MICROBIAL GENOMICS

RESEARCH ARTICLE

Davray *et al.*, *Microbial Genomics* 2021;7:000472

DOI 10.1099/mgen.0.000472

MICROBIOLOGY SOCIETY

OPEN DATA OPEN ACCESS

Plasmids encode niche-specific traits in *Lactobacillaceae*

Dimple Davray, Dipti Deo and Ram Kulkarni*

Abstract

Species belonging to the family *Lactobacillaceae* are found in highly diverse environments and play an important role in fermented foods and probiotic products. Many of these species have been individually reported to harbour plasmids that encode important genes. In this study, we performed comparative genomic analysis of publicly available data for 512 plasmids from 282 strains represented by 51 species of this family and correlated the genomic features of plasmids with the ecological niches in which these species are found. Two-thirds of the species had at least one plasmid-harbouring strain. Plasmid abundance and GC content were significantly lower in vertebrate-adapted species as compared to nomadic and free-living species. Hierarchical clustering highlighted the distinct nature of plasmids from the nomadic and free-living species than those from the vertebrate-adapted species. EggNOG-assisted functional annotation revealed that genes associated with transposition, conjugation, DNA repair and recombination, exopolysaccharide production, metal ion transport, toxin–antitoxin system, and stress tolerance were significantly enriched on the plasmids of the nomadic and in some cases nomadic and free-living species. On the other hand, genes related to anaerobic metabolism, ABC transporters and the major facilitator superfamily were over-represented on the plasmids of the vertebrate-adapted species. These genomic signatures correlate with the comparatively nutrient-depleted, stressful and dynamic environments of nomadic and free-living species and nutrient-rich and anaerobic environments of vertebrate-adapted species. Thus, these results indicate the contribution of the plasmids in the adaptation of lactobacilli to their respective habitats. This study also underlines the potential application of these plasmids in improving the technological and probiotic properties of lactic acid bacteria.

DATA SUMMARY

Nucleotide sequences of plasmids of *Lactobacillus* strains for which complete genome sequences were available were retrieved from the NCBI genome (https://www.ncbi.nlm.nih.gov/genome) and PATRIC 3.5.41 databases on 31 March 2019. The dataset includes 512 nucleotide sequences of plasmids of 282 strains belonging to the genus *Lactobacillus* before its reclassification into several genera [1]. Details of the plasmids are given in Table S1 (available in the online version of this article).

INTRODUCTION

Lactobacilli are gram-positive, aerotolerant, non-sporulating bacteria which belong to the lactic acid bacteria (LAB) group wherein lactic acid is the major metabolic end product during glucose fermentation [2]. *Lactobacillus* is a major genus of the family *Lactobacillaceae*, and the highly diverse nature of its species has recently resulted in reclassification of this genus into 23 genera [1]. Many lactobacilli have been associated with humans in a variety of ways. Their presence in the gastrointestinal tract has earned extraordinary attention due to the health-promoting properties of a few of the genera in the family. These bacteria are also important in the food industry for the production of fermented dairy products by providing taste, texture and antibacterial activity [3]. Furthermore, lactobacilli have applications in the production of industrially important chemicals such as lactic acid [4].

Lactobacilli have been isolated from a wide range of habitats and have been found to have highly diverse traits. Based on phylogenetic analysis, source of isolation, prevalence of detection, optimal growth temperature, substrate utilization

Received 08 August 2020; Accepted 22 October 2020; Published 09 November 2020

Author affiliations: [1]Symbiosis School of Biological Sciences, Symbiosis International (Deemed University), Lavale, Pune 412115, India.

***Correspondence:** Ram Kulkarni, ramskulkarni@gmail.com; ram.kulkarni@ssbs.edu.in

Keywords: Clusters of Orthologous Groups; comparative genomics; ecological niche; *Lactobacillaceae*; plasmid-encoded trait; stress resistance genes.

Abbreviations: ARG, antibiotic resistance gene; CDS, coding sequence; FLL, free-living lactobacilli; HCL, hierarchical clustering; HMRG, heavy metal resistance gene; LAB, lactic acid bacteria; MCL, Markov clustering; MFS, major facilitator superfamily; NGS, next-generation sequencing; NL, nomadic lactobacilli; PTS, phosphotransferase system; VAL, vertebrate-adapted lactobacilli.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Twenty-one supplementary tables and two supplementary figures are available with the online version of this article.

and environmental stress tolerance, species of the earlier *Lactobacillus* genus have been classified as host-adapted (vertebrate and insect-adapted), nomadic and free-living [5]. Some of the genomic features of these species are correlated with these habitats. For example, the vertebrate-adapted lactobacilli (VAL) have undergone genome reduction by losing a substantial number of genes involved in carbohydrate metabolism and amino acid and cofactor biosynthesis because they thrive in a nutrient-rich environment [5–7]. On the other hand, free-living lactobacilli (FLL) and nomadic lactobacilli (NL) have larger genome sizes because they encode various enzymes which utilize a broad spectrum of carbohydrates [8]. Furthermore, *Lactobacillus helvticus* DPC4571 and *Lactobacillus acidophilus* NCFM were found to have genes important for adaptation to their habitats, namely dairy and gut, respectively [9]. Lastly, *Lactiplantibacillus plantarum*, a hallmark NL species found in most diverse habitats, was also observed to have a highly diverse genome with no unique environmental signature [10]. Another classification of lactobacilli is based on the carbohydrate fermentation pathways, as homofermentative and heterofermentative. In the homofermentative species, the phosphotransferase system (PTS) is the preferred transport route for sugars, which are fermented via the Emden–Meyerhoff pathway. On the other hand, in the heterofermentative species, the PTS is not functional and the phosphoketolase pathway is the preferred catabolic pathway [11, 12].

Plasmids have been extensively found in lactobacilli and reported to carry the genes that give them favourable traits and permit them to survive in competitive surroundings [13]. Many plasmids have been sequenced from individual strains and found to encode important functions such as oxidative stress response, antibiotic resistance, bacteriophage resistance, chloride and potassium transport, and bacteriocin production [14, 15]. Only a handful of reports are available on the importance of plasmids in the niche adaptation of lactobacilli. In *L. plantarum* P-8 isolated from fermented dairy products, loss of the plasmids important for dairy adaptation was found when the strain was administered in rats and humans [16]. A plasmid from *Lacticaseibacillus paracasei* NFBC338 isolated from the human gut encoded a gene involved in adherence to human intestinal epithelial cells [17]. In *Levilactobacillus brevis*, brewery isolates were found to have unique genes on the plasmids supporting growth under harsh conditions. At the same time, a large proportion of the brewery plasmidome was also found to be shared with insect isolates [18]. Similarly, in *Lactococcus lactis*, another important bacterium of the LAB group, plant cell-wall-modifying enzymes were encoded by the plasmids in plant-derived isolates [19]. Apart from these scarce strain- and species-specific reports, the contribution of plasmids in niche adaptation has not been holistically studied in the family *Lactobacillaceae*. In the last decade, next generation sequencing (NGS) has created extensive data on the genome sequences of many strains of lactobacilli. In the current study, we report on an analysis of plasmid sequence data of species of the genus *Lactobacillus* before its division into several genera [1], available in the public domain, to explore their genomic content and their possible involvement in niche adaptation.

Impact Statement

Bacteria of the family *Lactobacillaceae* are present in a wide range of habitats and play an important role in human health, fermented foods and chemical industries. A few studies have demonstrated the presence of plasmids in individual strains of *Lactobacillaceae* species encoding various traits. Extensive data of genome sequences of lactobacilli are becoming available; however, no comprehensive analysis of the plasmid-encoded genes and determining their biological relevance across lactobacilli has been undertaken at a larger scale. In this study, we explored the genomic content of 512 plasmids of *Lactobacillaceae* species and correlated it to the three types of these species according to their ecological niches – vertebrate-adapted, free-living and nomadic. Comparatively lower plasmid abundance and GC content in the vertebrate-adapted species could be linked to the presence of these species in the nutrient-rich environment. The genomic content of the plasmids was consistent with the respective lifestyle adopted by lactobacilli suggesting that the plasmids might enhance the niche-specific fitness of the strains. The plethora of important genes present on the plasmids may also make them a highly useful tool in improving the probiotic, technological and food-related properties of lactobacilli.

METHODS

General characterization of the plasmids

The coding sequences (CDSs) of the 512 plasmids were extracted from the GenBank database, along with their NCBI annotation (https://www.ncbi.nlm.nih.gov/nucleotide) [20]. Plasmids for which the encoded genes were not annotated in NCBI were subjected to analysis using Prodigal software (version 2.6) to predict the CDSs.

Functional annotation

Functional annotation and classification of the CDSs were performed using the eggNOG4.5 database (http://eggnogdb.embl.de/). For subgrouping of the genes under each COG category, their annotations in the eggNOG4.5, NCBI and KEGG databases were considered.

Markov clustering analysis

All-against-all bi-directional BLAST alignment was performed on the obtained CDSs with at least 50% amino acid identity and 50% query coverage [20]. The BLAST output was analysed using Markov clustering (MCL) in the mclblastine v12-0678 pipeline to classify proteins into families [21]. Hierarchical

clustering was computed in TM4 MeV Suite [22] based on the presence and absence of the protein families in plasmids by using an average linkage algorithm and Manhattan distance metric parameters. HCL analysis was visualized in the Interactive Tree of life (ITOL) [23] by importing a Newick tree from the TM4 MeV Suite. For identification of the plasmids which are possibly shared between the species from different habitats, a similarity matrix was created for the plasmids based on the presence and absence of the MCL families using the Sørensen–Dice coefficient. GenBank files of the plasmids with Sørensen–Dice similarity coefficient of 0.8 or more were compared to each other using the EasyFig program [24].

Identification of other important genes

To identify antibiotic resistance genes, initially, amino acid sequences of 2404 antibiotic resistance genes (ARGs) were downloaded from the CARD database (http://arpcard.mcmaster.ca). Plasmid CDSs were aligned against retrieved ARG sequences through standalone BLAST (BLAST P,ver.2.9.0, https://ftp.ncbi.nlm.nih.gov/blast/executables/LATEST/) with a threshold of query coverage of >70%, amino acid sequence identity of >30%, and *e*-value $<10^{-5}$. Exopolysaccharide (EPS) gene clusters were identified as described previously [25]. Bacteriocin operons were identified using the web version of BAGEL4 (http://bagel4.molgenrug.nl) using default parameters.

Statistical analysis

GraphPad Prism 8 was used to perform statistical analysis. All comparisons between habitats were performed by Kruskal–Wallis test with Dunn's post-hoc test. Wilcoxon's signed rank test was used to compare GC content of the plasmids with that of the chromosome of the same strain and species.

RESULT AND DISCUSSION

Data collection and general characterization

As of March 2019, the NCBI prokaryotic genome database was found to have complete genome sequences for 282 strains under the genus defined as *Lactobacillus* before its reclassification into several genera [1]. With the current classification, these data include 24 of the 31 genera under the family *Lactobacillaceae*. Collectively, nucleotide sequences of a total of 512 plasmids were found in the database for all these strains (Tables 1 and S1). Based on the classification of Duar *et al.* [5], in our dataset, the greatest number of strains (110) belonged to the NL group followed by VAL (86), FLL (43) and insect-adapted (six) groups. Forty strains belonged to species for which no clear information on their possible native habitat was available; hence, they were not considered for further analysis. Plasmids from insect-adapted strains were excluded from subsequent analyses as only three plasmids were represented by this category of strains. Similarly, based on the glycolytic pathway, 216 strains belonged to homofermentative and 51 to heterofermentative species [11, 12].

Table 1. General features of the plasmids from completely sequenced *Lactobacillaceae* species

Basic features of plasmids	Values
Number and proportion of the strains having plasmids	155 (54.7%)
Number and proportion of the species having plasmids	38 (66.6%)
Total number of plasmids	512
GC content range (%)	28.9–76.5
Size range (Kb)	0.83–759.6
Total number of coding sequences (CDSs)	19064
Range of CDSs per plasmid	1–809

The proportion of strains having plasmids was lowest for VAL (32.2%) and highest for FLL (76.7%) (Fig. 1a). The average number of plasmids per strain was significantly lower for VAL (one) than NL (2.3) and FLL (2.7) (Fig. 1, Table S1). On the other hand, the average number of plasmids per strain for each species was significantly higher only in FLL as compared to VAL, although this average was three-fold higher for NL than for VAL (Fig. 1c). This suggests that some of these differences could be driven by strains of only a few species, whereas some others are possibly common to a majority of the species of that habitat. Overall, the lower plasmid abundance in VAL as compared to NL and FLL is in agreement with the similar observation for the chromosomal genome size [5]. Such reduction is considered to be due to the loss of several biosynthetic pathway genes as the bacteria are living in a nutrient-rich environment in the hosts. Furthermore, loss of plasmids in a dairy isolate of *Lactobacillus* when administered in human and rat has also been reported. Such loss has been postulated to be because of the superfluous nature of the genes encoded by the lost plasmid in the gut environment as well as stress response of the bacteria during the passage through the harsh gastric environment [16]. There was no difference in average plasmid size per strain across various habitats (Fig. 1d), suggesting that the low number of plasmids in VAL is not compensated for by their larger size in these strains.

The average GC content of the plasmids per strain was significantly lower for VAL (37.7%) than NL (39.7%) and FLL (39.7%) (Fig. 1e). At the species level, this difference was significant only between VAL and NL (Fig. 1f). This observation is similar to the lower chromosomal GC content of the host-associated strains, which is thought to be because of non-adaptive loss of the DNA repair system leading to a mutational bias toward A and T [5]. The plasmid GC contents per strain and per species for NL and FLL were lower than the respective chromosomal GC contents; by contrast, for VAL, plasmids and chromosomes had similar GC content (Fig. 1e, f). This observation was also consistent with a similar comparison within the

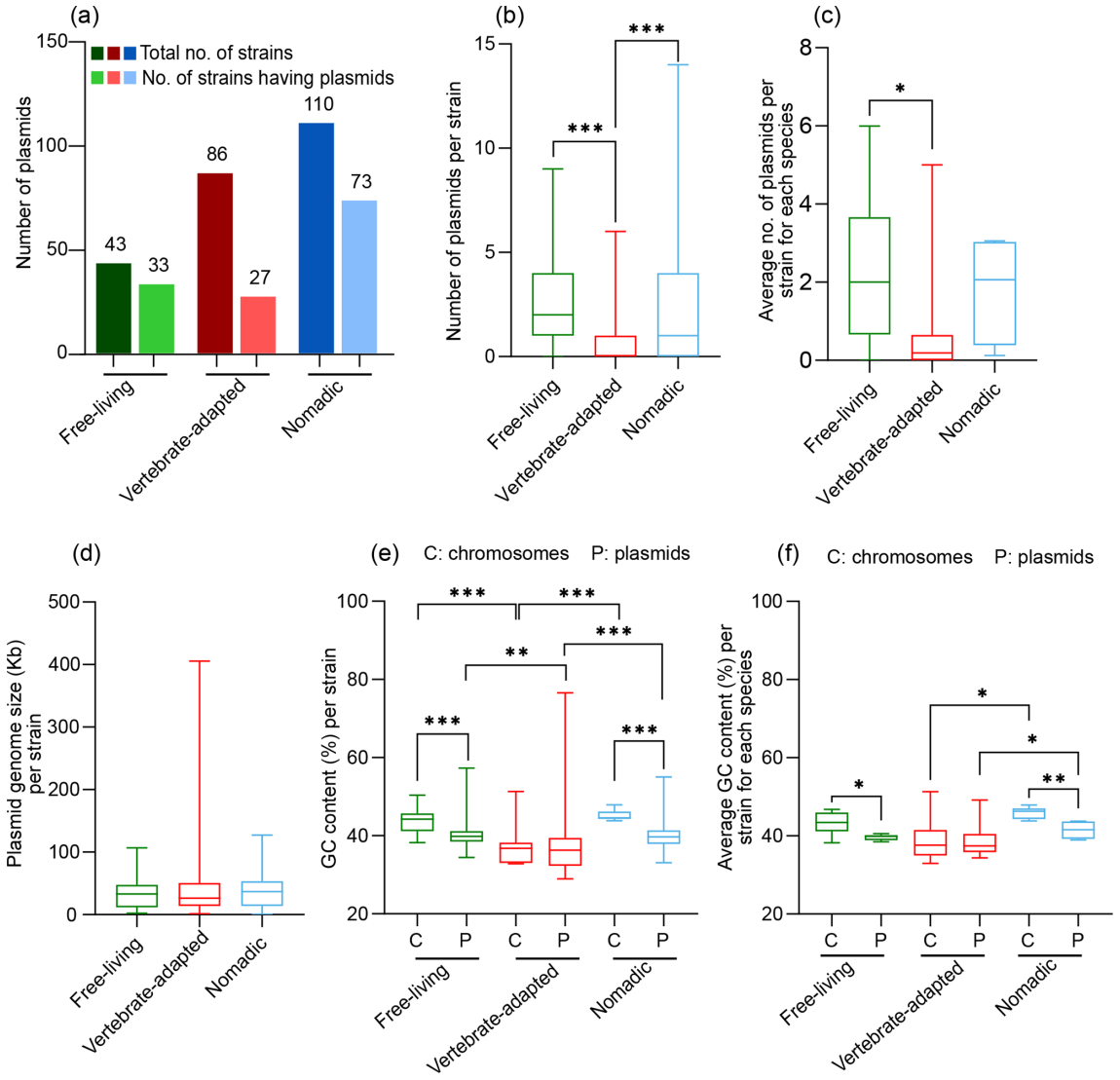

Fig. 1. General features of the plasmids in *Lactobacillaceae* species from various habitats, indicating the abundance of plasmids across habitats (a), number of plasmids per strain (b), average number of plasmids per strain for each species (c), average size of the plasmids per strain (d), GC content per strain (e) and average GC content per strain for each species (f). Statistical analyses were performed to compare the data across various habitats using a Kruskal–Wallis test with Dunn's post-hoc test. The GC contents of plasmids and chromosome of the same strain (e) and of the same species (f) were compared using the Wilcoxon signed rank test (***P <0.001; **P <0.01; *P <0.05).

individual strains wherein the average plasmid GC content was lower than the respective chromosomal GC content in all the NL strains and in 27 of 33 FLL strains. It has been reported that the maintenance of a plasmid with higher GC content is metabolically expensive for bacterial cells with lower chromosomal GC content and is thus evolutionary unfavourable [26]. It has also been proposed that plasmids can be transferred from chromosomally GC-poor hosts or environments to relatively GC-rich hosts [27]. Thus, the lower GC content of plasmids from NL and FLL than their chromosomes is not surprising and is indeed an indication of their possible acquisition by horizontal gene transfer. On the other hand, it has been suggested that with time the nucleotide composition of plasmids might become closer to that of the host genome because of the dependency of plasmid replication on the host cell machinery [28, 29]. Thus, the comparable GC content of VAL chromosomes and plasmids probably suggests their long association with each other. No correlation was observed between GC content and size of the plasmids (data not shown) in contrast to earlier observations [30].

Hierarchical clustering (HCL) analysis

To understand the extent of similarity between the plasmids based on their genomic content, their functional grouping

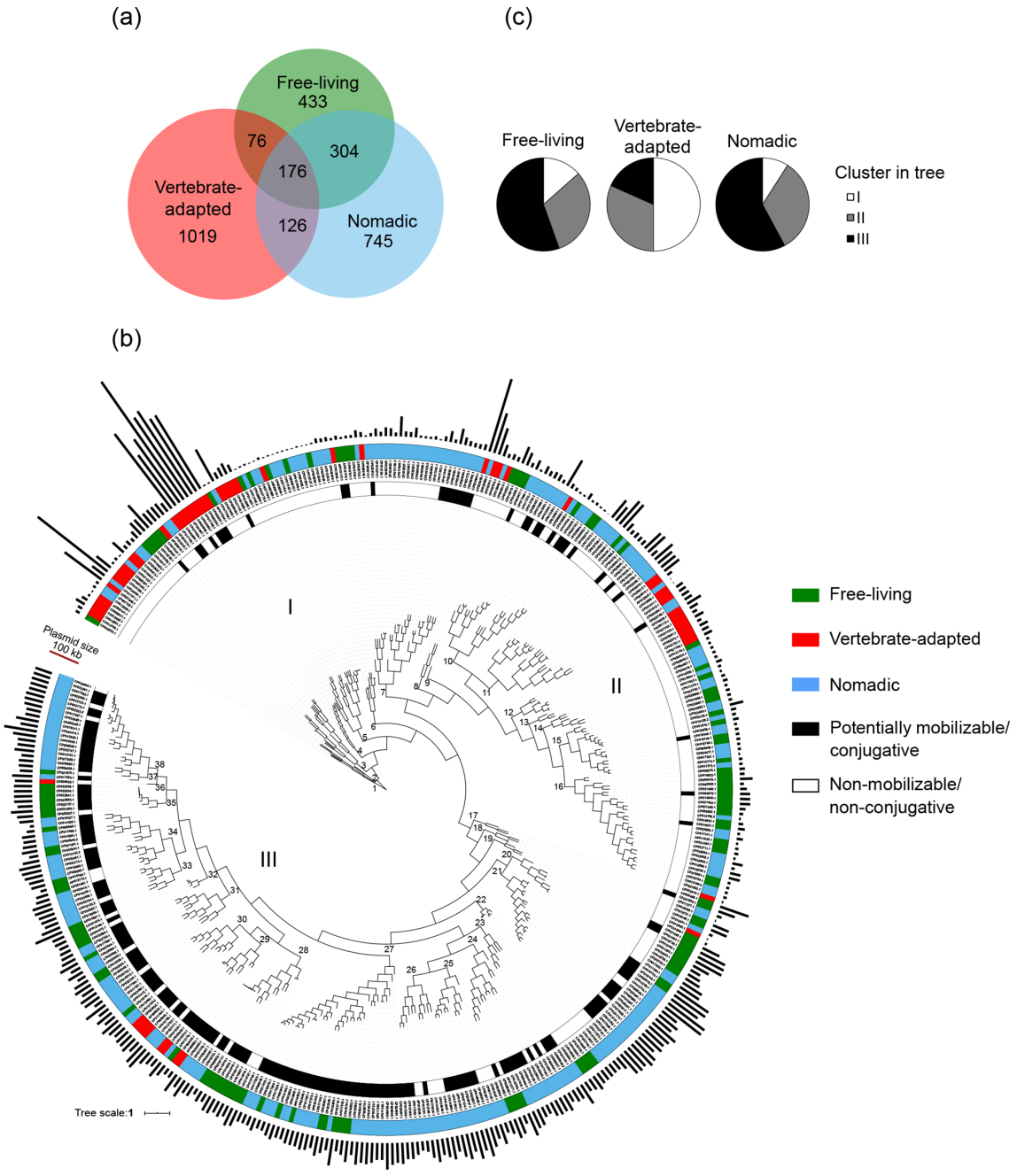

Fig. 2. HCL analysis of *Lactobacillaceae* plasmids. (a) Sharing of protein families encoded by plasmids as identified by MCL across lactobacilli from various habitats. (b) HCL tree displaying clustering of the plasmids based on their protein family composition. Black bars at the edge represent plasmid genome size. Plasmids having mobilization/conjugation genes were identified based on eggNOG annotations. (c) Pie chart representing the proportion of plasmids from each habitat under clusters I, II and III of the tree shown in (b).

was performed by HCL analysis. Based on MCL analysis, the CDSs (total 19064) could be classified into a total of 3380 protein families, with about half of them (1560) having a single member. This observation is similar to that on *L. lactis* plasmids, where also approximately half (413 out of 885) of the protein families were singleton [31]. Plasmids from VAL had the highest proportion (73%) of unique families, followed by NL (55%) and FLL (44%) (Fig. 2a). This probably suggests the uniqueness of the plasmids in VAL, which might have arisen during the long course of association of plasmids with these species as speculated based on the GC content. The highest sharing of families of NL with VAL and FLL could be due to the fact that NL is found in a wide range of environments, including those associated with the hosts as well as those similar to FLL [5].

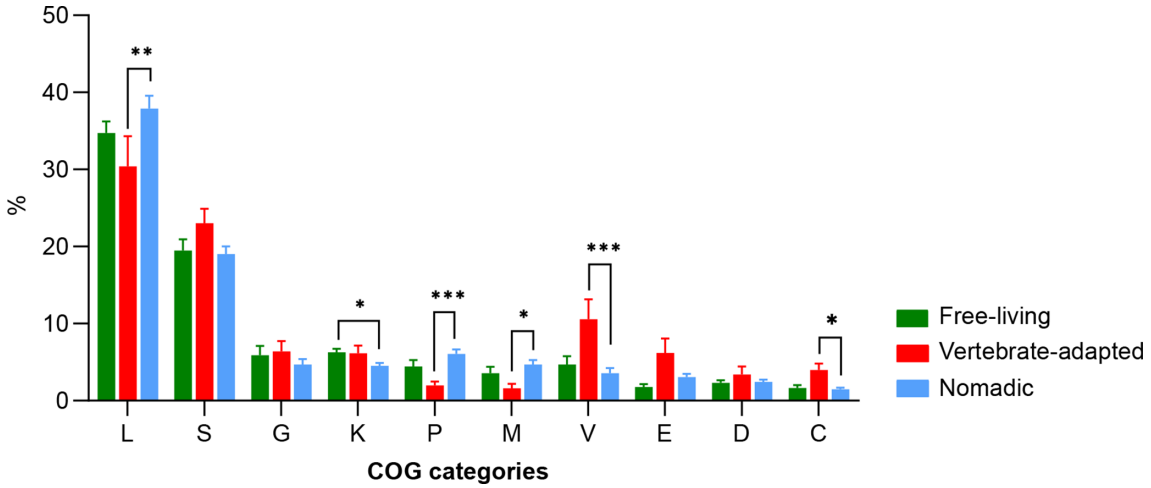

Fig. 3. Average proportion of each of the top ten COG categories in *Lactobacillaceae* plasmids per strain from various habitats. L, Replication, repair and recombination; S, Function unknown; G, Carbohydrate transport and metabolism; K, Transcription; P, Inorganic ion transport and metabolism; M, Cell wall/membrane/envelope biogenesis; V, Defence mechanisms; E, Amino acid transport and metabolism; D, Cell cycle control, cell division, chromosome partitioning; and C, Energy production and conservation. Statistical analysis was performed to compare the data across various habitats using the Kruskal–Wallis test with Dunn's post-hoc test (***P <0.001; **P <0.01; *P <0.05). Error bars represent the standard error of measurement.

The data obtained on protein families from MCL analysis were used to perform HCL to cluster the plasmids based on the presence and absence of protein families (Fig. S1). The analysis depicted three broad clusters (I, II and III) of the plasmids (Fig. 2b). Although there was no strict distinct clustering based on habitat, some habitat-wise observations could be made. For example, half of the plasmids from VAL were present in cluster I whereas the majority of the plasmids from NL and FLL (57.9% and 55.5%, respectively) were present in cluster III. On the other hand, almost equal proportions of plasmids from VAL (31.7%), NL (33.1%) and FLL (31.1%) were present in cluster II (Fig. 2c). Within the three main clusters, 38 subclusters were observed (Fig. 2b, Table S2). Of 100 strains having two or more plasmids, 63 were such that none of the plasmids from the same strain belonged to the same subcluster. This probably indicates the non-redundant genomic content of multiple plasmids within a strain in the majority of lactobacilli. A similar observation was reported for *L. lactis* IL594 where multiple plasmids in the same strain were found to be diverse and proposed to function in a synergistic manner [32].

Based on the Sørensen–Dice similarity coefficient, a measure of the extent of sharing of the protein families between plasmids, 15 plasmid pairs with a Sørensen–Dice coefficient >0.8 were identified (Table S3). Of these plasmid pairs, four having a Sørensen–Dice coefficient of 1 had only one protein family and a sequence identity of <85% with the partner plasmid with query coverage of <60%, and hence were considered dissimilar. Most of the remaining pairs (6) had plasmids from species belonging to dissimilar habitats. BLAST analysis of the plasmids found to be similar to each other in this way using the EasyFig program suggested their highly similar gene content (Fig. S2). This suggests inter-species as well as inter-habitat sharing of some of the plasmids in lactobacilli. Interestingly, none of the plasmids from these pairs belonged to VAL, underlining the uniqueness of plasmids from VAL. This further supports an evolutionary model proposing that VAL have evolved independently in the confined vertebrate-associated environment [5].

Functional annotation

To predict the putative functions associated with the plasmid-encoded genes, they were classified based on orthology using eggNOG. Based on this, 63.2% (12120) of CDSs were annotated into 20 Clusters of Orthologous Groups (COGs) functional categories. Replication, recombination and repair was the largest category with 32.5% of the CDSs followed by Function unknown (20.7% of CDSs), Carbohydrate transport and metabolism (6.2%) and Transcription (5.3%) (Fig. 3, Table S4). Furthermore, the genes were subgrouped under each eggNOG category based on their putative functions based on eggNOG, NCBI and KEGG annotations. Genes under the Function unknown category were manually evaluated for their annotations in these databases, and based on this some of the genes were binned into individual COG categories. For example, genes annotated as LPXTG-motif cell wall anchor domain protein that were found under the Function unknown category were manually placed under Cell wall/membrane/envelope biogenesis.

To test the hypothesis that plasmids play an important role in environmental adaptations in *Lactobacillaceae*, we investigated whether the plasmids encode genes with varying putative functions across different lifestyles. The fraction of genes involved in Replication, recombination and repair (L), Cell wall/membrane/envelope biogenesis (M), and Inorganic

ion transport and metabolism (P) was significantly higher (P <0.05) in the plasmids from NL as compared to VAL (Fig. 3). Furthermore, the genes associated with Defence mechanisms (V) and Energy production and conservation (C) were significantly over-represented (P <0.001, and P <0.05 respectively) in VAL compared with NL (Fig. 3). Similarly, the average proportion of genes involved in Transcription (K) was significantly higher (P <0.05) in the plasmids from FLL as compared to NL.

Replication, recombination and repair

Various subcategories under Replication, recombination and repair (L) were transposase (1928 CDSs), DNA replication (614 CDSs), resolvase (323 CDSs), integrase (256 CDSs), mobilization/conjugation (221 CDSs), DNA repair (191 CDSs), and DNA repair and recombination (99 CDSs) (Fig. 4a, Table S5). Within these, the genes encoding transposases, resolvases, DNA repair and recombination proteins, and mobilization/conjugation proteins were significantly over-represented on the plasmids from NL than those from VAL based on the strain-level average. Of these genes, resolvase and transposase genes were significantly overrepresented on the plasmids from FLL as compared to VAL at the species level (Fig. 4a, Table S6). Previous studies have demonstrated that transposases are more abundant in bacteria living in extreme environments and that they can help in the evolution of fitter phenotypes by accelerating the rate of mutations [33, 34]. In the context of lactobacilli, NL and FLL can be considered relatively extreme as these species occur in a wider range of harsh environmental conditions [5]. Thus, a higher proportion of mobile genetic elements in the plasmids of NL and FLL may give them genomic plasticity for adaptation to dynamic environments. Because NL and FLL live under more extreme conditions, they are likely to face more DNA damage, justifying the higher proportion of DNA repair genes in their plasmids. This speculation is also supported by earlier observations reporting a higher proportion of DNA repair genes in thermophilic microorganisms [35]. Furthermore, the low proportion of DNA repair genes in VAL could also be correlated with the reduced genome size of these bacteria. The correlation between a smaller proteome and loss of DNA repair genes has previously been established in several other bacteria [36].

Carbohydrate transport and metabolism

In the current study, 673 genes (6.2% of the total plasmid-encoded genes) associated with Carbohydrate transport and metabolism (G) were found. In addition, carbohydrate metabolism (332 CDSs), PTS (177 CDSs) and major facilitator superfamily transporter (131 CDSs) were identified as major subcategories (Fig. 4c, Table S5). Fourty-nine complete PTS transporters were found in the 35 plasmids across 30 strains (Table S7). Although there was no significant difference between habitats for the average proportion of PTS genes, this proportion per species was significantly higher for NL as compared to VAL and FLL. This was also reflected in the fact that the proportion of species having a complete PTS on plasmids was 22.2, 9.1 and 83.3% for FLL, VAL and NL, respectively (Table S7). NL are metabolically versatile and able to utilize a wide range of sugars. Thus, the broader distribution of complete PTS on the plasmids from NL underlines the contribution of plasmids in adaptation to diverse environments. All 30 strains having a complete PTS were homofermentative, which is consistent with the reported loss of PTS genes from the chromosomes of heterofermentative species [11]. These results suggest that plasmids can significantly contribute to the physiology of homofermentative species by allowing sugar uptake via PTSs.

The abundance of the genes from the KEGG category 'starch and sucrose metabolism' in *L. brevis*, *L. sakei*, *Secundilactobacillus paracollinoides* and *L. paracasei*, which are FLL can be clearly attributed to the presence of maltose phosphorylase and β-phospho-glucomutase in their plasmids, which allow utilization of maltose. Furthermore, a large number of maltose/moltooligosaccharide transporters were encoded on the plasmids of FLL (Table S5). Plasmids having such maltose utilization genes had *lacI* in their upstream region in the six strains of *L. sakei*, an FLL species (Fig. 4b, Table S8). The LacI transcriptional regulator was shown to be involved in maltose utilization in *L. acidophilus*, *Enterococcus* and *Bifidobacterium* [37]. One of the major habitats for FLL is fermented cereals [5], which are likely to be rich in maltose released by starch hydrolysis. This suggests that the plasmids might contribute to the efficient utilization of maltose in FLL.

The complete lactose operon (*lacTEFG*) was found on the plasmids of six strains of *L. paracasei* and one strain of *Lacticaseibacillus rhamnosus* (Table S9). A previous study has shown that this operon also helps *Lacticaseibacillus casei* in uptake of *N*-acetyllactosamine in human milk [38]. AraC was present upstream of the rhamnose utilization genes, namely MFS transporter, rhamnulokinase, L-rhamnose mutarotase, L-rhamnose isomerase and rhamnulose-1-phosphate aldolase, in five plasmids from *Ligilactobacillus salivarius* (Fig. 4b, Table S10). The AraC transcriptional regulator was shown to be involved in rhamnose metabolism in *L. acidophilus* and *L. plantarum* [39]. These results suggest that plasmids might play an important role in the growth and physiology of *Lactobacillaceae* species by allowing the uptake of various sugars.

Genes encoding major facilitator superfamily (MFS) proteins were found in significantly higher proportion on the plasmids of VAL as compared to NL and FLL (Fig. 4c). MFS proteins are involved in the transport of a large number of molecules across the membrane in response to chemiosmotic gradients. These include sugars, vitamins, antibiotics, amino acids, nucleosides and ions. In general, most of the MFS transporters in LAB are multidrug transporters [40]. Thus, their overrepresentation in the plasmids of VAL might be related to antibiotic usage in humans and other animals. Additionally, some of the MFS transporters are also involved in bile salt resistance in *Lactobacillus* [41], which is relevant for VAL.

Inorganic ion transport and metabolism

Within the Inorganic ion transport and metabolism (P) category, metal/ion transporters (435 CDSs), other transporters

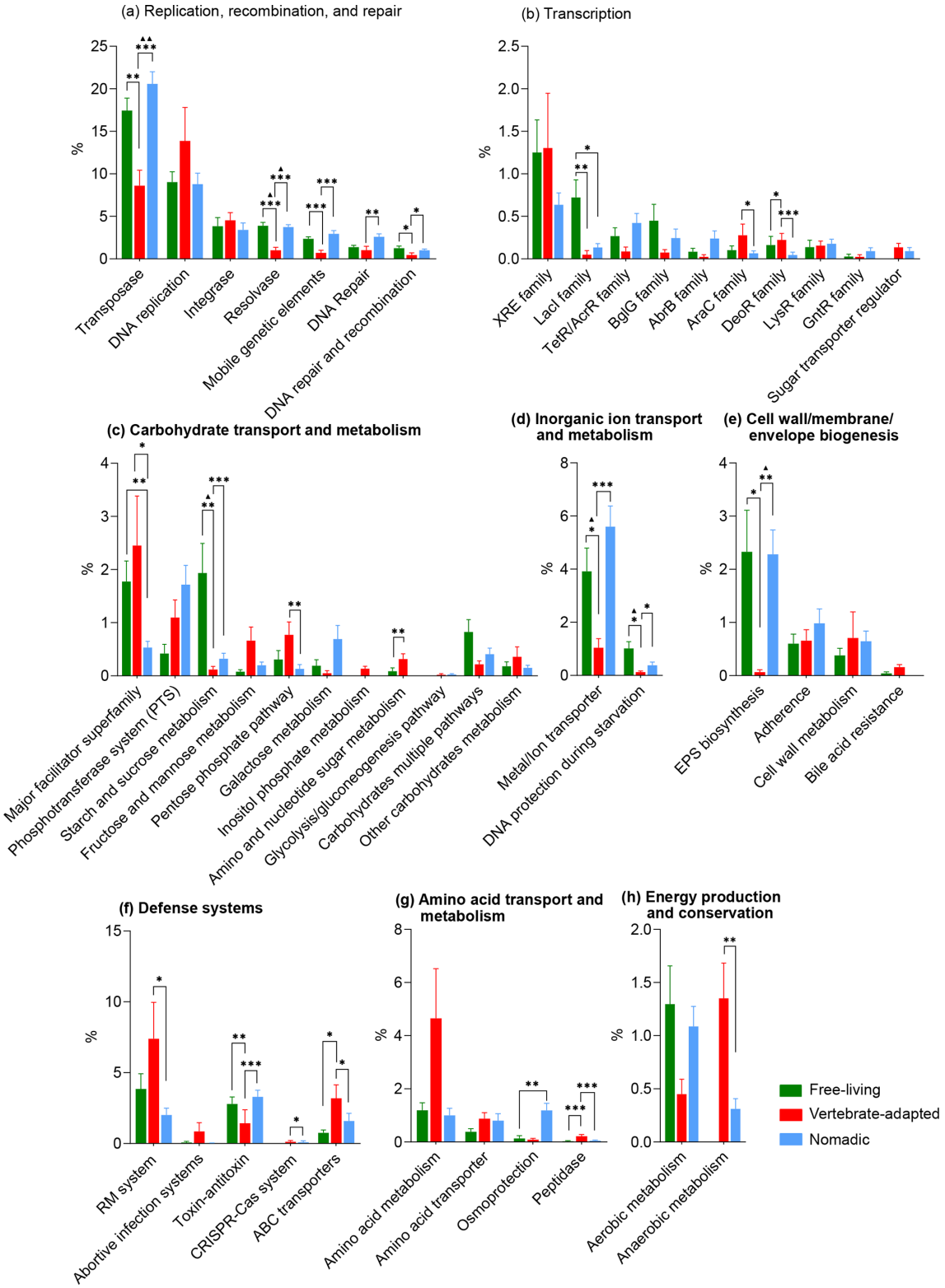

Fig. 4. Average proportion of each of the COG subcategories in *Lactobacillaceae* plasmids per strain from various habitats. Statistical analyses were performed to compare the data across various habitats using the Kruskal–Wallis test with Dunn's post-hoc test (***P <0.001; **P <0.01; *P <0.05). Error bars represent the standard error of measurement. Solid black triangles indicate a significant difference by the same test between the habitats for the species-level average (Table S6) (▲▲▲P <0.001; ▲▲P <0.01; ▲P <0.05).

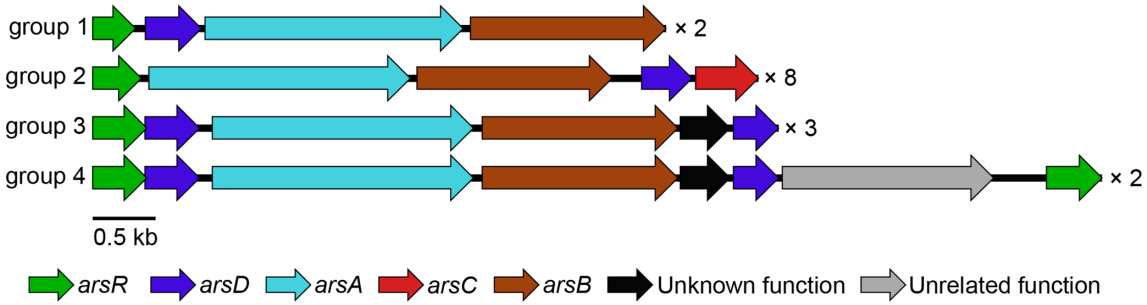

Fig. 5. Arsenic resistance gene clusters identified in 15 *Lactobacillaceae* plasmids. Numbers on the right side of the clusters indicate the number of plasmids showing that type of arrangement of the cluster. *arsR*, a regulator; *arsA*, ATPase component; *arsB*, secondary transporter; *arsD*, transcriptional regulator; and *arsC*, arsenate reductase. The plasmids represented in each group were from, group 1: *L. sakei* WiKim0074 (CP025207.1), *L. curvatus* ZJUNIT8 (CP029967.1); group 2: *L. hokkaidonensis* LOOC 260 (AP014681.1), *L. buchneri* NRRL B-30929 (CP002653.1), *L. plantarum* CLP0611 (CP019723.1), *L. brevis* ZLB004 (–) (CP021460.1), *L. plantarum* HAC01 (–) (CP029350.1), *L. plantarum* DR7 (–) (CP031319.1), *L. plantarum* ATG-K6 (CP032465.1), *L. plantarum* ATG-K8 (–) (CP032467.1); group 3: *L. plantarum* TMW 1.277 (CP017364.1), *L. paraplantarum* DSM 10667 (CP032747.1), *L. plantarum* WCFS1 (CR377166.1); group 4: *L. plantarum* MF1298 (–) (CP013150.1 and CP013151.1). Details are available in Table S11.

(87 CDSs) and DNA protection during starvation protein (*dps*) (46 CDSs) were identified as the dominant genes (Table S5). Metal/ion transporter genes were significantly more abundant in the plasmids from NL and FLL than VAL based on the strain-level average (Fig. 4d, Table S6). This difference was also reflected in species-level averages between FLL and VAL. Under this category, a large number of genes putatively involved in heavy metal resistance (HMRGs) were found, including those conferring resistance to arsenic (98 genes), cadmium (49 genes), copper (19 genes) and cobalt (13 genes) on the plasmids of 69 strains (Table S5). Such HMRGs, especially those involved in arsenic resistance, were highly enriched on the plasmids of NL such as *L. plantarum*, *L. paracasei* and *L. paraplantarum* as compared to VAL (65 and 4 genes, respectively). No comparative phenotypic data on the heavy metal resistance of *Lactobacillaceae* species belonging to various habitats are available. However, the above observation matches a finding in *Acinetobacter* and *Ornithilibacillus* where environmental isolates were more resistant to heavy metals than clinical isolates [42, 43]. Furthermore, the gene clusters involved in arsenate resistance were detected in 15 plasmids from NAL and FLL (Fig. 5, Table S11). Eight plasmids had complete clusters encoding all five proteins, namely a regulator (*arsR*), ATPase component (*arsA*), secondary transporter (*arsB*), transcriptional regulator (*arsD*) and arsenate reductase (*arsC*). The other seven plasmids lacked *arsC*; however, chromosomal *arsC* can complement such an incomplete cluster, as shown previously for *L. plantarum* WCFS1 [44].

Several genes related to uptake of potassium were found in the plasmids of *L. plantarum* and *Lactiplantibacillus pentosus,* which are NL (35 genes), and *L. brevis,* an FLL (10 genes); by contrast, none of the plasmids from VAL had this gene (Table S5). This observation is consistent with an earlier report suggesting the potassium channels are not essential in host-dependent bacteria and are found mostly in the free-living and metabolically versatile bacteria [45]. Genes encoding a manganese transport protein were found to be enriched in plasmids of NL and mostly had the *dtxR* family transcriptional regulator on the upstream side (Table S12). Other genes enriched on the plasmids from NL mainly include chloride channel protein, calcium-transporting ATPase, magnesium transporter, and Na^+/H^+ antiporter (Table S5). These proteins are involved in functions such as acid resistance, salt tolerance, metal ion homeostasis and oxidative stress tolerance [46–48], probably suggesting their contribution to adaptation of NL to dynamic environments.

The *dps* gene per strain was 10- and 3-fold more abundant on the plasmids from FLL (found in four species) and NL (two species), respectively, as compared to VAL (one species) (Fig. 4d). Dps is involved in tolerance towards oxidative, pH, heat and irradiation stresses as well as metal toxicity and attachment to abiotic surfaces during the stationary phase in *Escherichia coli* [49] and has been thought to protect *L. plantarum* GB-LP2 from oxidative damage [50]. A higher abundance of this gene in the plasmids from FLL and NL reflects the association of these strains with open and harsh environments where exposure to such stresses is comparatively higher.

Cell wall/membrane/envelope biogenesis

Within this category, heteropolysaccharide (HePS) biosynthesis (246 CDSs), adherence (89 CDSs), and cell wall biosynthesis and degradation (77 CDSs) were the major subcategories. HePS biosynthesis was represented by 124 CDSs of glycosyltransferases (GTs), 21 CDSs of tyrosine kinase modulator (*epsB*), 53 genes of polysaccharide synthesis proteins putatively involved in HePS biosynthesis, and 48 genes involved in the biosynthesis of activated sugar precursors (Fig. 4e, Table S5). The proportion of these CDSs per strain involved in HePS biosynthesis was significantly 34- and 35-fold higher in the plasmids of NL and FLL, respectively, as compared to VAL. The significant difference between VAL

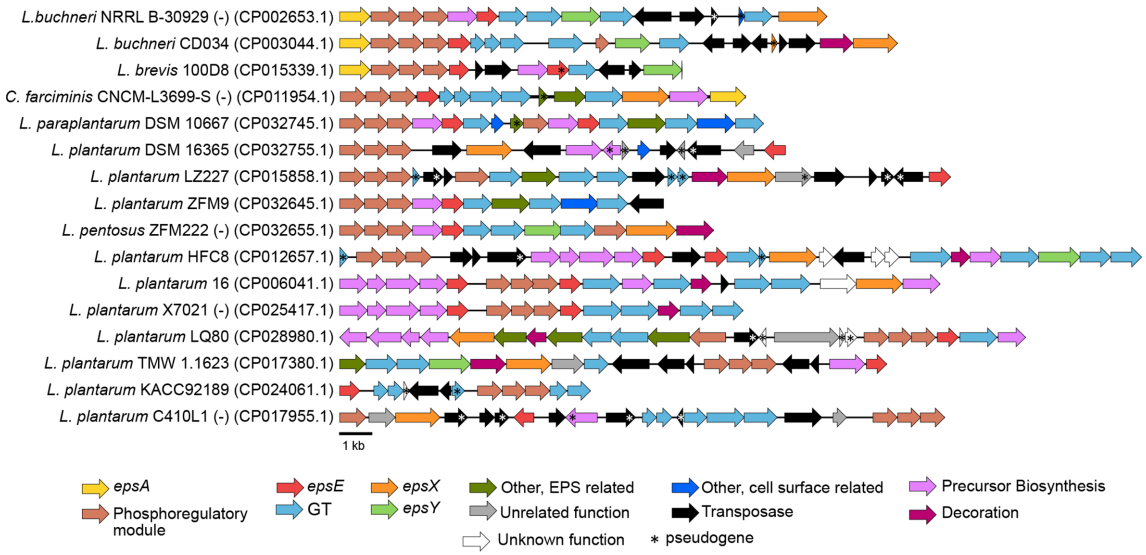

Fig. 6. EPS gene clusters identified in 16 *Lactobacillaceae* plasmids. The clusters encoded by the negative strand are represented by a minus sign in parentheses after the strain name. GenBank accession numbers of the plasmid are given in parentheses after the strain name.

and NL was also reflected in the species-level comparison (Fig. 4d, Table S6). As some of the GTs were also represented in other COG categories such as Carbohydrate transport and metabolism and Function unknown, their comprehensive identification in the *Lactobacillus* plasmids was performed via the dbCAN2 meta server (Table S13). A total of 99 GTs belonging to nine families were found on 47 plasmids with 33 plasmids having two or more GTs, with a dominance of the GT2 family (52% of the CDSs).

We further analysed the plasmids for the presence of the HePS biosynthesis gene clusters. Of 512 plasmids, HePS gene clusters were found on 16 plasmids of 16 strains represented by six species (Fig. 6, Table S14). Of these, four clusters had all the genes required for HePS biosynthesis [25]. Although a large number (12) of clusters analysed in this way were incomplete, they can contribute to HePS production in collaboration with other EPS clusters present in the chromosomes of the same strain. Such contribution of multiple HePS clusters within a strain to various properties of HePS has been demonstrated in *L. plantarum* WCFS1 [51]. The absence of *epsA* in the HePS clusters from *L. plantarum* plasmids is consistent with our earlier similar observation in *L. plantarum* chromosomes [25]. HePS is important in many VAL for colonization in the gut and in a few cases withstanding low pH and high bile concentration in the stomach [52]. The higher proportion of HePS-related genes in the plasmids from NL and FLL than VAL probably suggests that HePS might be even more important for NL and FLL in adapting to more diverse and harsh environments. This hypothesis is also supported by a recent genome analysis showing that capsular exopolysaccharide-encoding genes occur more frequently in environmental bacteria than in pathogenic bacteria [53].

A large number (143) of glycosyl hydrolases (GHs) were also detected in the plasmids and were classified within 19 CAZy families (Table S15) according to the dbCAN2 meta server. Contributions of the GH1 (16.7%) and GH13 (15.3%) families were highest followed by GH25 (13.9%). The proportion of GH genes on the plasmids was significantly higher in FLL and NL as compares to VAL. Almost all the GH68 (fructansucrases) and GH70 (glucansucrase) members, which are involved in the production of homopolysaccharides, were found only on the plasmids of *L. plantarum*. Moreover, some of the *L. plantarum* strains had multiples of such genes. For example, *L. plantarum* 16 had GH68 as well as GH70 on the same plasmid and a complete heteropolysaccharide biosynthesis gene cluster on another plasmid. Similarly, *L. plantarum* subsp. *plantarum* TS12 had two GH68, one on each of the two plasmids, and also had a GH70 on another plasmid. This is consistent with the higher occurrence of genes involved in HePS biosynthesis in the plasmids of NL as described above. It is interesting to note the absence of glucansucrase on the plasmids of VAL species but their presence in the chromosome of *Limosilactobacillus reuteri*, which is a VAL [54]. By contrast, we found an absence of glucansucrase in the chromosomes of *L. plantarum* 16 and *L. plantarum* subsp. *plantarum* TS12, which had these genes on the respective plasmids. This observation probably suggests the obligatory requirement for glucansucrase in the lifestyle adapted by limosilactobacilli and its conditional requirement in *L. plantarum*. The higher abundance of the lysozyme family (GH23 and GH25) and *N*-acetylglucosaminidase (GH3 and GH73) genes in the plasmids of FLL and NL might help them in hydrolysing the cells walls of competing bacteria in the open environment. Such a higher abundance of lysozymes encoding genes in the environmental isolates than the host-associated isolates has

also been reported in *E. coli* [55]. Similarly, a higher metabolic versatility of NL and FLL is highlighted by the abundance in their plasmids of GH1 and GH2, which enable utilization of various oligosaccharides, and GH13 and GH65, which enable utilization of starch.

We analysed whether *Lactobacillaceae* plasmids encode any of the proteins required for adhesion to the intestinal epithelium [56]. Based on annotated data and manual curation, 91 genes putatively involved in adhesion were observed in the lactobacilli plasmids. These genes mainly encoded cell wall anchor domain protein (39 CDSs), sortase (34 CDSs) and collagen-binding surface protein (16 CDSs) (Table S4). There was no significant difference between the habitats for the proportion of the plasmidome encoding these genes collectively. However, a higher number of the genes for collagen-binding surface protein were found on the plasmids of VAL (nine CDSs) as compared to NL (five CDSs) and FLL (two CDSs). This protein is involved in adhesion of the lactobacilli to the mammalian extracellular matrix [57]. The proportion of sortase and cell-wall anchor domain proteins was higher in the plasmids from NL (37 CDSs) than VAL (five CDSs). Sortases and sortase-dependent proteins have been mostly studied for their role in interactions with the host cells. However, large numbers of these genes have already been reported in NL, such as *L. plantarum*, and have been correlated with interactions with the dynamic environments [58].

Defence systems

The proportion of Defence system (V) genes was significantly higher in the plasmids of VAL species as compared to NL. The major subcategories observed were restriction-modification (R-M) system (275 CDSs), toxin–antitoxin (TA) (263 CDSs), ABC transporters (123 CDSs) and CRISPR-Cas-associated protein (14 CDSs) (Fig. 4f, Table S5). Of the four major types (I–IV) of R-M systems, the type I R-M system was found to be the most abundant with 175 CDSs. Although the fraction of type I R-M system genes was overrepresented on the plasmids of VAL, all three essential subunits (R, M, S) were present in highest numbers in FLL (nine) and NL (eight) as compared to VAL (4). The average proportion of TA genes on the plasmids was significantly higher in FLL (found in seven species) and NL (four species) as compared to VAL (four species). TA loci are involved in regulating gene expression under stressful conditions and are thought to be more important to free-living bacteria because they are more frequently under nutritional stress than host-associated bacteria [59], confirming the above observation.

The proportion of genes encoding ABC transporters was 4.2- and 3-fold higher in the plasmids of VAL than those from FLL and NL, respectively (Fig. 4f). Most of the ABC transporters under this COG category appear to be involved in the transport of antimicrobial peptides and antibiotics outside the cell (Table S5). A higher proportion of such genes in the plasmids of VAL is similar to the overrepresentation of MFS transporters as discussed above and bacteriocin-encoding genes (see section on antibacterial activity) in similar plasmids. This observation also suggests the contribution of plasmids in defence against other bacteria in the more competitive environment and in effluxing antibiotics.

Amino acid transport and metabolism

A total of 387 CDSs were found under Amino acid transport and metabolism (E) with no significant difference across the habitats. Amino acids metabolism (161 CDSs), osmoprotectant transport system (72 CDSs) and peptidase (16 CDSs) were the major subcategories (Fig. 4g, Table S5). *L. salivarius* plasmids uniquely had several genes involved in amino acid metabolism such as 3-dehydroquinate dehydratase, serine dehydratase and cysteine desulfurase (Table S5). The genes required for the conversion of methionine to cysteine, including cysteine synthase, cystathionine lyase and serine acetyltransferase, were found only in a few *L. paracasei*, *L. casei*, *L. rhamonosus* and *Limosilactobacillus fermentum* plasmids and this is consistent with this well-characterized pathway in *L. parcasei* [60].

Some plasmids encoded peptidases and their fraction was significantly higher in the plasmids from VAL as compared to NL and FLL. However, these genes were confined to the plasmids from *L. salivarius* in the VAL group. A total of 68 glycine betaine transporter genes were found on the plasmids, 64 of which belonged to NL (*L. plantarum* and *L. pentosus*, Table S4). Glycine betaine transporters are involved in providing osmoprotection to the bacteria [61], suggesting that plasmids may play a similar role in NL.

Energy production and conservation

The category Energy production and conservation (C) was significantly over-represented in the plasmids of VAL as compared to NL. We further classified the genes under this category as those associated with aerobic metabolism (98 CDSs), anaerobic metabolism (71 CDSs) and others (97 CDSs) (Fig. 4h, Table S5). Plasmids from VAL had a significantly higher proportion of genes associated with anaerobic metabolism than NL, and no FLL members had such genes present on their plasmids (Fig. 4h). Although two genes, pyruvate formate lyase and acetaldehyde dehydrogenase, were present only in *L. salivarius*, other genes such as lactate dehydrogenase, alcohol dehydrogenase and fumarate reductase were each present in two or more VAL species such as *L. salivarius*, *L. amylolyticus*, *L. amylovorus*, *L. reuteri* and *L. gasseri*. There was no significant difference between the habitats regarding the proportion of plasmids with genes involved in aerobic metabolism. However, genes encoding oxygen-utilizing enzymes, namely NADH oxidase and pyruvate oxidase, were present only in NL (*L. plantarum* and *L. paraplantarum*) and FLL (*Paucilactobacillus hokkaidonensis*, *L. brevis*, *S. paracollinoides*, *Lentilactobacillus parabuchneri*). Although *Lactobacillus* species are generally considered to be aerotolerant, because of the anaerobic gut environment, most VAL such as *L. johnsonii* and *L. gasseri* are considered to be strictly anaerobic [62, 63]. On the other hand, NL and FLL exhibit respiratory growth [64]. Thus, NADH oxidase can help FLL and NL in the regeneration of NAD^+ using oxygen as an electron acceptor. Furthermore, *L. lactis* NADH oxidase

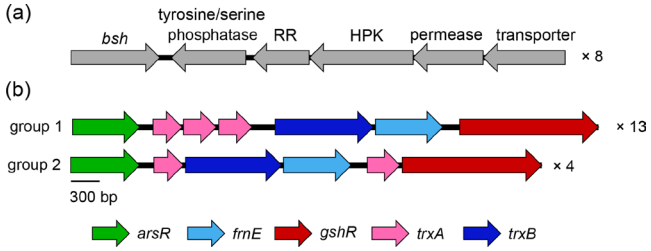

Fig. 7. Organization of the (a) bile and (b) oxidative stress resistance genes in *Lactobacillaceae* plasmids. Numbers on the right side of the clusters indicate the number of plasmids showing that type of arrangement of the cluster. The bile resistance genes (a) were: *bsh*, choloylglycine hydrolase; phosphatase, protein tyrosine/serine phosphatase; RR, two-component response regulator; HPK, signal transduction histidine kinase; permease, ABC transporter permease protein; and transporter, ABC-type multidrug transport system ATPase component. The bile resistance genes were from *L. salivarius* UCC118 (CP000234.1), *L. salivarius* CECT 5713 (CP002037.1), *L. salivarius* JCM1046 (CP007647.1), *L. salivarius* str. Ren (CP011404.1), *L. salivarius* CICC 23174 (CP017108.1), *L. salivarius* ZLS006 (CP020859.1), *L. salivarius* BCRC 12574 (CP024065.1), *L. salivarius* BCRC 14759 (CP024068.1). The oxidative stress resistance genes (b) were: *arsR*, transcriptional regulator; *trxA*, thioredoxin; *trxB*, thioredoxin reductase; *frnE*, disulfide isomerase; and *gshR*, glutathione reductase. The oxidative stress resistance genes were from, group 1: *L. plantarum* LP3 (CP017068.1), *L. plantarum* CAUH2 (–) (CP015129.2), *L. brevis* LMT1-73 (CP033887.1), *L. plantarum* LMT1-48 (CP033892.1), *L. brevis* SRCM101174 (CP021481.1), *L. brevis* 100D8 (CP015340.1), *L. brevis* KB290 (–) (AP012172.1), *L. brevis* SRCM101106 (CP021676.1), *L. plantarum* ATG-K2 (–) (CP032463.1), *L. plantarum* DSR M2 (CP022293.1), *L. sakei* WiKim0063 (CP022710.1), *L. sakei* WiKim0072 (–) (CP025138.1), *L. sakei* WiKim0074 (CP025207.1); group 2: *L. plantarum* ZFM55 (CP032360.1), *L. plantarum* ZFM9 (CP032643.1), *L. plantarum* SRCM102022 (CP021502.1), *L. plantarum* KACC 92189 (CP024060.1).

was also shown to be inactivated under anaerobic conditions or low pH [65], justifying its dispensability in VAL. Similarly, the anaerobic metabolism genes which were dominant in the plasmids from VAL can assist in maintaining redox balance under anaerobic conditions in these species.

Stress resistance

To confront a variety of harsh conditions, lactobacilli are equipped with an arsenal of stress signalling pathways [66]. To gain insight into the possible contribution of plasmids in the stress tolerance in *Lactobacillaceae*, we explored whether any of the various stress tolerance-associated genes described earlier in *Lactobacillus* [67–71] were present on the plasmids by evaluating the eggNOG, NCBI, and KEGG annotations. In total, 228 stress resistance genes related to oxidative stress resistance (104 CDSs), multiple stress resistance (*clp*, 39 CDSs), universal stress resistance protein (*uspA*, 26 CDSs), stress response regulator protein (*gls24*, 15 CDSs), stress response membrane protein (GlsB/YeaQ/YmgE family, 11 CDSs), low pH tolerance genes (13 CDSs) and bile tolerance (10 CDSs) were detected on 103 plasmids across 62 different strains (Table S16).

Within oxidative stress tolerance, genes encoding thioredoxin (TrxA, 33 CDSs), thioredoxin reductase (TrxB, 20 CDSs), glutathione reductase (GshR, 28 CDSs) and disulfide isomerase (FrnE, 20 CDSs) were detected in *Lactobacillaceae* plasmids (Table S17). The functions of *gshR*, *trxA* and *trxB* have already been established in NL and FLL [14, 72]. *FrnE* codes for a disulfide bond formation protein (DsbA) belonging to the thioredoxin family which plays a role in oxidative stress tolerance and has been characterized in *Deinococcus radiodurans* [73]. Interestingly, 17 plasmids which were solely from NL and FLL had all the four genes *trxA*, *trxB*, *gshR* and *frnE* present in the same order and had *arsR* upstream of *trxA* (Fig. 7b, Table S18), suggesting the possible involvement of this transcriptional regulator in expression of all the downstream genes associated with the oxidative stress response. Such a contribution of an ArsR family transcriptional regulator in the expression of *trxA2* upon exposure to oxidative stress has been reported in the cyanobacterium *Anabaena* sp. PCC 7120 [74]. This observation along with the collective overrepresentation of these genes in the plasmids of FLL as compared to VAL indicates the contribution of these plasmids in relieving the oxidative stress in FLL and NL, which are more exposed to aerobic conditions.

A total of 39 Clp proteinase genes (*clpL*, *clpV*, *clpX*, *clpC*) were detected with their fraction significantly higher in the plasmids of FLL as compared to those from VAL (Table S16). The primary function of the Clp complex is to degrade abnormal or misfolded proteins under stressful conditions [75, 76] with its increased expression observed in *L. plantarum* WCFS1 under heat stress conditions [77]. The total of 52 genes encoding other stress response proteins, namely UspA, Gls24 and GlsB/YeaQ/YmgE family protein, were found on the *Lactobacillaceae* plasmids of which only two ORFs were encoded by the plasmids of VAL. The gene *uspA* is considered a universal stress resistance gene and was shown to respond to various stresses such as heat shock, oxidants, DNA damage and nutrient starvation in *L. plantarum* [78]. Similarly, *gls24* encodes a stress response regulator protein and is induced under copper stress in *Enterococcus hirae* [79].

Bsh encoding choloylglycine hydrolase involved in the deconjugation of bile salts in the mammalian gut (disabling their inhibitory effect) was present only in the plasmids from VAL (eight genes) and FLL (two genes), although was confined only to one species in each habitat (Table S4). Interestingly, all these plasmids also had a two-component system encoded by histidine kinase and response regulator genes, and ABC transporter genes downstream of *bsh* but encoded by the opposite strand (Fig. 7a, Table S19). An operon containing *bsh* and the two-component system genes was shown to be involved in bile salt resistance in *L. acidophilus* [80]. In addition to these, an ABC transporter locus was shown to be present next to such a bile resistance operon in *L. gasseri* [81]. Thus, in light of these facts and previous studies showing a strong correlation between bile salt resistance genotype, phenotype and the habitat of the LAB [82], it appears that plasmids play an important role in survival at least in some lactobacilli during passage through the gastrointestinal tract.

Antibiotic resistance

Since antibiotic resistance is one of the safety concerns associated with LAB used for human consumption [83], we examined the extent of occurrence of ARGs on *Lactobacillaceae* plasmids using the Comprehensive Antibiotic Resistance Database (CARD). A total of 58 ARGs were found on 48 plasmids in 42 strains (Table S20). These ARGs encoded resistance for lincosamides (26 *lmrB*, two *lnuA* and one *lsaA*), tetracycline (15), multiple antibiotics (six), aminoglycoside (two), chloramphenicol (two), erythromycin (one), pleuromutilin (two), streptogramin (one) and vancomycin (one). The presence of *lmrB* has been previously shown in *L. gasseri* UFVCC 1091 [84] and was also found to be involved in bacteriocin secretion and immunity in *L. lactis* [85]. Campedelli *et al*. [86] reported the presence of 18 tetracycline resistance genes in the raw data on the genomes of 182 type strains of *Lactobacillus*. Thus, the presence of 15 tetracycline resistance genes in 512 plasmids in the current set appears to be a much higher proportion. Of these, eight genes coded for the efflux pumps [Tet (L) and Tcr3], whereas seven coded for ribosomal protection proteins [TetM, TetW, and TetB(P)]. The majority of the tetracycline resistance genes (eight) were found on the plasmids of VAL and this phenomenon could be related to antibiotic use as tetracycline is one of the most commonly used antibiotics in humans. This hypothesis is also supported by an earlier report suggesting that tetracycline resistance genes in the human intestinal bifidobacteria might have been acquired from intestinal pathogens [87]. Although it is important that LAB for the food and probiotic applications should be free of antibiotic resistance and the associated genes, some previous reports have shown the absence of a correlation between the antibiotic resistance phenotype and genotype [86, 88]. Thus, the mechanism of antibiotic resistance in the strains lacking related ARGs and that of the antibiotic susceptibility in spite of having the ARGs needs to be studied in detail in *Lactobacillaceae*.

ARGs are often reported to co-occur with HMRGs in many bacteria especially under the high concentration of heavy metals in the environment [89]. Thus, we analysed the co-occurrence of ARGs with HMRGs. No significant difference was found in the proportion of plasmids having ARGs with or without HMRGs (logistic regression, P <0.05, data not shown). Such a low extent of co-occurrence of these genes has also been reported earlier in *Lactobacillus* and *Lactococcus* [90].

Antibacterial activity

Lactobacilli are known to be important producers of bacteriocins which show antimicrobial activity against closely related bacteria. To identify bacteriocin-encoded ORFs in lactobacilli plasmids, the plasmid genomes were screened via the BAGEL4 database [91]. A total of 18 plasmids of 18 strains were found to encode genes for bacteriocin operons. Twelve plasmids contained complete operons collectively for class IIa (four plasmids), IIb (five plasmids) and IIc (three plasmids) bacteriocins and encoded genes for structural peptides and immunity and transport proteins (Table S21a). In total, 27 structural genes were found on the plasmids, and nine of them appear to encode novel bacteriocins (identity <85%) (Table S21b). The highest number (24) of structural genes encoded class II bacteriocins, whereas only one plasmid (*Lactobacillus gallinarum* HFD4) encoded three bacteriolysin genes (class III) (Table S8b). This is consistent with an earlier observation in *Lactobacillus* [92]. Although the presence of structural genes alone in a bacterium is not sufficient for the production and secretion of bacteriocins, complementation of such genes with the accessory genes from other strains was shown to support the secretion of the novel bacteriocins [93]. This suggests the potential application of *Lactobacillaceae* plasmids in the production of novel bacteriocins. In the context of habitats, the proportion of strains having complete bacteriocin operons as well as those having structural genes in the plasmids was highest in VAL (18.5 and 22.2%, respectively) than NL (9.5 and 13.6%, respectively) and none of the FLL had complete bacteriocin operons or structural genes. A similar observation was reported earlier with a high prevalence of bacteriocin-encoding genes in *Lactobacillus* strains isolated from animal and human guts than those isolated from other sources [92]. This also suggests that plasmids might offer a competitive advantage to the lactobacilli, mostly in the host-adapted environment.

CONCLUSION

Plasmids have been shown to be important for niche adaptation in many individual bacteria. We demonstrated this phenomenon in *Lactobacillaceae* by large-scale comparative genomic analysis of publicly available plasmid sequence data. The genomic content of the plasmids was consistent with the respective lifestyle adopted by lactobacilli, suggesting that the plasmids might enhance the niche-specific fitness of the strains. Whether the plasmid-encoded genes are as redundant as the chromosomal genes remains to be determined. Provided that numerous bacterial genes have no proven function assigned to them, at least some of the smaller plasmids with a very few genes offer a ready-made tool to understand the function of these genes. Specifically, after incorporating a suitable selection marker into the plasmids and their transformation into a model LAB, the altered phenotype can give a readout of the function of such genes. Additionally, some of the genes present on the plasmids, such as those conferring resistance to antibiotics, heavy metals, oxidative stress, bacteriocins, etc., as well as those ascribing certain biochemical traits such as utilization of specific sugars can be developed as selection markers. Considering the arsenal of the genes present on the plasmids and the commercial importance of *Lactobacillaceae* species, plasmids have huge potential in improving the properties of other strains of these bacteria. For example, the plasmids with PTSs can enable utilization of a broad range of sugars, those with the stress resistance genes can enable bacterial sustenance under industrial as well as gastrointestinal settings, those with EPS biosynthesis-related genes can improve food-related properties, those with the genes involved in redox balancing can assist in the production of value-added chemicals, and so on. Taken together, this

study not only provides a bird's eye view of the plasmidome of *Lactobacillaceae* and fundamental insights into their ecological importance but also opens new avenues to understand the biological functions and develop novel biotechnological applications of the plasmids.

Funding information
We are grateful to the Symbiosis Centre for Research and Innovation (SCRI), Symbiosis International (Deemed University) for financial support. The academic activities related to antimicrobial resistance at the host institute were supported through ERASMUS+grant 598515-EP P-1-2018-1-IN-EPPKA2-CBHE-JP.

Author contributions
Conceptualization, R.K.; Methodology, D.D. (Dimple Davray); Formal analysis, D.D. (Dimple Davray), D.D. (Dipti Deo); Investigation, D.D. (Dimple Davray); Resources, R.K.; Writing – original draft preparation, D.D. (Dimple Davray); Writing – review and editing, R.K.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Zheng J, Wittouck S, Salvetti E, Franz CMAP, Harris HMB *et al*. A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. *Int J Syst Evol Microbiol* 2020;70:2782–2858.
2. Slover CM. *Lactobacillus*: a Review. *Clin Microbiol Newsl* 2008;30:23–27.
3. Evivie SE, Huo G-C, Igene JO, Bian X. Some current applications, limitations and future perspectives of lactic acid bacteria as probiotics. *Food Nutr Res* 2017;61:1318034.
4. Hatti-Kaul R, Chen L, Dishisha T, Enshasy HE, El EH. Lactic acid bacteria: from starter cultures to producers of chemicals. *FEMS Microbiol Lett* 2018;365
5. Duar RM, Lin XB, Zheng J, Martino ME, Grenier T *et al*. Lifestyles in transition: evolution and natural history of the genus *Lactobacillus*. *FEMS Microbiol Rev* 2017;41:S27–48.
6. Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B *et al*. Comparative genomics of the lactic acid bacteria. *Proc Natl Acad Sci U S A* 2006;103:15611–15616.
7. Zheng J, Guan Z, Cao S, Peng D, Ruan L *et al*. Plasmids are vectors for redundant chromosomal genes in the *Bacillus cereus* group. *BMC Genomics* 2015;16:6.
8. Danner H, Holzer M, Mayrhuber E, Braun R. Acetic acid increases stability of silage under aerobic conditions. *Appl Environ Microbiol* 2003;69:562–567.
9. O'Sullivan O, O'Callaghan J, Sangrador-Vegas A, McAuliffe O, Slattery L *et al*. Comparative genomics of lactic acid bacteria reveals a niche-specific gene set. *BMC Microbiol* 2009;9:50.
10. Martino ME, Bayjanov JR, Caffrey BE, Wels M, Joncour P *et al*. Nomadic lifestyle of *Lactobacillus plantarum* revealed by comparative genomics of 54 strains isolated from different habitats. *Environ Microbiol* 2016;18:4974–4989.
11. Gänzle MG. Lactic metabolism revisited: metabolism of lactic acid bacteria in food fermentations and food spoilage. *Current Opinion in Food Science*, 2. Elsevier Ltd; 2015. pp. 106–117.
12. Zheng J, Ruan L, Sun M, Gänzle M. A genomic view of lactobacilli and pediococci demonstrates that phylogeny matches ecology and physiology. *Appl Environ Microbiol* 2015;81:7233–7243.
13. Folli C, Levante A, Percudani R, Amidani D, Bottazzi S *et al*. Toward the identification of a type I toxin-antitoxin system in the plasmid DNA of dairy *Lactobacillus rhamnosus*. *Sci Rep* 2017;7:1–13.
14. Zhai Z, Yang Y, Wang J, Wang G, Ren F *et al*. Complete genome sequencing of *Lactobacillus plantarum* CAUH2 reveals a novel plasmid pCAUH203 associated with oxidative stress tolerance. *3 Biotech* 2019;9:116.
15. Cui Y, Hu T, Qu X, Zhang L, Ding Z *et al*. Plasmids from food lactic acid bacteria: diversity, similarity, and new developments. *Int J Mol Sci* 2015;16:13172–13202.
16. Song Y, He Q, Zhang J, Qiao J, Xu H *et al*. Genomic variations in probiotic *Lactobacillus plantarum* P-8 in the human and rat gut. *Front Microbiol* 2018;9:893.
17. D C, RP R GF. C S. Sequence analysis of the plasmid genome of the probiotic strain *Lactobacillus paracasei* NFBC338 Which Includes the Plasmids pCD01 and pCD02. *Plasmid* 2005;54.
18. Fraunhofer ME, Geißler AJ, Behr J, Vogel RF. Comparative genomics of *Lactobacillus brevis* reveals a significant plasmidome overlap of brewery and insect isolates. *Curr Microbiol* 2019;76:37–47.
19. Fallico V, McAuliffe O, Fitzgerald GF, Ross RP. Plasmids of raw milk cheese isolate *Lactococcus lactis subsp. lactis* biovar diacetylactis DPC3901 suggest a plant-based origin for the strain. *Appl Environ Microbiol* 2011;77:6451–6462.
20. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.
21. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 2002;30:1575–1584.
22. Howe E, Holton K, Nair S, Schlauch D, Sinha R. *Mev: Multiexperiment Viewer. In: Biomedical Informatics for Cancer Research*. Boston, MA: Springer US; 2010. pp. 267–277.
23. Letunic I, Bork P. Interactive tree of life (iTOL) V3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 2016;44:W242–245.
24. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. *Bioinformatics* 2011;27:1009–1010.
25. Deo D, Davray D, Kulkarni R. A diverse repertoire of exopolysaccharide biosynthesis gene clusters in *Lactobacillus* revealed by comparative analysis in 106 sequenced genomes. *Microorganisms* 2019;7:E444
26. Dietel A-K, Merker H, Kaltenpoth M, Kost C. Selective advantages favour high genomic AT-contents in intracellular elements. In: Blokesch M (editor). *PLOS Genet*, 15(4); 2019. p. e1007778.
27. Nishida H. Comparative analyses of base compositions, DNA sizes, and dinucleotide frequency profiles in archaeal and bacterial chromosomes and plasmids. *Int J Evol Biol* 2012;2012:1–5.
28. Lawrence JG, Ochman H. Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci U S A* 1998;95:9413–9417.
29. Rocha EPC, Danchin A. Base composition bias might result from competition for metabolic resources. vol. 18, trends in genetics. *Trends Genet* 2002:291–294.
30. Almpanis A, Swain M, Gatherer D, McEwan N. Correlation between bacterial G+C content, genome size and the G+C content of associated plasmids and bacteriophages. *Microb genomics* 2018;4.
31. Kelleher P, Mahony J, Bottacini F, Lugli GA, Ventura M *et al*. The *Lactococcus lactis* pan-Plasmidome. *Front Microbiol* 2019;10:707.
32. Górecki RK, Koryszewska-Bagińska A, Gołębiewski M, Żylińska J, Grynberg M *et al*. Adaptative potential of the *Lactococcus lactis* IL594 strain encoded in its 7 plasmids. *PLoS One* 2011;6:e22238.
33. Li J, Hua ZS, Huang LN, Li J, Shi SH *et al*. Microbial communities evolve faster in extreme environments. *Sci Rep* 2014;4:6205.
34. Nelson WC, Wollerman L, Bhaya D, Heidelberg JF. Analysis of insertion sequences in thermophilic cyanobacteria: exploring the mechanisms of establishing, maintaining, and withstanding high insertion sequence abundance. *Appl Environ Microbiol* 2011;77:5458–5466.
35. Cobo-Simón M, Tamames J. Relating genomic characteristics to environmental preferences and ubiquity in different microbial taxa. *BMC Genomics* 2017;18:499.

36. Acosta S, Carela M, Garcia-Gonzalez A, Gines M, Vicens L *et al*. DNA repair is associated with information content in bacteria, archaea, and DNA viruses. *J Hered* 2015;106:644–659.
37. Zhang X, Rogers M, Bierschenk D, Bonten MJM, Willems RJL *et al*. A LacI-Family Regulator Activates Maltodextrin Metabolism of *Enterococcus faecium*. In: Manganelli R (editor). *PLoS One*, ; 2013 8; e72285.
38. Bidart GN, Rodríguez-Díaz J, Pérez-Martínez G, Yebra MJ. The lactose operon from *Lactobacillus casei* is involved in the transport and metabolism of the human milk oligosaccharide core-2 N-acetyllactosamine. *Sci Rep* 2018;8:1–12.
39. Beekwilder J, Marcozzi D, Vecchi S, de Vos R, Janssen P *et al*. Characterization of Rhamnosidases from *Lactobacillus plantarum* and *Lactobacillus acidophilus*. *Appl Environ Microbiol* 2009;75:3447–3454.
40. Lorca G, Reddy L, Nguyen A, Sun EI, Tseng J *et al*. Lactic acid bacteria: comparative genomic analyses of transport systems. *Biotechnology of Lactic Acid Bacteria Novel Applications*; 2007. pp. 73–87.
41. Elkins CA, Moser SA, Savage DC. Genes encoding bile salt hydrolases and conjugated bile salt transporters in *Lactobacillus johnsonii* 100-100 and other *Lactobacillus* species. *Microbiology* 2001;147:3403–3412.
42. Dhakephalkar PK, Chopade BA. High levels of multiple metal resistance and its correlation to antibiotic resistance in environmental isolates of *Acinetobacter*. *Biometals* 1994;7:67–74.
43. Jiang X-W, Cheng H, Zheng B-W, Li A, Lv L-X *et al*. Comparative genomic study of three species within the genus *Ornithinibacillus*, reflecting the adaption to different habitats. *Gene* 2016;578:25–31.
44. van Kranenburg R, Golic N, Bongers R, Leer RJ, de Vos WM *et al*. Functional analysis of three plasmids from *Lactobacillus plantarum*. *Appl Environ Microbiol* 2005;71:1223–1230.
45. Loukin SH, MMC K, Zhou XL, Haynes WJ, Kung C. *Microbial K+ Channels. vol. 125, Journal of General Physiology*. The Rockefeller University Press; 2005. pp. 521–527.
46. Merchant AT, Spatafora GA. A role for the DtxR family of metalloregulators in gram-positive pathogenesis. *Mol Oral Microbiol* 2014;29:1–7.
47. Yao W, Yang L, Shao Z, Xie L, Chen L. Identification of salt tolerance-related genes of *Lactobacillus plantarum* D31 and T9 strains by genomic analysis. *Ann Microbiol* 2020;70:1–14.
48. Böhmer N, König S, Fischer L. A novel manganese starvation-inducible expression system for *Lactobacillus plantarum*. *FEMS Microbiol Lett* 2013;342:37–44.
49. Goulter-Thorsen RM, Taran E, Gentle IR, Gobius KS, Dykes GA. CsgA production by *Escherichia coli* O157:H7 alters attachment to abiotic surfaces in some growth environments. *Appl Environ Microbiol* 2011;77:7339–7344.
50. Kwak W, Kim K, Lee C, Lee C, Kang J *et al*. Comparative analysis of the complete genome of *Lactobacillus plantarum* GB-LP2 and potential candidate genes for host immune system enhancement. *J Microbiol Biotechnol* 2016;26:684–692.
51. Remus DM, van Kranenburg R, van Swam II, Taverne N, Bongers RS *et al*. Impact of 4 *Lactobacillus plantarum* capsular polysaccharide clusters on surface glycan composition and host cell signaling. *Microb Cell Fact* 2012;11:149.
52. Khalil ES, Abd Manap MY, Mustafa S, Alhelli AM, Shokryazdan P. Probiotic properties of exopolysaccharide-producing *Lactobacillus* strains isolated from tempoyak. *Molecules* 2018;23:E398
53. Rendueles O, Garcia-Garcerà M, Néron B, Touchon M, Rocha EPC. Abundance and co-occurrence of extracellular capsules increase environmental breadth: implications for the emergence of pathogens. *PLoS Pathog* 2017;13:e1006525.
54. Kralj S, van Geel-Schutten GH, Dondorff MMG, Kirsanovs S, van der Maarel MJEC *et al*. Glucan synthesis in the genus *Lactobacillus*: isolation and characterization of glucansucrase genes, enzymes and glucan products from six different strains. *Microbiology* 2004;150:3681–3690.
55. Cohan FM, Kopac SM. *Microbial Genomics: E. coli Relatives Out of Doors and Out of Body*, 21. Current Biology, Cell Press; 2011. pp. R587–589.
56. Sengupta R, Altermann E, Anderson RC, McNabb WC, Moughan PJ *et al*. The role of cell surface architecture of lactobacilli in host-microbe interactions in the gastrointestinal tract. *Mediators Inflamm* 2013;2013:237921.
57. Vélez MP, De Keersmaecker SCJ, Vanderleyden J. Adherence factors of *Lactobacillus* in the human gastrointestinal tract. Vol. 276, FEMS microbiology letters. *FEMS Microbiol Lett* 2007:140–148.
58. Kleerebezem M, Boekhorst J, van Kranenburg R, Molenaar D, Kuipers OP *et al*. Complete genome sequence of *Lactobacillus plantarum* WCFS1. *Proc Natl Acad Sci U S A* 2003;100:1990–1995.
59. Pandey DP, Gerdes K. Toxin-Antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Res* 2005;33:966–976.
60. Wüthrich D, Irmler S, Berthoud H, Guggenbühl B, Eugster E *et al*. Conversion of methionine to cysteine in *Lactobacillus paracasei* depends on the highly mobile *cysK-ctl-cysE* gene cluster. *Front Microbiol* 2018;9:2415.
61. Zhao S, Zhang Q, Hao G, Liu X, Zhao J *et al*. The protective role of glycine betaine in *Lactobacillus plantarum* ST-III against salt stress. *Food Control* 2014;44:208–213.
62. Maresca D, Zotta T, Mauriello G. Adaptation to aerobic environment of *Lactobacillus johnsonii/gasseri* strains. *Front Microbiol* 2018;9:157.
63. Riedl RA, Atkinson SN, Burnett CML, Grobe JL, Kirby JR. The gut microbiome, energy homeostasis, and implications for hypertension. *Curr Hypertens Rep* 2017;19:27.
64. Zotta T, Parente E, Ricciardi A. Aerobic metabolism in the genus *Lactobacillus*: impact on stress response and potential applications in the food industry. Vol. 122, Journal of Applied Microbiology. *Blackwell Publishing Ltd* 2017:857–869.
65. Chambellon E, Rijnen L, Lorquet F, Gitton C, van Hylckama Vlieg JET *et al*. The D-2-hydroxyacid dehydrogenase incorrectly annotated PanE is the sole reduction system for branched-chain 2-keto acids in *Lactococcus lactis*. *J Bacteriol* 2009;191:873–881.
66. Papadimitriou K, Alegría Ángel, Bron PA, de Angelis M, Gobbetti M *et al*. Stress physiology of lactic acid bacteria. *Microbiol Mol Biol Rev* 2016;80:837–890.
67. Biswas S, Keightley A, Biswas I. Characterization of a stress tolerance-defective mutant of *Lactobacillus rhamnosus* LRB. *Mol Oral Microbiol* 2019;34:153–167.
68. Fontana A, Falasconi I, Molinari P, Treu L, Basile A *et al*. Genomic comparison of *Lactobacillus helveticus* strains highlights probiotic potential. *Front Microbiol* 2019;10:1380.
69. Daranas N, Badosa E, Francés J, Montesinos E, Bonaterra A. Enhancing water stress tolerance improves fitness in biological control strains of *Lactobacillus plantarum* in plant environments. In: Jogaiah S (editor). *PLoS One*, 13(1); 2018. p. e0190931.
70. Jia FF, Zheng HQ, Sun SR, Pang XH, Liang Y *et al*. Role of luxS in Stress Tolerance and Adhesion Ability in *Lactobacillus plantarum* KLDS1.0391. *Biomed Res Int* 2018;2018.
71. Lebeer S, Vanderleyden J, De Keersmaecker SCJ. Genes and molecules of lactobacilli supporting probiotic action. *Microbiol Mol Biol Rev* 2008;72:728–764.
72. Jänsch A, Korakli M, Vogel RF, Gänzle MG. Glutathione reductase from *Lactobacillus sanfranciscensis* DSM20451T: contribution to oxygen tolerance and thiol exchange reactions in wheat sourdoughs. *Appl Environ Microbiol* 2007;73:4469–4476.
73. Khairnar NP, Joe MH, Misra HS, Lim SY, Kim DH. FrnE, a cadmium-inducible protein in *Deinococcus radiodurans*, is characterized as a disulfide isomerase chaperone in vitro and for its role in oxidative stress tolerance in vivo. *J Bacteriol* 2013;195:2880–2886.
74. Ehira S, Ohmori M. The redox-sensing transcriptional regulator RexT controls expression of thioredoxin A2 in the *cyanobacterium Anabaena sp*. strain PCC 7120. *J Biol Chem* 2012;287:40433–40440.

75. Krüger E, Witt E, Ohlmeier S, Hanschke R, Hecker M. The Clp proteases of *Bacillus subtilis* are directly involved in degradation of misfolded proteins. *J Bacteriol* 2000;182:3259–3265.
76. Michel A, Agerer F, Hauck CR, Herrmann M, Ullrich J *et al*. Global regulatory impact of ClpP protease of *Staphylococcus aureus* on regulons involved in virulence, oxidative stress response, autolysis, and DNA repair. *J Bacteriol* 2006;188:5783–5796.
77. Russo P, de la Luz Mohedano M, Capozzi V, de Palencia PF, López P *et al*. Comparative proteomic analysis of *Lactobacillus plantarum* WCFS1 and ΔctsR mutant strains under physiological and heat stress conditions. *Int J Mol Sci* 2012;13:10680–10696.
78. Gury J, Seraut H, Tran NP, Barthelmebs L, Weidmann S *et al*. Inactivation of padR, the repressor of the phenolic acid stress response, by molecular interaction with USP1, a universal stress protein from *Lactobacillus plantarum*, in *Escherichia coli*. *Appl Environ Microbiol* 2009;75:5273–5283.
79. Stoyanov JV, Mancini S, Lu ZH, Mourlane F, Poulsen KR *et al*. The stress response protein Gls24 is induced by copper and interacts with the CopZ copper chaperone of *Enterococcus hirae*. *FEMS Microbiol Lett* 2010;302:69–75.
80. Pfeiler EA, Azcarate-Peril MA, Klaenhammer TR. Characterization of a novel bile-inducible operon encoding a two-component regulatory system in *Lactobacillus acidophilus*. *J Bacteriol* 2007;189:4624–4634.
81. Azcarate-Peril MA, Altermann E, Goh YJ, Tallon R, Sanozky-Dawes RB *et al*. Analysis of the genome sequence of *Lactobacillus gasseri* ATCC 33323 reveals the molecular basis of an autochthonous intestinal organism. *Appl Environ Microbiol* 2008;74:4610–4625.
82. O'Flaherty S, Briner Crawley A, Theriot CM, Barrangou R. The *Lactobacillus* bile salt hydrolase repertoire reveals niche-specific adaptation. In: Ellermeier CD (editor). *mSphere*, 3(3); 2018.
83. Barlow M. What antimicrobial resistance has taught us about horizontal gene transfer. vol. 532, methods in molecular biology. *Methods Mol Biol* 2009:397–411.
84. Rodrigues da Cunha L, Fortes Ferreira CLL, Durmaz E, Goh YJ, Sanozky-Dawes R *et al*. Characterization of *Lactobacillus gasseri* isolates from a breast-fed infant. *Gut Microbes* 2012;3:15–24.
85. Kaatz GW, McAleese F, Seo SM. Multidrug resistance in *Staphylococcus aureus* due to overexpression of a novel multidrug and toxin extrusion (MATE) transport protein. *Antimicrob Agents Chemother* 2005;49:1857–1864.
86. Campedelli I, Mathur H, Salvetti E, Clarke S, Rea MC *et al*. Genus-wide assessment of antibiotic resistance in *Lactobacillus* spp. *Appl Environ Microbiol* 2019;85
87. Wang N, Hang X, Zhang M, Liu X, Yang H. Analysis of newly detected tetracycline resistance genes and their flanking sequences in human intestinal bifidobacteria. *Sci Rep* 2017;7:6267.
88. Rozman V, Mohar Lorbeg P, Accetto T, Bogovič Matijašić B. Characterization of antimicrobial resistance in lactobacilli and bifidobacteria used as probiotics or starter cultures based on integration of phenotypic and *in silico* data. *Int J Food Microbiol* 2020;314:108388.
89. Chen J, Li J, Zhang H, Shi W, Liu Y. Bacterial heavy-metal and antibiotic resistance genes in a copper tailing dam area in northern China. *Front Microbiol* 2019;10:1–12.
90. Pal C, Bengtsson-Palme J, Kristiansson E, Larsson DGJ. Co-occurrence of resistance genes to antibiotics, biocides and metals reveals novel insights into their co-selection potential. *BMC Genomics* 2015;16:1–14.
91. de Jong A, van Hijum SAFT, Bijlsma JJE, Kok J, Kuipers OP. BAGEL: a web-based bacteriocin genome mining tool. *Nucleic Acids Res* 2006;34:W273-9–.
92. Collins FWJ, O'Connor PM, O'Sullivan O, Gómez-Sala B, Rea MC *et al*. Bacteriocin gene-trait matching across the complete *Lactobacillus* pan-genome. *Sci Rep* 2017;7:1–14.
93. Collins FWJ, Mesa-Pereira B, O'Connor PM, Rea MC, Hill C *et al*. Reincarnation of Bacteriocins From the *Lactobacillus* Pangenomic Graveyard. *Front Microbiol* 2018;9:1298.