from the NCBI database that we analyzed by using BAGEL4 to detect Area Of Interest (AOI)-containing putative bacteriocins encoding genes.

Firstly, *L. crispatus* genomes were clustered by using a core genome multilocus sequence typing (cgMLST) scheme obtained using chewBBACA package based on 24 complete genomes from NCBI database (Supplementary Table 1)[30]. cgMLST analysis provided a comprehensive overview of the genetic diversity of *L. crispatus* strains, highlighting differences regarding bacteriocins genes presence and their diversification. Sixteen genomes of animal-derived strains were also included that clustered together in the phylogenetic tree, compared to human strains, confirming a host-specific evolutionary divergence. Unfortunately, no indication about the host source was available for 4 genomes.

Of the 95 genomes, four were excluded from the analysis, and not included in the previously phylogenetic tree, because they lacked putative bacteriocin-encoding genes (Supplementary Table 1). The bacteriocin most represented was *Enterolysin A*, which was detected in 87 out of 91 genomes (95.6%). Next, genes encoding for *Helveticin J* (94.5%), *Bacteriocin Helveticin J* (93.2%), *SakT alpha* (51.6%), *Penocin A* (44.0%), *LAPs* (19.8%), *Bovicin 255* (13.2%) were recognized. Among the less represented were *Amylovorin*, *Lactacin F*, *Lanthipeptide class I* (5.5%), *Lanthipeptide class IV* (2.2%) and *Thermophilin A*, *Lactococcin 972*, *CoagulinA* and *Ericin S* that were detected only in the 1.1% of the genomes (Fig. 1, Supplementary Table 1). Intriguingly, BAGEL4 detected orthologous genes for *Enterolysin A*, *Helveticin J*, *Bacteriocin Helveticin J* and *Penocin A*: *Enterolysin A* 1 (78/91, 85.7%) and *Enterolysin A* 2 (82/91, 90.1%); *Helveticin J* 1 (53/91, 58.2%), *Helveticin J* 2 (11/91, 12.0%), *Helveticin J* 3 (25/91, 27.4%), *Helveticin J* 4 (13/91, 14.3%) and *Helveticin J* 5 (2/91, 2.19%); *Bacteriocin Helveticin J* 1 (33/91, 36.2%), *Bacteriocin Helveticin J* 2 (41/91, 45.0%) and *Bacteriocin Helveticin J* 3 (9/91, 9.78%); *Penocin A* 1 (23/91,

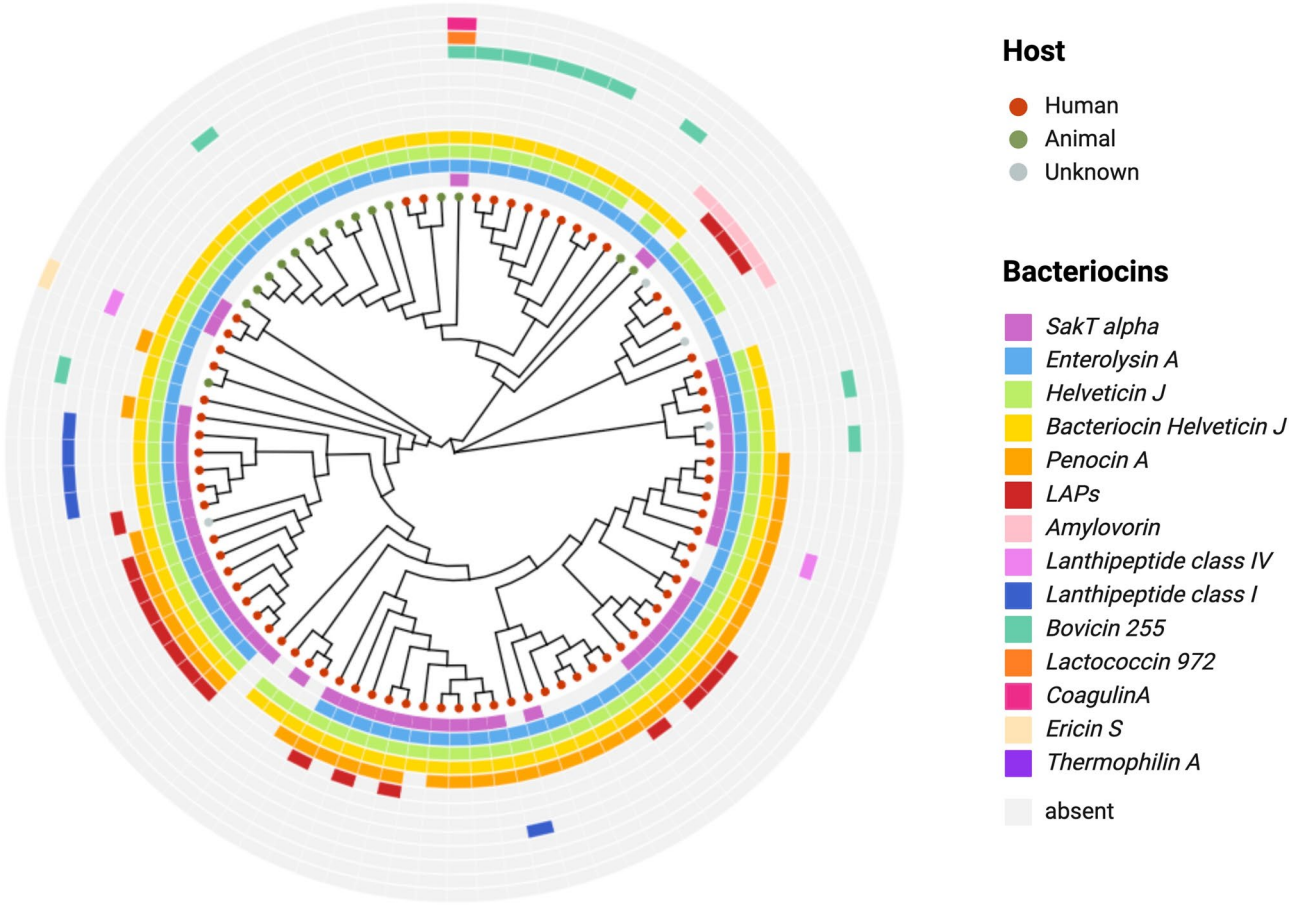

Fig. 1. Phylogenetic tree of *Lactobacillus crispatus* genomes and associated bacteriocins. The circular phylogenetic tree represents the relationship between 91 genomes of *Lactobacillus crispatus.* Branch lengths are indicative of evolutionary distances and the inner circle indicates the host, with red dots representing human-derived strains, green dots for animal-derived strains, and grey dots for strains with unknown origins. The outer colored blocks denote the presence of bacteriocin genes for every genome, and the colors correspond to the bacteriocin name listed in the legend. The bacteriocins includes: *SakT alpha*, *Enterolysin A*, *Helveticin J*, *Bacteriocin Helveticin J*, *Penocin A*, *LAPs*, *Amyrovicin*, *Lanthipeptide class IV*, *Lanthipeptide class I*, *Bovicin 255*, *Lacticoccin 972*, *Coagulin*, *Ericin S* and *Thermophilin A*. Light gray spaces indicate the absence of these bacteriocin genes.

25.2%) and *Penocin A* 2 (39/91, 42.8%). Only *Enterolysin A* and *Helveticin J* were found in more copies in these selected genomes (Fig. 2) further supporting their functional role to sustain bacterial fitness.

According to the presence of the most represented bacteriocin encoding genes, *L. crispatus* strains showed an overall genetic similarity. Conversely, the presence of less represented bacteriocin genes may explain differences among isolates from different hosts or biological niches. Notably, animal strains shared some bacteriocins with human strains, probably representing the most conserved and ancestral genes, such as *Enterolysin A*, *Helveticin J* and *Bacteriocin Helveticin J*, while only a small percentage of animal strains showed the presence of other bacteriocins, including *SakT alpha* and *Bovicin 255* (2/16, 12.5%), *Amylovorin*, *Lactococcin 972* and *Coagulin A* (1/16, 6.25%).

Furthermore, we investigated the genetic variability encompassing the most represented bacteriocin-encoding genes. For each bacteriocin, we performed a multiple alignment to identify single nucleotide polymorphisms (SNPs) and indels. The most abundant allele sequence was considered as the wild-type allele. Finally, we assessed the degree of conservation as the ratio of non-synonymous to synonymous substitutions (dN/dS) observing an overall negative selection and suggesting that bacteriocins encoding genes are conserved.

SakT alpha gene, detected in 47/91 genomes, showed a dN/dS ratio of 1 suggesting an overall negative selection: 82.6% of the detected genes were similar, while 2 alleles were identified with a non-synonymous mutation or a 5' deletion (Fig. 3a and Supplementary Table 2). Both *Enterolysin A*1 and A2 genes showed dN/dS ratio of 0.38 and 0.78. Interestingly, we found 22 different alleles for *Enterolysin A*1 based on SNPs (16 up to 58 detected SNPs were non-synonymous) and indels combinations, while *Enterolysin A*2 showed 13 alleles (7 up to 16 identified SNPs were non-synonymous) that were detected in 44/82 genomes (53.6%) (Fig. 3b and Supplementary Table 3). Furthermore, only for the *Enterolysin* A1 gene we detected deletions in three alleles. *Helveticin J* genes showed four major orthologous genes: while *Helveticin J* 1 appeared mostly conserved, while *Helveticin J* 3 showed dN/dS > 1 with several mutated alleles (Fig. 3c and Supplementary Table 4). Finally, *Bacteriocin Helveticin J* showed a dN/dS ratio less than 1 even though these genes had the greatest number of SNPs (Fig. 3d and Supplementary Table 5).

Our findings suggest that bacteriocin encoding genes in *L. crispatus* can be grouped in two types: those highly conserved (dN/dS < 1) that form the core set and those less conserved, showing some degree of polymorphism (dN/dS > 1) that form the "cloud" set of bacteriocin genes.

Discussion

L. crispatus species have been isolated from various hosts and biological niches, ranging from healthy poultry gut to various districts of the human body including the oral cavity, rectum, and urinary tract, suggesting within host and within species diversity[33,36]. In this context, nucleotide variations within protein-encoding genes, shared across non-identical *L. crispatus* isolates, may significantly contribute to strain diversification and impair their intra-species competition within the VMC[39].

L. crispatus, as commonly *Lactobacillus* species, produce and secrete several soluble compounds, able to acidify the colonized environment and diverse peptides[40–46]. Some of these peptides are known as bacteriocins and they are classified based on their size, microbial target, mechanism of action, secretion mechanism, and immunity mechanisms[45,47]. Bacteriocins are ribosomally synthesized and show a wide range of structural diversity, from simple linear peptides to complex post-translationally modified peptides (RiPPs)[47].

Of note, bacteriocins secreted by lactobacilli differ in length, genetic origin, biochemical features, molecular weight (MW), cellular receptors and interaction with the immune system[45]. During the last decade, *Lactobacillus*-synthesized bacteriocins have received a lot of attention due to their high potential, especially as natural food preservatives and novel antibacterial agents in the prevention or treatment of bacterial infections, making them attractive for various applications[48].

To facilitate bacteriocins research, BAGEL4 database has been available, updating BAGEL3 and becoming fully independent from ORF-calling. Gene clusters of interest (AOI) are discovered using the core-peptide database through HMM motifs[31].

As phylogenetic divergence has been described from comparative genomic studies, we highlighted how bacteriocins encoding genes were conserved among some *L. crispatus* strains.

The genomic analysis we performed suggests that a core of bacteriocin genes may play a key role in warranting *L. crispatus* adaptation to specific host biological niches, as the vaginal microenvironment, while a less pervasive set of bacteriocin genes may have evolved to confer peculiar antimicrobial activities useful to compete with specific pathobionts[49].

These bacteriocins may likely reflect adaptive changes driven by selective pressures in different ecological niches prompting for their strain variability and divergence affecting host range or interactions with other microbial species. Indeed, a dN/dS ratio < 1, which we observed for the majority of bacteriocin-encoding genes (e.g., Enterolysin A, Helveticin J, and Bacteriocin Helveticin J), is indicative of purifying (negative) selection. This suggests that amino acid-changing mutations are being selectively removed, implying that these genes encode functionally important and evolutionarily conserved and constrained proteins. Core bacteriocins, being conserved and under purifying selection, likely represent evolutionarily optimized traits that contribute to stable colonization and dominance in host-associated microbiota. Their wide presence suggests they are crucial for a persistent host colonization, bacterial fitness and defence against ubiquitous competitors, particularly in the acidic vaginal environment.

We hypothesize that highly conserved bacteriocins (under purifying selection) may serve a fundamental function in colonizing the environmental niche independently of the host. These conserved bacteriocins may be seen as the core set of bacteriocins deployed by *L. crispatus* as defense mechanism in CST-I dominated communities. The less conserved set of bacteriocins are also found more sporadically (e.g. SakT alpha, Penocin A variants), and may have evolved to maintain a degree of variability and "plasticity" useful to support adaptation

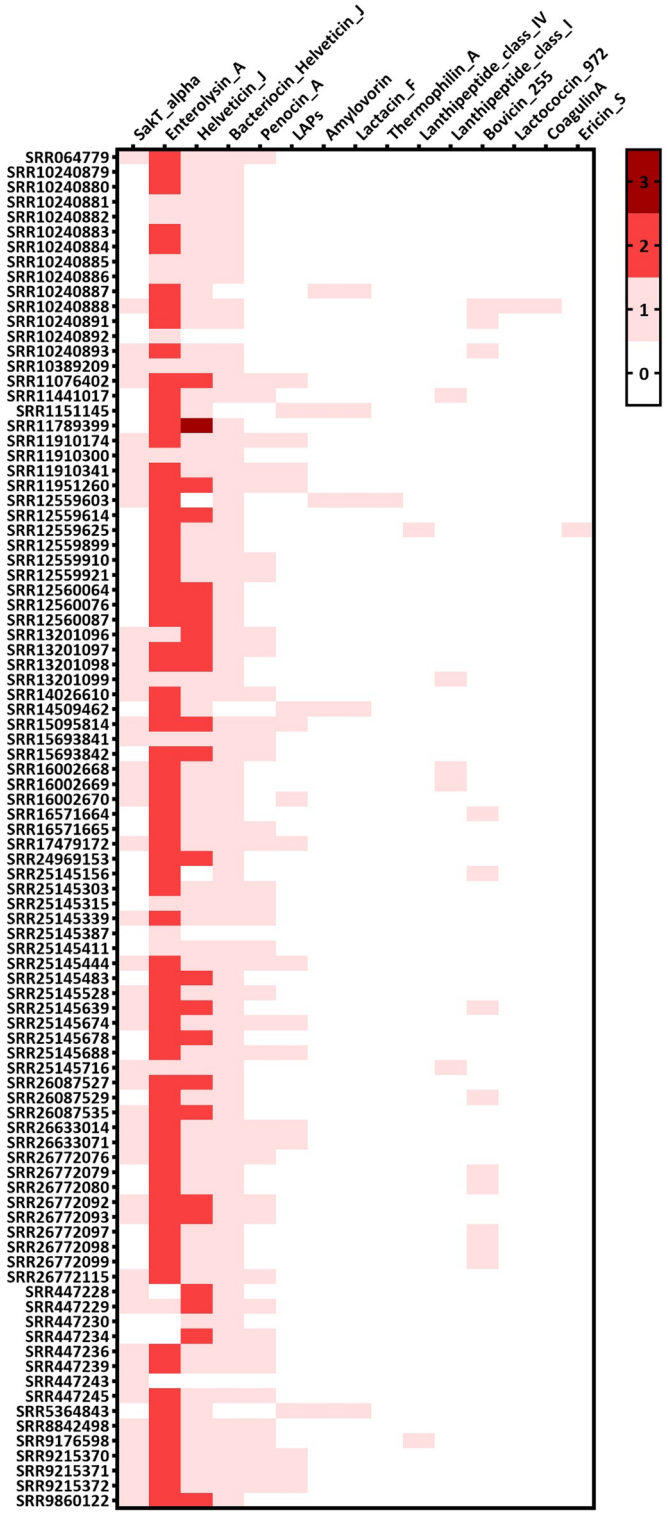

Fig. 2. Heat map of copy number variation of bacteriocin genes in *Lactobacillus crispatus* genomes. The heat map represents the copy number variation of different bacteriocin genes across different *L. crispatus* genomes. The intensity of the red color corresponds to the presence of multiple copies of a specific bacteriocin. The deeper red indicates the maximum number of copies, i.e. four. White indicates the absence of the bacteriocin gene in the genome.

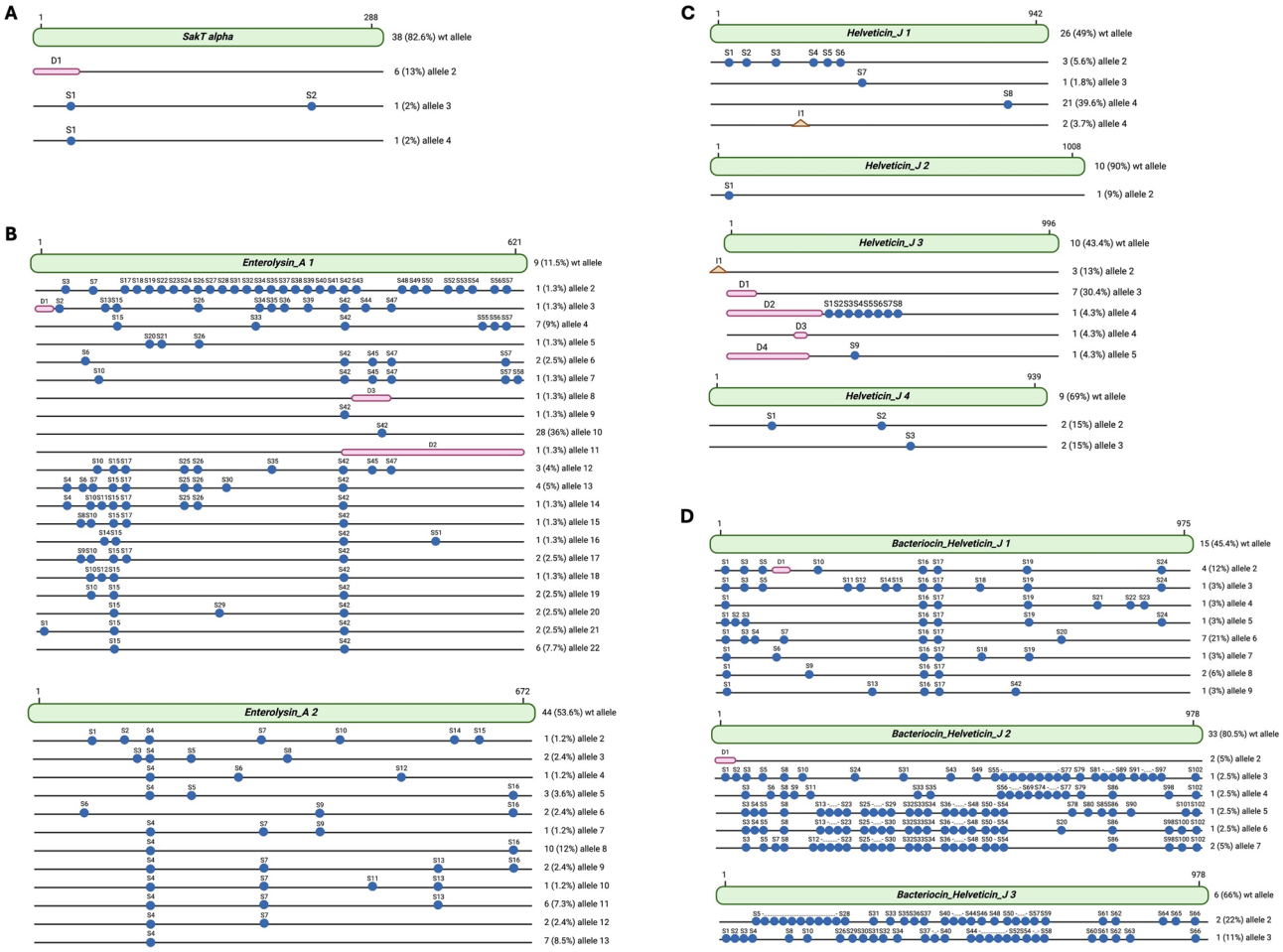

Fig. 3. Mutations and allelic variation of the most represented bacteriocins. The figure illustrates the sequence variations observed in different bacteriocin genes found in *L. crispatus* genomes using BAGEL4. Each panel (**a**–**d**) corresponds to a different bacteriocin gene, with the wild-type allele shown at the top as the most represented reference. *Enterolisyn A*, *Helveticin J* and *Bacteriocin Helveticin J* have different non-overlapping sequences and are therefore identified in different genes. Below the reference gene, the following rows represent different alleles with specific mutations indicated by blue circles for point mutations, pink bars for deletions and orange triangles for insertions. Each allele is associated with a frequency (percentage) in the total number of genomes containing the bacteriocin.

to different microbial ecosystems (different host) or a different environmental niche, such as the nutrient-rich gut or diverse urinary tract. In line with a broader definition proposed by Koonin[49], these bacteriocin genes may be defined as a set of "cloud" genes showing variability in presence and sequence that can be seen as a repertoire of antimicrobial peptides that provide ecological fine-tuning and adaptability. Their diversity and sporadic distribution could support strain-specific interactions or responses to localized microbial threats, such as particular pathobionts.

For example, bacteriocins with higher dN/dS ratios, such as Helveticin J3, may undergo adaptive evolution, potentially reflecting functional diversification in response to different host niches or microbial threats, also cooperating with other secreted compounds under the acidic and low-nutrient conditions of the vagina[50].

Although, core versus cloud bacteriocins distinction remains hypothetical and speculative and awaits functional validation, some studies prompt towards this hypothesis. Enterolysin A has been shown to have bacteriolytic activity against a broad range of Gram-positive bacteria, including *Clostridium difficile* and *Listeria monocytogenes*[51], while Helveticin J exhibits bactericidal activity against other *Lactobacillus* species and enteric pathogens, contributing to niche dominance[52]. These functional studies support our hypothesis that conserved bacteriocins may provide a baseline competitive advantage to *L. crispatus* strains by enabling inhibition of common vaginal or gut competitors, especially in *Lactobacillus*-dominated CSTs. On the other side, further studies will be needed to elucidate the specific functions of these genes. The detection of more Helveticin like encoding genes (Helveticin J and Bacteriocin Helveticin J, and their orthologous genes), may suggest a different activity.

The combination of core and cloud bacteriocin genes may explain the differences between the *L. crispatus* activity in protecting or restoring the healthy vaginal microbiota. Future studies, using knockout mutants or

overexpression strains, on the biological properties of these peptides and their impact on microbiota and host may shed light on this complex interplay.

Data availability

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Received: 6 February 2025; Accepted: 20 June 2025
Published online: 01 July 2025

References

1. Han, Y. & Ren, Q. L. Does probiotics work for bacterial vaginosis and vulvovaginal candidiasis. *Curr. Opin. Pharmacol.* **61**, 83–90 (2021).
2. Liu, P., Lu, Y., Li, R. & Chen, X. Use of probiotic lactobacilli in the treatment of vaginal infections: in vitro and in vivo investigations. *Front. Cell. Infect. Microbiol.* **13**, 1153894 (2023).
3. Smith, S. B. & Ravel, J. The vaginal microbiota, host defence and reproductive physiology. *J. Physiol.* **595** (2), 451–463 (2017).
4. Stoyancheva, G., Marzotto, M., Dellaglio, F. & Torriani, S. Bacteriocin production and gene sequencing analysis from vaginal *Lactobacillus* strains. *Arch. Microbiol.* **196** (9), 645–653 (2014).
5. Kobatake, E. & Kabuki, T. S-Layer protein of Lactobacillus helveticus SBT2171 promotes human beta-Defensin 2 expression via TLR2-JNK signaling. *Front. Microbiol.* **10**, 2414 (2019).
6. Abramov, V. M. et al. S-layer protein 2 of *Lactobacillus crispatus* 2029, its structural and Immunomodulatory characteristics and roles in protective potential of the whole bacteria against foodborne pathogens. *Int. J. Biol. Macromol.* **150**, 400–412 (2020).
7. Deka, N., Hassan, S., Seghal Kiran, G. & Selvin, J. Insights into the role of vaginal Microbiome in women's health. *J. Basic. Microbiol.* **61** (12), 1071–1084 (2021).
8. Coudray, M. S. & Madhivanan, P. Bacterial vaginosis-A brief synopsis of the literature. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **245**, 143–148 (2020).
9. Ravel, J., Moreno, I. & Simon, C. Bacterial vaginosis and its association with infertility, endometritis, and pelvic inflammatory disease. *Am. J. Obstet. Gynecol.* **224** (3), 251–257 (2021).
10. Lev-Sagie, A. et al. Vaginal Microbiome transplantation in women with intractable bacterial vaginosis. *Nat. Med.* **25** (10), 1500–1504 (2019).
11. Yockey, L. J. et al. Screening and characterization of vaginal fluid donations for vaginal microbiota transplantation. *Sci. Rep.* **12** (1), 17948 (2022).
12. Jawanda, I. K., Soni, T., Kumari, S. & Prabha, V. The evolving facets of vaginal microbiota transplantation: reinvigorating the unexplored frontier amid complex challenges. *Arch. Microbiol.* **206** (7), 306 (2024).
13. Ravel, J. et al. Vaginal Microbiome of reproductive-age women. *Proc. Natl. Acad. Sci. U S A.* **108** (Suppl 1(Suppl 1), 4680–4687 (2011).
14. Chee, W. J. Y., Chew, S. Y. & Than, L. T. L. Vaginal microbiota and the potential of Lactobacillus derivatives in maintaining vaginal health. *Microb. Cell. Fact.* **19** (1), 203 (2020).
15. Ma, Z. S. & Li, L. Quantifying the human vaginal community state types (CSTs) with the species specificity index. *PeerJ* **5**, e3366 (2017).
16. De Seta, F., Campisciano, G., Zanotta, N., Ricci, G. & Comar, M. The vaginal community state types Microbiome-Immune network as key factor for bacterial vaginosis and aerobic vaginitis. *Front. Microbiol.* **10**, 2451 (2019).
17. Argentini, C. et al. Evaluation of modulatory activities of Lactobacillus crispatus strains in the context of the vaginal microbiota. *Microbiol. Spectr.* **10** (2), e0273321 (2022).
18. France, M., Alizadeh, M., Brown, S., Ma, B. & Ravel, J. Towards a deeper Understanding of the vaginal microbiota. *Nat. Microbiol.* **7** (3), 367–378 (2022).
19. Di Pierro, F. et al. Oral administration of *Lactobacillus crispatus* M247 to papillomavirus-infected women: results of a preliminary, uncontrolled, open trial. *Minerva Obstet. Gynecol.* **73** (5), 621–631 (2021).
20. Di Pierro, F. B., Cattivelli, A., Soldi, D. & Elli, S. M. *Lactobacillus crispatus* M247: a possible tool to counteract CST IV. *Nutrafoods* 17 (2018).
21. Strus, M. et al. Studies on the effects of probiotic *Lactobacillus* mixture given orally on vaginal and rectal colonization and on parameters of vaginal health in women with intermediate vaginal flora. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **163** (2), 210–215 (2012).
22. Dellino, M. et al. *Lactobacillus crispatus* M247 oral administration: is it really an effective strategy in the management of papillomavirus-infected women? *Infect. Agent Cancer*. **17** (1), 53 (2022).
23. Molina, M. A., Andralojc, K. M., Huynen, M. A., Leenders, W. P. J. & Melchers, W. J. G. In-depth insights into cervicovaginal microbial communities and HrHPV infections using high-resolution Microbiome profiling. *NPJ Biofilms Microbiomes*. **8** (1), 75 (2022).
24. Baud, A. et al. Microbial diversity in the vaginal microbiota and its link to pregnancy outcomes. *Sci. Rep.* **13** (1), 9061 (2023).
25. McComb, E., Holm, J., Ma, B. & Ravel, J. Complete genome sequence of *Lactobacillus crispatus* CO3MRSI1. *Microbiol. Resour. Announc* **8**(5) (2019).
26. Oliveira de Almeida, M. et al. Characterization of the first vaginal *Lactobacillus crispatus* genomes isolated in Brazil. *PeerJ* **9**, e11079 (2021).
27. Colombini, L. et al. The mobilome of *Lactobacillus crispatus* M247 includes two novel genetic elements: Tn7088 coding for a putative bacteriocin and the siphovirus prophage PhiM247. *Microb Genom* **9**(12) (2023).
28. Chen, S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using Fastp. *Imeta* **2** (2), e107 (2023).
29. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* **13** (6), e1005595 (2017).
30. Silva, M. et al. ChewBBACA: a complete suite for gene-by-gene schema creation and strain identification. *Microb. Genom.* **4**(3) (2018).
31. van Heel, A. J. et al. BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res.* **46** (W1), W278–W281 (2018).
32. van der Veer, C. 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* **7** (1), 49 (2019).
33. Zhang, Q. et al. Comparative genomics of *Lactobacillus crispatus* from the gut and vagina reveals genetic diversity and lifestyle adaptation. *Genes (Basel)* **11**(4) (2020).

34. Costantini, P. E. et al. Insight into phenotypic and genotypic differences between vaginal *Lactobacillus crispatus* BC5 and *Lactobacillus gasseri* BC12 to unravel nutritional and stress factors influencing their metabolic activity. *Microb. Genom* **7**(6). (2021).
35. Pan, M., Hidalgo-Cantabrana, C., Goh, Y. J., Sanozky-Dawes, R. & Barrangou, R. Comparative analysis of *Lactobacillus gasseri* and *Lactobacillus crispatus* isolated from human urogenital and Gastrointestinal tracts. *Front. Microbiol.* **10**, 3146 (2019).
36. Mancabelli, L. 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. Microbiol* **87**(8) (2021).
37. Hulyalkar, N. V. et al. Complete genome sequences of three *Lactobacillus crispatus* strains isolated from the urine of postmenopausal women. *Microbiol. Resour. Announc*. **10** (48), e0101721 (2021).
38. Fontana, F. et al. Probiogenomics analysis of 97 *Lactobacillus crispatus* strains as a tool for the identification of promising Next-Generation probiotics. *Microorganisms* **9**(1). (2020).
39. Tarracchini, C. et al. The core genome evolution of *Lactobacillus crispatus* as a driving force for niche competition in the human vaginal tract. *Microb. Biotechnol.* **16** (9), 1774–1789 (2023).
40. Sablon, E., Contreras, B. & Vandamme, E. Antimicrobial peptides of lactic acid bacteria: mode of action, genetics and biosynthesis. *Adv. Biochem. Eng. Biotechnol.* **68**, 21–60 (2000).
41. Aroutcheva, A. et al. Defense factors of vaginal lactobacilli. *Am. J. Obstet. Gynecol.* **185** (2), 375–379 (2001).
42. O'Hanlon, D. E., Moench, T. R. & Cone, R. A. Vaginal pH and microbicidal lactic acid when lactobacilli dominate the microbiota. *PLoS One*. **8** (11), e80074 (2013).
43. Sara, M. & Sleytr, U. B. S-Layer proteins. *J. Bacteriol.* **182** (4), 859–868 (2000).
44. Fagan, R. P. & Fairweather, N. F. Biogenesis and functions of bacterial S-layers. *Nat. Rev. Microbiol.* **12** (3), 211–222 (2014).
45. Darbandi, A. et al. Bacteriocins: properties and potential use as antimicrobials. *J. Clin. Lab. Anal.* **36** (1), e24093 (2022).
46. Fernandes, A. & Jobby, R. Bacteriocins from lactic acid bacteria and their potential clinical applications. *Appl. Biochem. Biotechnol.* **194** (10), 4377–4399 (2022).
47. Sugrue, I., Ross, R. P. & Hill, C. Bacteriocin diversity, function, discovery and application as antimicrobials. *Nat. Rev. Microbiol.* **22** (9), 556–571 (2024).
48. De Giani, A. et al. Identification of a bacteriocin-like compound from Lactobacillus plantarum with antimicrobial activity and effects on normal and cancerogenic human intestinal cells. *AMB Express*. **9** (1), 88 (2019).
49. Koonin, E. V., Makarova, K. S. & Wolf, Y. I. Evolution of microbial genomics: conceptual shifts over a quarter century. *Trends Microbiol.* **29** (7), 582–592 (2021).
50. Kovachev, S. Defence factors of vaginal lactobacilli. *Crit. Rev. Microbiol.* **44** (1), 31–39 (2018).
51. Nilsen, T., Nes, I. F. & Holo, H. Enterolysin A, a cell wall-degrading bacteriocin from *Enterococcus faecalis* LMG 2333. *Appl. Environ. Microbiol.* **69** (5), 2975–2984 (2003).
52. Joerger, M. C. & Klaenhammer, T. R. Characterization and purification of helveticin J and evidence for a chromosomally determined bacteriocin produced by *Lactobacillus helveticus* 481. *J. Bacteriol.* **167** (2), 439–446 (1986).

Acknowledgements
Not applicable.

Author contributions
This study was designed by GS, GD and FDM. Data collection was performed by GS and FDM. Data analysis and figure preparation were performed by GS and FDM. GS and FDM wrote the initial draft. GD and FDM revised the final manuscript. All authors participated in the discussion and interpretation of the results. All authors have read and agreed to the published version of the manuscript. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Funding
We acknowledge EU funding for the MUR PNRR Extended Partnership initiative on Emerging Infectious Diseases (Project number PE00000007, INF-ACT).

Declarations

Ethics approval and consent to participate
Ethical approval is not required as no clinical data were used.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-08321-7.

Correspondence and requests for materials should be addressed to F.D.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.