Plasmids encode niche-specific traits in *Lactobacillaceae*

Dimple Davray, Dipti Deo and Ram Kulkarni*

Symbiosis School of Biological Sciences, Symbiosis International (Deemed University), Lavale, Pune 412115, India

***Corresponding Author**

Dr. Ram Kulkarni

ram.kulkarni@ssbs.edu.in or ramskulkarni@gmail.com;

Tel.: +91 20 28116477

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

Abstract

The species of 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 harbor plasmids that encode important genes. In this study, we performed comparative genomic analysis of the publically available data of 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-third of the species had at least one plasmid-harboring strain. Plasmid abundance and GC content were significantly lower in the host-adapted species as compared to the nomadic and free-living species. Hierarchical clustering (HCL) highlighted the distinct nature of plasmids from the nomadic and free-living species than those from the host-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 major facilitator superfamily were found to be overrepresented on the plasmids of the host-adapted species. These genomic signatures are correlated to the comparatively nutrient-depleted, stressful and dynamic environments of nomadic and free-living species and nutrient-rich and anaerobic environments of the host-adapted species. Thus, these results indicate the contribution of the plasmids in the adaptation of lactobacilli to the respective habitats. This study also underlines the potential application of these plasmids in improving the technological and probiotic properties of lactic acid bacteria.

Impact statement

The bacteria of the family *Lactobacillaceae* are present in the 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 the individual strains of *Lactobacillaceae* species encoding various traits. Extensive data of genome sequences of the 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 – host-adapted, free-living and nomadic. Comparatively lower plasmid abundance and GC content in the host-adapted species could be correlated 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 can also make them a highly useful tool in improving the probiotic, technological and food-related properties of lactobacilli.

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 31st March 2019. The dataset includes 512 nucleotide sequences of plasmids of 282 strains belonging to genus *Lactobacillus* before its reclassification into several genera (1). Details of the plasmids have been given in Table S1.

1. Introduction

Lactobacilli are gram-positive, aerotolerant, and 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 *Lactobacillaceae* family and the highly diverse nature of its species has recently resulted in reclassification of this genus into 25 genera (1). Lactobacilli have been associated with humans in a variety of ways. Their presence in the gastrointestinal tract has earned extraordinary attention due to their health-promoting properties. These bacteria are also important in the food industry for the production of fermented dairy products wherein these bacteria provide the 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 the properties such as phylogenetic analysis, source of isolation, the prevalence of detection, optimal growth temperature, substrate utilization, and environmental stress tolerance, the species of the earlier *Lactobacillus* genus have been classified as host-adapted, nomadic and free-living (5). Some of the genomic features of these species are correlated with these habitats. For example, the host-adapted lactobacilli (HAL) 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), (6), (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, *L. helvticus* DPC4571 and *L. acidophilus* NCFM were found to have genes important for the adaptation to their habitats, *viz*., dairy and gut, respectively (9).

Lastly, *L. 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 preferred to transport the sugars which are fermented via the Emden-Meyerhoff pathway. On the other hand, in the heterofermentative species, PTS are 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 favorable traits and permit them to survive in competitive surroundings (13). Many plasmids have been sequenced from the 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 *L. paracasei* NFBC338 isolated from the human gut encoded a gene involved in adherence to the human intestinal epithelial cells. Similarly, in *Lactococcus lactis*, another important bacterium of the LAB group, plant cell wall modifying enzymes were encoded by the plasmids in the plant-derived isolates (17). However, apart from these scarce strain and species-specific reports, the contribution of plasmids in niche-adaptation has not been holistically studied in the *Lactobacillaceae* family. 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 an analysis of the plasmid sequence data of the species of genus defined as *Lactobacillus*

before its division into several genera (1), available in the public domains to explore their genomic content and their possible involvement in the niche adaptation.

2. Material and Methods

2.1 General characterization of the plasmids

The coding sequences (CDSs) of the 512 plasmids were extracted from the GenBank database, where they have been automatically annotated by the NCBI against the non-redundant protein databases using BLASTP v2.2.26 (https://www.ncbi.nlm.nih.gov/nucleotide) (18). The CDSs which were not available in the database were predicted using Prodigal software (version 2.6).

2.2 Functional annotation

The functional annotation and classification of the CDSs were performed using eggNOG 4.5 databases (http://eggnogdb.embl.de/). For subgrouping of the genes under each COG category, their annotations in eggNOG4.5, NCBI and KEGG databases were considered.

2.3 Markov clustering analysis

The all-against-all bi-directional blast alignment was performed on the obtained CDSs with at least 50% amino acid identity and 50% query coverage (18). The blast output was analyzed using markov clustering (MCL) in the mclblastine v12-0678 pipeline to classify proteins into families (19). Further, Hierarchical clustering was computed in TM4 MeV Suite (20) based on the presence and absence of the protein families in plasmids. The HCL analysis was visualized in the Interactive Tree of life (ITOL) (21) by importing Newick tree from TM4 MeV Suite.

2.4 Identification of other important genes

To identify antibiotic resistance gene, initially, amino acid sequences of 2,404 antibiotic resistance (ARGs) genes were downloaded from the CARD database

(http://arpcard.mcmaster.ca). Plasmids CDSs were aligned against retrieved ARGs sequences through standalone blast (BLASTP,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 $<10^{-5}$. Further, all the putative ARGs hits were annotated querying the eggNOG 4.5 and the NCBI non-redundant protein database to verify their putative functions. Annotation from both the databases was considered to determine the putative ARGs to avoid any false results. The exopolysaccharide (EPS) gene clusters were identified as described earlier (22). Bacteriocin operons were identified using the web version of BAGEL4 (http://bagel4.molgenrug.nl) using default parameters.

2.5 Statistical analysis

GraphPad Prism 8 was used to perform statistical analysis. Comparison between the habitats was performed by ANOVA using Kruskal-Wallis test with Dunn's post hoc test.

3. Result and Discussion

3.1 Data collection and general characterization

Till March 2019, 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, this data includes 24 of the 26 genera under the *Lactobacillaceae* family. Collectively, nucleotide sequences of a total of 512 plasmids were found in the database for all these strains (Table 1, Table S1). Based on the classification by Duar et al. (5), in our dataset, the maximum number of strains (110) belonged to NL group followed by HAL (90) and FLL (43) groups. Forty strains belonged to the species for which no clear information on their possible native habitat was available; hence they were not considered for further analysis. Similarly, based on the glycolytic pathway, 216 strains belonged to homofermentative and 51 to heterofermentative species (11), (12).

The proportion of strains having plasmids was the smallest for HAL (32.2%) and the largest for FLL (76.7%) (Fig. 1a). The average number of plasmids per strain was significantly lower for HAL (1) than NL (2.3) and FLL (2.7) (Fig. 1b, Table S1). This observation is consistent with reduced chromosomal genome size in HAL as compared to NL and FLL (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 the 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 the average

plasmid size per strain across various habitats (Fig. 1c), suggesting that the low number of plasmids in HAL is not compensated by their larger size in these strains.

The average GC content of the plasmids was significantly lower for HAL (37.7%) than NL (39.7%) and FLL (39.7%). This observation is similar with the lower chromosomal GC content of the host-associated strains, which is thought to be because of non-adaptive loss of DNA repair system leading to the mutational bias toward A and T (5). The plasmid GC contents of NL and FLL were lower than the respective chromosomal GC contents; whereas, in HAL, plasmids and chromosomes had similar GC content (Fig. 1d). It has been reported that the maintenance of plasmid with higher GC content is metabolically expensive for the bacterial cells with lower chromosomal GC content and is thus evolutionary unfavourable (23). It has also been proposed that plasmids can be transferred from chromosomally GC-poor hosts or environments to relatively GC-rich hosts (24). 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 is 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 (25), (26). Thus, the comparable GC content of HAL 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 the earlier observation (27).

3.2 Hierarchical clustering (HCL) analysis

To understand the extent of similarity between the plasmids based on the genomic content, their functional grouping was performed by HCL analysis. Based on markov clustering analysis (MCL), the coding sequences (total 19,069) could be classified into a total of 3,382 protein

families, with about half of them (1,562) having a single-member. This observation is similar to that on *Lc. lactis* plasmids, where also approximately half (413 out of 885) of the protein families were singleton (28). Plasmids from HAL 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 HAL, which might have arisen during the long course of association of plasmids with these species as speculated from the GC content. The highest sharing of families of NL with HAL 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).

The data obtained on protein families from MCL analysis was 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 the habitats, certain habitat-wise observations were made. For example, half of the plasmids from HAL were present in the 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 HAL (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). Out of 100 strains having two or more plasmids, 63 strains were such that none of the plasmids from the same strain belonged to the same subcluster. This observation probably indicates the non-redundant genomic content of multiple plasmids within a strain in the majority of lactobacilli. Similar observation was reported in *Lc. lactis* IL594 where multiple plasmids in the same strain were found to be diverse and proposed to function in the synergistic manner (29).

3.3 Functional annotation

To predict the putative functions associated with the plasmid-encoded genes, they were classified on the basis of orthology using eggNOG. Based on this, 63.2% (12,065) of CDSs were annotated into 34 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% CDSs), Carbohydrate transport and metabolism (6.2% CDSs) and Transcription (5.3% CDSs) (Fig. 3, Table S3). Furthermore, the genes were subgrouped under each eggNOG category based on their putative functions based on eggNOG, NCBI, and KEGG annotations. The genes under Function unknown category were manually evaluated for their annotations in these databases and based on this some of the genes were binned into the 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 put under Cell wall/membrane/envelope biogenesis.

To test the hypothesis that plasmids play an important role in the environmental adaptations in *Lactobacillaceae*, we investigated whether the plasmids encode genes with varying putative functions across the 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 HAL (Fig. 3). Furthermore, the genes associated with Defense mechanisms (V) and Energy production and conservation (C) were significantly overrepresented ($p < 0.001$, and $p < 0.05$ respectively) in HAL in comparison 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.

3.3.1 Replication, recombination, and repair

Various subcategories under Replication, recombination, and repair (L) were transposase (1,928 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 S4). Within these, the genes encoding transposases, resolvases, DNA repair and recombination proteins, and mobilization/conjugation proteins were significantly overrepresented on the plasmids from NL and than those from HAL. Previous studies have demonstrated that transposases are more abundant in the bacteria living in the extreme environment and they can help in the evolution of fitter phenotypes by accelerating the rate of mutations (30), (31). 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, higher proportion of mobile genetic elements in the plasmids of NL and FLL may give them genomic plasticity for adaptation to the dynamic environments. Since NL and FLL live under more extreme conditions, they are likely to face more DNA damage, which can justify 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 the thermophilic microorganisms (32). Furthermore, the low proportion of DNA repair genes in HAL could also be correlated to the reduced genome size of these bacteria. The correlation between smaller proteome and loss of the DNA repair genes has already been established in several other bacteria (33).

3.3.2 Carbohydrate transport and metabolism

In the current study total of 673 genes (6.2% of the total plasmid-encoded genes) associated with Carbohydrate transport and metabolism (G) were found. Further, carbohydrates metabolism (332

CDSs), phosphotransferase system (PTS) (177 CDSs) and major facilitator superfamily transporter (131 CDSs) were identified as the major subcategories (Fig. 4c, Table S4). Fourty-nine complete PTS transporters were found in the 35 plasmids across 30 strains (Table S5a). Although there was no significant difference between the habitats for the average proportion of PTS genes per strain, this proportion per species was significantly higher for the NL as compared to HAL and FLL. This observation was also reflected in the fact that the proportion of species having complete PTS on plasmids was 22.2%, 9.1%, and 83.3% for FLL, HAL, and NL, respectively (Table S5b). 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 the 30 strains having 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 the homofermentative species by allowing sugar uptake via PTS.

The abundance of starch and sucrose metabolism genes in 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 S4). Plasmids having such maltose utilization genes had *lacI* on their upstream region in the six strains of *L. sakei* which is one of the FLL (Fig. 4b, Table S6). LacI transcriptional regulator was shown to be involved in the maltose utilization in *L. acidophilus*, *Enterococcus*, and *Bifidobacterium* (34). 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 *L. rhamnosus* (Table S7). A previous study has shown that this operon also helps *L. casei* in the uptake of N-acetyllactosamine in human milk (35). AraC was present upstream of the rhamnose utilization genes, *viz.*, MFS transporter, rhamnulokinase, L-rhamnose mutarotase, L-rhamnose isomerase, rhamnulose-1-phosphate aldolase in five plasmids from *L. salivarius* (Fig. 4b, Table S8). AraC transcriptional regulator was shown to be involved in the rhamnose metabolism in *L. acidophilus* and *L. plantarum* (36). 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 HAL 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 gradient. These include sugars, vitamins, antibiotics, amino acids, nucleosides, ions, etc. In general, majority of the MFS transporters in LAB are multidrug transporters (37). Thus, their overrepresentation in the plasmids of HAL might be related to the antibiotic usage in humans and other animals. Additionally, some of the MFS transporters are also involved in the bile salt resistance in *Lactobacillus* (38) which is relevant for HAL.

3.3.3 Inorganic ion transport and metabolism

Within the Inorganic ion transport and metabolism (P) category metal/ion transporters (435 CDSs), other transporters (87 CDSs) and DNA protection during starvation protein (*dps*) (46 CDSs) were identified as the dominant genes (Table S4). Metal/ion transporters genes were five

fold more abundant in the plasmids from NL than HAL (Fig. 4d). Under this category, a large number of genes putatively involved in heavy metal resistance (HMRGs) were found including those ascribing resistance to arsenic (98 genes), cadmium (49 genes), copper (19 genes) and cobalt (13 genes) on the plasmids of 69 strains (Table S4). Such heavy metal resistance genes, especially those involved in arsenic resistance, were highly enriched on the plasmids from NL as compared to HAL (65 and 4 genes, respectively). No comparative phenotypic data on the heavy metal resistance of the *Lactobacillaceae* species belonging to various habitats is available. However, the above observation is correlated to a finding in the case of *Acinetobacter* and *Ornithilibacillus* where in the environmental isolates were more resistant to heavy metals than the clinical isolates (39), (40). Further, the gene clusters involved in arsenate resistance were detected in 15 plasmids from NAL and FLL (Fig. 5, Table S9). Eight plasmids had complete clusters encoding all five proteins, *viz*., 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 incomplete cluster as shown earlier in the case of *L. plantarum* WCFS1 (41).

Several genes related to uptake of the potassium were found in the plasmids from NL (35 genes) and FLL (10 genes); whereas, none of the plasmids from HAL had this gene (Table S4). This observation is consistent with the earlier report suggesting the potassium channels are not essential in the host-dependent bacteria and are found mostly in the free-living and metabolically versatile bacteria (42). Genes encoding manganese transport protein were found to be enriched in plasmids of NL and mostly had *dtxR* family transcriptional regulator on the upstream side. (Table S10). Other genes enriched on the plasmids from NL mainly include chloride channel protein, calcium-transporting ATPase, magnesium transporter, and Na+/H+ antiporter (Table

S4). These proteins are involved in functions such as acid resistance, salt tolerance, metal ion homeostasis, and oxidative stress tolerance (43), (44), (45), probably suggesting their contribution to adaptation of NL to the dynamic environments.

The *dps* gene was 10 and 3 fold more abundant on the plasmids from FLL and NL, respectively, as compared to HAL (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 *E. coli* (46) and has been thought to protect *L. plantarum* GB-LP2 from the oxidative damage (47). Higher abundance of this gene in the plasmids from FLL and NL reflects the association of these strains with the open and harsh environment where exposure of such stresses is comparatively higher.

3.3.4 Cell wall/membrane/envelope biogenesis

Within this category, EPS biosynthesis (246 CDSs), adherence (89 CDSs), and cell wall biosynthesis and degradation (77 CDSs) were the major subcategories. EPS 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 EPS biosynthesis, and 48 genes involved in the biosynthesis of activated sugar precursors (Fig. 4e, Table S4). The proportion of these CDS involved in EPS biosynthesis was 34 and 35 fold higher in the plasmids of NL and FLL, respectively, as compared to HAL. 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 using dbCAN2 meta server (Table S11). A total of 99 GTs belonging to nine families were found on 47 plasmids with 33 plasmids having two or more GTs with the dominance of GT2 family (52% of the CDSs).

We further analyzed the plasmids for the presence of the ESP biosynthesis gene clusters. Out of 512 plasmids, EPS genes clusters were found on 16 plasmids of 16 strains represented by six species (Fig. 6, Table S12). Of these, four clusters had all the genes required for EPS biosynthesis (22). Although a large number (12) of clusters analyzed in this way were incomplete, they can contribute to the EPS production in collaboration with other EPS clusters present in the chromosomes of the same strain. Such contribution of multiple EPS clusters within a strain to various properties of EPS has been demonstrated in *L. plantarum* WCFS1 (48). The absence of *epsA* in the EPS clusters from *L. plantarum* plasmids is consistent with our earlier similar observation in case of *L. plantarum* chromosomes (22). EPS is important in HAL for withstanding low pH and high bile concentration in the stomach and later for colonization in the gut (49). The higher proportion of EPS related genes in the plasmids from NL and FLL than HAL probably suggests that EPS might be even more important for NL and FLL in adapting to the more diverse and harsh environments. This hypothesis is also supported by a recent genome analysis showing that the capsular exopolysaccharide encoding genes occur more frequently in the environmental bacteria than the pathogens (50).

A large number (143) of glycosyl hydrolases (GH) were also detected under in the plasmids and were classified in 19 CAZy families (Table S13) according to the dbCAN2 meta server. The contributions of GH1 (16.7%) and GH13 (15.3%) families were maximum followed by GH25 (13.9%). The proportion of the GH genes on the plasmids was significantly higher in FLL and NL as compare to HAL. Almost all the GH68 (levansucrases) and GH70 (dextransucrase) 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 multiple 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 the genes involved in the heteropolysaccharides in the plasmids of NL as described above. 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 hydrolyzing the cells walls of the 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* (51). Similarly, 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 analyzed whether *Lactobacillaceae* plasmids encode any of the proteins required for adhesion to the intestinal epithelium (52). 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 HAL (9 CDSs) as compare to NL (5 CDSs) and FLL (2 CDSs). This protein is involved in adhesion of the lactobacilli to the mammalian extracellular matrix (53). The proportion of sortase and cell-wall anchor domain proteins was higher in the plasmids from NL (37 CDSs) than HAL (5 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 to the interactions with the dynamic environments (54).

3.3.5 Defense systems

The proportion of Defense systems (V) genes was significantly higher in the plasmids of HAL 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 S4). 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 HAL, all the three essential subunits (R, M, S) were present in the highest numbers in FLL (9) and NL (8) as compared to the HAL (4). The proportion of TA genes on the plasmids was significantly higher in FLL and NL as compared to HAL. TA loci are involved in regulating the gene expression under stress conditions and are thought to be more important to the free-living bacteria since they are more frequently under nutritional stress than the host-associated bacteria (55) justifying the above observation.

The proportion of genes encoding ABC transporters was 4.2 and 3 fold higher in the plasmids of HAL 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 the antimicrobial peptides and antibiotics outside the cell (Table S4). Higher proportion of such genes in the plasmids of HAL is similar to the overrepresentation of the MFS transporters as discussed above and bacteriocin-encoding genes (see section antibacterial activity) in the similar plasmids. This observation also suggests the contribution of plasmids in defence against other bacteria in the more competitive environment and effluxing the antibiotics.

3.3.6 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 S4). *L. salivarius* plasmids uniquely had several genes involved in the amino acid metabolism such as 3-dehydroquinate dehydratase, serine dehydratase, and cysteine desulfurase (Table S4). 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 L. fermentum* plasmids and this is consistent with this well-characterized pathway in *L. parcasei* (56).

Some plasmids encoded peptidases and their fraction was significantly higher in the plasmids from HAL as compared to NL and FLL. HAL are more auxotrophic for many amino acids (57) and the above results suggest the contribution of plasmids in coping up with this disability by digesting the proteins present in the nutrient-rich environment and further uptaking them with the help of a large number (39) of amino acid transporters found in HAL. A total of 68 glycine betaine transporter genes were found on the plasmids, of which 64 belonged to NL (Table S4). Glycine betaine transporters are involved in providing osmoprotection to the bacteria (58) suggesting that plasmids may play a role in the similar function in NL.

3.3.7 Energy production and conservation

The category Energy production and conservation (C) was significantly over-represented in the plasmids of HAL 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 S4). Plasmids from HAL had a significantly higher proportion

of the genes associated with anaerobic metabolism than NL and none of FLL had such genes present on their plasmids (Fig. 4h). There was no significant difference between the habitats for the proportion in the plasmids of the genes involved in the aerobic metabolism. However, genes encoding the oxygen utilizing enzymes *viz*., NADH oxidase and pyruvate oxidase were present only in NL and FLL. Although *Lactobacillus* species are generally considered to be aerotolerant, because of the anaerobic gut environment, most of HAL such as *L. johnsonii* and *L. gasseri* are considered strictly anaerobic (59), (60). On the other hand, NL and FLL exhibit respiratory growth (61). Thus, NADH oxidase can help FLL and NL in the regeneration of NAD^+ using oxygen as an electron acceptor. Furthermore, *Lc. lactis* NADH oxidase was also shown to be inactivated under anaerobic conditions or low pH (62) justifying its dispensability in HAL. Similarly, the genes such as lactate dehydrogenase, alcohol dehydrogenase, pyruvate formate lyase, fumarate reductase, which were dominant in the plasmids from HAL can assist in the maintaining the redox balance under the anaerobic conditions in these species.

3.3.8 Stress resistance

To confront a variety of harsh conditions, lactobacilli are equipped with an arsenal of stress signalling pathways (63). To get insights 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* (64), (65), (66), (67), (68) 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 S14).

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 the *Lactobacillaceae* plasmids (Table S15). The functions of *gshR*, *trxA*, and *trxB* have already been established in NL and FLL (14), (69). *FrnE* encodes 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* (70). 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 S16) suggesting the possible involvement of this transcriptional regulator in the 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 the exposer to the oxidative stress has been reported in *Cyanobacterium Anabaena sp.* PCC 7120 (71). This observation along with the collective overrepresentation of these genes in the plasmids of FLL as compared to HAL indicates the contribution of these plasmids in relieving the oxidative stress in FLL and NL which is more exposed to the 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 HAL (Table S14). The primary function of the Clp complex is to degrade the abnormal or misfolded proteins under stressful conditions (72), (73) with its increased expression observed in *L. plantarum* WCFS1 under heat stress condition (74). The total of 52 genes encoding other stress response proteins *viz*., 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 HAL. The gene *uspA* is considered as 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* (75). Similarly, gene *gls24*, encodes a stress response regulator protein and is induced under copper stress in *Enterococcus hirae* (76).

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 HAL (8 genes) and FLL (2 genes) (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* though encoded by the opposite strand (Fig. 7a, Table S17). An operon containing *bsh* and the two-component system genes was shown to be involved in bile salt resistance in *L. acidophilus* (77). In addition to these, an ABC transporter locus was shown to be present next to such bile resistance operon in *L. gasseri* (78). Thus, in light of these facts and few other previous studies showing strong correlation between the bile salt resistance genotype, phenotype, and the habitat of the LAB (79), it appears that plasmids play an important role in survival of at least in some lactobacilli during the passage through GI tract.

3.3.9 Antibiotic resistance

Since antibiotic resistance is one of the safety concerns associated with LAB used for human consumption (80), we examined the extent of the occurrence of ARGs on the *Lactobacillaceae* plasmids using Comprehensive Antibiotic Resistance Database (CARD). A total of 58 ARGs were found on 48 plasmids in 42 *Lactobacillus* strains (Table S18). These ARGs encoded resistance for lincosamides (26 *lmrB*, two *lnuA*, and one *lsaA*), tetracycline (15), multiple antibiotics (6), aminoglycoside (2), chloramphenicol (2), erythromycin (1), pleuromutilin (2),

streptogramin (1), and vancomycin (1). The presence of *lmrB* has been previously shown in *L. gasseri* UFVCC 1091 (81) and was also found to be involved in bacteriocin secretion and immunity in *Lc. lactis* (82). Campedelli et al. (83) reported the presence of 18 tetracycline resistance genes in the raw data of 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 encoded for the efflux pumps (Tet (L) and Tcr3); whereas seven encoded for ribosomal protection proteins (TetM, TetW, and TetB(P)). Majority of the tetracycline resistance genes (eight) were found on the plasmids of HAL and this phenomenon could be related to the usage of the antibiotics as tetracycline is one of the most commonly used antibiotics in humans. This hypothesis is also supported by an earlier report suggesting that the tetracycline resistance genes in the human intestinal bifidobacteria might have been acquired from the intestinal pathogens (84). Although it is important that LAB for the food and probiotic applications should be free of antibiotic resistance and the associated genes, some of the previous reports have shown the absence of correlation between the antibiotic resistance phenotype and genotype (83), (85). Thus, the mechanism of antibiotic resistance in the strains lacking the related ARGs and that of the antibiotic susceptibility in spite of having the ARGs needs to be deeply studied 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 (86). Thus, we analyzed 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* (87).

3.3.10 Antibacterial activity

Lactobacilli are known as one of the 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 with the help of BAGEL4 database (88). 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 (4 plasmids), IIb (5 plasmids), and IIc (3 plasmids) bacteriocins and encoded genes for structural peptides and immunity and transport proteins (Table S19a). In total, 27 structural genes were found on the plasmids, and nine of them appear to encode novel bacteriocin (identity < 85%) (Table S19b). The highest number of the structural genes (24) encoded class II bacteriocins, whereas only one plasmid (*L. gallinarum* HFD4) encoded three bacteriolysin genes (class III) (Table S8b). This is consistent with the earlier observation in *Lactobacillus* (89). Although the presence of structural genes alone in a bacterium is not sufficient for the production and secretion of the bacteriocins, complementation of such genes with the accessory genes from other strains was shown to support the secretion of the novel bacteriocins (90). 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 HAL (18.5% and 22.2%, respectively) than NL (9.5% and 13.6%, respectively) and none of the FLL had the complete bacteriocin operons or structural genes. A similar observation was reported earlier with a high prevalence of bacteriocin encoding genes in the *Lactobacillus* strains isolated from animal and human gut than those isolated from other sources (89). This observation also suggests that plasmids might offer a competitive advantage to the lactobacilli, mostly in the host-adapted environment.

4. Conclusion

Plasmids have been shown to be important for the niche-adaptation in many individual bacteria. We demonstrated this phenomenon in *Lactobacillaceae* by large scale comparative genomic analysis of the publically 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 redundant as the chromosomal ones remains to be determined. Provided that numerous bacterial genes have no proved function assigned to them, at least some of the smaller plasmids with a very few genes offer a readymade tool to understand the function of these genes. Specifically, after incorporating suitable selection marker into the plasmids and their transformation into a model LAB, the altered phenotype can give 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 the selection markers. Considering the arsenal of the genes present on the plasmids and the commercial importance of the *Lactobacillaceae* species, plasmids have huge potential in improving the properties of other strains of these bacteria. For example, the plasmids with PTS can enable utilization of broad range of sugars, those with the stress resistance genes can enable bacterial sustainance under industrial as well as gastrointestinal settings, those with EPS biosynthesis related genes can improve the 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 invent novel biotechnological applications of the plasmids.

Authors and contributors

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.

Funding information

We are grateful to the Symbiosis Centre for Research and Innovation (SCRI), Symbiosis International (Deemed University) for financial support.

References

1. Zheng J, Wittouck S, Salvetti E, Franz CMAP, Harris HMB, Mattarelli P, 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(4):2782–858.
2. Slover CM. *Lactobacillus*: a Review. Clin Microbiol Newsl. 2008 Feb 15;30(4):23–7.
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 Jan 3;61(1):1318034.
4. Hatti-Kaul R, Chen L, Dishisha T, Enshasy H El. Lactic acid bacteria: from starter cultures to producers of chemicals. FEMS Microbiol Lett. 2018 Oct 1;365(20).
5. Duar RM, Lin XB, Zheng J, Martino ME, Grenier T, Pérez-Muñoz ME, et al. Lifestyles in transition: evolution and natural history of the genus *Lactobacillus*. FEMS Microbiol Rev. 2017;41(1):S27–48.
6. Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, Koonin E, et al. Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci U S A. 2006 Oct 17;103(42):15611–6.
7. Zheng J, Guan Z, Cao S, Peng D, Ruan L, Jiang D, et al. Plasmids are vectors for redundant chromosomal genes in the group. BMC Genomics. 2015 Jan 22;16(1):6.
8. Danner H, Holzer M, Mayrhuber E, Braun R. Acetic acid increases stability of silage under aerobic conditions. Appl Environ Microbiol. 2003 Jan 1;69(1):562–7.

9. O'Sullivan O, O'Callaghan J, Sangrador-Vegas A, McAuliffe O, Slattery L, Kaleta P, et al. Comparative genomics of lactic acid bacteria reveals a niche-specific gene set. BMC Microbiol. 2009;9.
10. Martino ME, Bayjanov JR, Caffrey BE, Wels M, Joncour P, Hughes S, et al. Nomadic lifestyle of *Lactobacillus plantarum* revealed by comparative genomics of 54 strains isolated from different habitats. Environ Microbiol. 2016 Dec 1;18(12):4974–89.
11. Gänzle MG. Lactic metabolism revisited: Metabolism of lactic acid bacteria in food fermentations and food spoilage. Vol. 2, Current Opinion in Food Science. Elsevier Ltd; 2015. p. 106–17.
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(20):7233–43.
13. Folli C, Levante A, Percudani R, Amidani D, Bottazzi S, Ferrari A, et al. Toward the identification of a type I toxin-antitoxin system in the plasmid DNA of dairy *Lactobacillus rhamnosus.* Sci Rep. 2017 Dec 1;7(1):1–13.
14. Zhai Z, Yang Y, Wang J, Wang G, Ren F, Hao Y. Complete genome sequencing of *Lactobacillus plantarum* CAUH2 reveals a novel plasmid pCAUH203 associated with oxidative stress tolerance. 3 Biotech. 2019 Apr 1;9(4).
15. Cui Y, Hu T, Qu X, Zhang L, Ding Z, Dong A. Plasmids from Food Lactic Acid Bacteria: Diversity, Similarity, and New Developments. Int J Mol Sci. 2015 Jun 10;16(12):13172–202.
16. Song Y, He Q, Zhang J, Qiao J, Xu H, Zhong Z, et al. Genomic Variations in Probiotic *Lactobacillus plantarum* P-8 in the Human and Rat Gut. Front Microbiol. 2018 May 8;9(MAY):893.
17. 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 Sep 15;77(18):6451–62.
18. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
19. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 2002;30(7):1575–84.
20. Howe E, Holton K, Nair S, Schlauch D, Sinha R, Quackenbush J. MeV: MultiExperiment viewer. In: Biomedical Informatics for Cancer Research. Springer US; 2010. p. 267–77.
21. 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(W1):W242–5.
22. 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(10).
23. Dietel A-K, Merker H, Kaltenpoth M, Kost C. Selective advantages favour high genomic AT-contents in intracellular elements. Blokesch M, editor. PLOS Genet. 2019 Apr 29;15(4):e1007778.

24. 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.
25. Lawrence JG, Ochman H. Molecular archaeology of the *Escherichia coli* genome. Proc Natl Acad Sci U S A. 1998 Aug 4;95(16):9413–7.
26. Rocha EPC, Danchin A. Base composition bias might result from competition for metabolic resources. Vol. 18, Trends in Genetics. Trends Genet; 2002. p. 291–4.
27. 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 Apr 1;4(4).
28. Kelleher P, Mahony J, Bottacini F, Lugli GA, Ventura M, Van Sinderen D. The *Lactococcus lactis* pan-Plasmidome. Front Microbiol. 2019 Apr 4;10(APR):707.
29. Górecki RK, Koryszewska-Bagińska A, Gołębiewski M, Żylińska J, Grynberg M, Bardowski JK. Adaptative Potential of the *Lactococcus lactis* IL594 Strain Encoded in Its 7 Plasmids. PLoS One. 2011 Jul 18;6(7):e22238.
30. Li SJ, Hua ZS, Huang LN, Li J, Shi SH, Chen LX, et al. Microbial communities evolve faster in extreme environments. Sci Rep. 2014 Aug 27;4.
31. 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 Aug 1;77(15):5458–66.
32. Cobo-Simón M, Tamames J. Relating genomic characteristics to environmental preferences and ubiquity in different microbial taxa. BMC Genomics. 2017 Jun 29;18(1):499.
33. Acosta S, Carela M, Garcia-Gonzalez A, Gines M, Vicens L, Cruet R, et al. DNA Repair Is Associated with Information Content in Bacteria, Archaea, and DNA Viruses. J Hered. 2015;106(5):644–59.
34. Zhang X, Rogers M, Bierschenk D, Bonten MJM, Willems RJL, van Schaik W. A LacI-Family Regulator Activates Maltodextrin Metabolism of *Enterococcus faecium*.PLoS One. 2013 Aug 7;8(8):e72285.
35. 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 Dec 1;8(1):1–12.
36. Beekwilder J, Marcozzi D, Vecchi S, De Vos R, Janssen P, Francke C, et al. Characterization of Rhamnosidases from *Lactobacillus plantarum* and *Lactobacillus acidophilus*. Appl Environ Microbiol. 2009 Jun;75(11):3447–54.
37. Lorca G, Reddy L, Nguyen A, Sun EI, Tseng J, Yen MR, et al. Lactic Acid Bacteria□: Comparative Genomic Analyses of Transport Systems. In: Biotechnology of Lactic Acid Bacteria Novel Applications. 2007. p. 73–87.

38. 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 Dec 1;147(12):3403–12.
39. Dhakephalkar PK, Chopade BA. High levels of multiple metal resistance and its correlation to antibiotic resistance in environmental isolates of Acinetobacter. Biometals. 1994 Jan;7(1):67–74.
40. Jiang XW, Cheng H, Zheng BW, Li A, Lv LX, Ling ZX, et al. Comparative genomic study of three species within the genus *Ornithinibacillus*, reflecting the adaption to different habitats. Gene. 2016 Mar 1;578(1):25–31.
41. Van Kranenburg R, Golic N, Bongers R, Leer RJ, De Vos WM, Siezen RJ, et al. Functional analysis of three plasmids from *Lactobacillus plantarum*. Appl Environ Microbiol. 2005 Mar 1;71(3):1223–30.
42. Loukin SH, Kuo MMC, Zhou XL, Haynes WJ, Kung C, Saimi Y. Microbial K+ channels. Vol. 125, Journal of General Physiology. The Rockefeller University Press; 2005. p. 521–7.
43. Anjali T. Merchant and Grace. A Spatafora. A role for the DtxR family of metalloregulators in gram-positive pathogenesis. Mol Oral Microbiol. 2014;29(1):1–7.
44. 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 Mar 3;70(1):1–14.
45. Böhmer N, König S, Fischer L. A novel manganese starvation-inducible expression system for *Lactobacillus plantarum.* FEMS Microbiol Lett. 2013 May;342(1):37–44.
46. 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 Oct;77(20):7339–44.
47. Kwak W, Kim K, Lee C, Lee C, Kang J, Cho K, 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 Apr 28;26(4):684–92.
48. Remus DM, van Kranenburg R, van Swam II, Taverne N, Bongers RS, Wels M, et al. Impact of 4 *Lactobacillus plantarum* capsular polysaccharide clusters on surface glycan composition and host cell signaling. Microb Cell Fact. 2012 Nov 21;11(1):149.
49. Khalil ES, Manap MYA, Mustafa S, Alhelli AM, Shokryazdan P. Probiotic properties of exopolysaccharide-producing *Lactobacillus* strains isolated from tempoyak. Molecules. 2018;23(2).
50. 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(7):e1006525.
51. Cohan FM, Kopac SM. Microbial genomics: *E. coli* relatives out of doors and out of body. Vol. 21, Current Biology. Cell Press; 2011. p. R587–9.

52. Sengupta R, Altermann E, Anderson RC, McNabb WC, Moughan PJ, Roy NC. The role of cell surface architecture of lactobacilli in host-microbe interactions in the gastrointestinal tract. Mediators Inflamm. 2013;2013.
53. 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. p. 140–8.
54. Kleerebezem M, Boekhorst J, Van Kranenburg R, Molenaar D, Kuipers OP, Leer R, et al. Complete genome sequence of *Lactobacillus plantarum* WCFS1. Proc Natl Acad Sci U S A. 2003 Feb 18;100(4):1990–5.
55. Deo Prakash Pandey and Kenn Gerdes. Toxin–antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. Nucleic Acids Research. 2005. p. 966–976.
56. Wüthrich D, Irmler S, Berthoud H, Guggenbühl B, Eugster E, Bruggmann R. Conversion of Methionine to Cysteine in *Lactobacillus paracasei* Depends on the Highly Mobile cysK-ctl-cysE Gene Cluster. Front Microbiol. 2018 Oct 17;9(OCT):2415.
57. Claesson MJ, Van Sinderen D, O'Toole PW. The genus *Lactobacillus* - A genomic basis for understanding its diversity. FEMS Microbiol Lett. 2007;269(1):22–8.
58. Zhao S, Zhang Q, Hao G, Liu X, Zhao J, Chen Y, et al. The protective role of glycine betaine in *Lactobacillus plantarum* ST-III against salt stress. Food Control. 2014 Oct 1;44:208–13.
59. Maresca D, Zotta T, Mauriello G. Adaptation to Aerobic Environment of *Lactobacillus johnsonii*/*gasseri* Strains. Front Microbiol. 2018 Feb 9;9(FEB):157.
60. Ruth A. Riedl, Samantha N. Atkinson, Colin M.L. Burnett, Justin L. Grobe and JRK. The Gut Microbiome, Energy Homeostasis, and Implications for Hypertension. Curr Hypertens Rep. 2017;19(4).
61. 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. p. 857–69.
62. Chambellon E, Rijnen L, Lorquet F, Gitton C, Van Hylckama Vlieg JET, Wouters JA, 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 Feb 1;191(3):873–81.
63. Papadimitriou K, Alegría Á, Bron PA, de Angelis M, Gobbetti M, Kleerebezem M, et al. Stress Physiology of Lactic Acid Bacteria. Microbiol Mol Biol Rev. 2016 Sep;80(3):837–90.
64. Biswas S, Keightley A, Biswas I. Characterization of a stress tolerance-defective mutant of *Lactobacillus rhamnosus* LRB. Mol Oral Microbiol. 2019 Aug 1;34(4):153–67.
65. Fontana A, Falasconi I, Molinari P, Treu L, Basile A, Vezzi A, et al. Genomic comparison of *Lactobacillus helveticus* strains highlights probiotic potential. Front Microbiol. 2019 Jun 26;10:1380.
66. 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. Jogaiah S, editor. PLoS One. 2018 Jan 5;13(1):e0190931.

67. Jia FF, Zheng HQ, Sun SR, Pang XH, Liang Y, Shang JC, et al. Role of luxS in Stress Tolerance and Adhesion Ability in *Lactobacillus plantarum* KLDS1.0391. Biomed Res Int. 2018;2018.
68. Lebeer S, Vanderleyden J, De Keersmaecker SCJ. Genes and Molecules of Lactobacilli Supporting Probiotic Action. Microbiol Mol Biol Rev. 2008;72(4):728–64.
69. Jänsch A, Korakli M, Vogel RF, Gänzle MG. Glutathione reductase from *Lactobacillus sanfranciscensis* DSM20451 T: Contribution to oxygen tolerance and thiol exchange reactions in wheat sourdoughs. Appl Environ Microbiol. 2007 Jul;73(14):4469–76.
70. 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(12):2880–6.
71. 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 Nov 23;287(48):40433–40.
72. 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 Jun;182(11):3259–65.
73. Michel A, Agerer F, Hauck CR, Herrmann M, Ullrich J, Hacker 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 Aug 15;188(16):5783–96.
74. Russo P, Mohedano M de la L, Capozzi V, de Palencia PF, López P, Spano G, et al. Comparative proteomic analysis of *Lactobacillus plantarum* WCFS1 and ΔctsR mutant strains under physiological and heat stress conditions. Int J Mol Sci. 2012 Sep;13(9):10680–96.
75. Gury J, Seraut H, Tran NP, Barthelmebs L, Weidmann S, Gervais P, 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 Aug 15;75(16):5273–83.
76. Stoyanov J V., Mancini S, Lu ZH, Mourlane F, Poulsen KR, Wimmer R, 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 Jan 1;302(1):69–75.
77. 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 Jul 1;189(13):4624–34.
78. Azcarate-Peril MA, Altermann E, Goh YJ, Tallon R, Sanozky-Dawes RB, Pfeiler EA, 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 Aug 1;74(15):4610–25.

79. O'Flaherty S, Briner Crawley A, Theriot CM, Barrangou R. The *Lactobacillus* Bile Salt Hydrolase Repertoire Reveals Niche-Specific Adaptation. Ellermeier CD, editor. mSphere. 2018 May 30;3(3).
80. Barlow M. What antimicrobial resistance has taught us about horizontal gene transfer. Vol. 532, Methods in molecular biology. Methods Mol Biol; 2009. p. 397–411.
81. da Cunha LR, Ferreira CLLF, Durmaz E, Goh YJ, Sanozky-Dawes RB, Klaenhammer TR. Characterization of *Lactobacillus gasseri* isolates from a breast-fed infant. Gut Microbes. 2012;3(1).
82. 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 May 1;49(5):1857–64.
83. Campedelli I, Mathur H, Salvetti E, Clarke S, Rea MC, Torriani S, et al. Genus-wide assessment of antibiotic resistance in *Lactobacillus spp.* Appl Environ Microbiol. 2019 Jan 1;85(1).
84. 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 Dec 1;7(1).
85. 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 Feb 2;314:108388.
86. 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.
87. 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):1–14.
88. Anne de Jong, Sacha A. F. T. van Hijum*, Jetta J. E. Bijlsma JK and OPK. BAGEL: a web-based bacteriocin genome mining tool. Nucleic Acids Research. 2006.
89. Collins FWJ, O'Connor PM, O'Sullivan O, Gómez-Sala B, Rea MC, Hill C, et al. Bacteriocin Gene-Trait matching across the complete *Lactobacillus* Pan-genome. Sci Rep. 2017 Dec 1;7(1):1–14.
90. Collins FWJ, Mesa-Pereira B, O'Connor PM, Rea MC, Hill C, Ross RP. Reincarnation of Bacteriocins From the *Lactobacillus* Pangenomic Graveyard. Front Microbiol. 2018 Jul 2;9:1298.

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)	19,064
Number range of coding sequences per plasmid	1-809

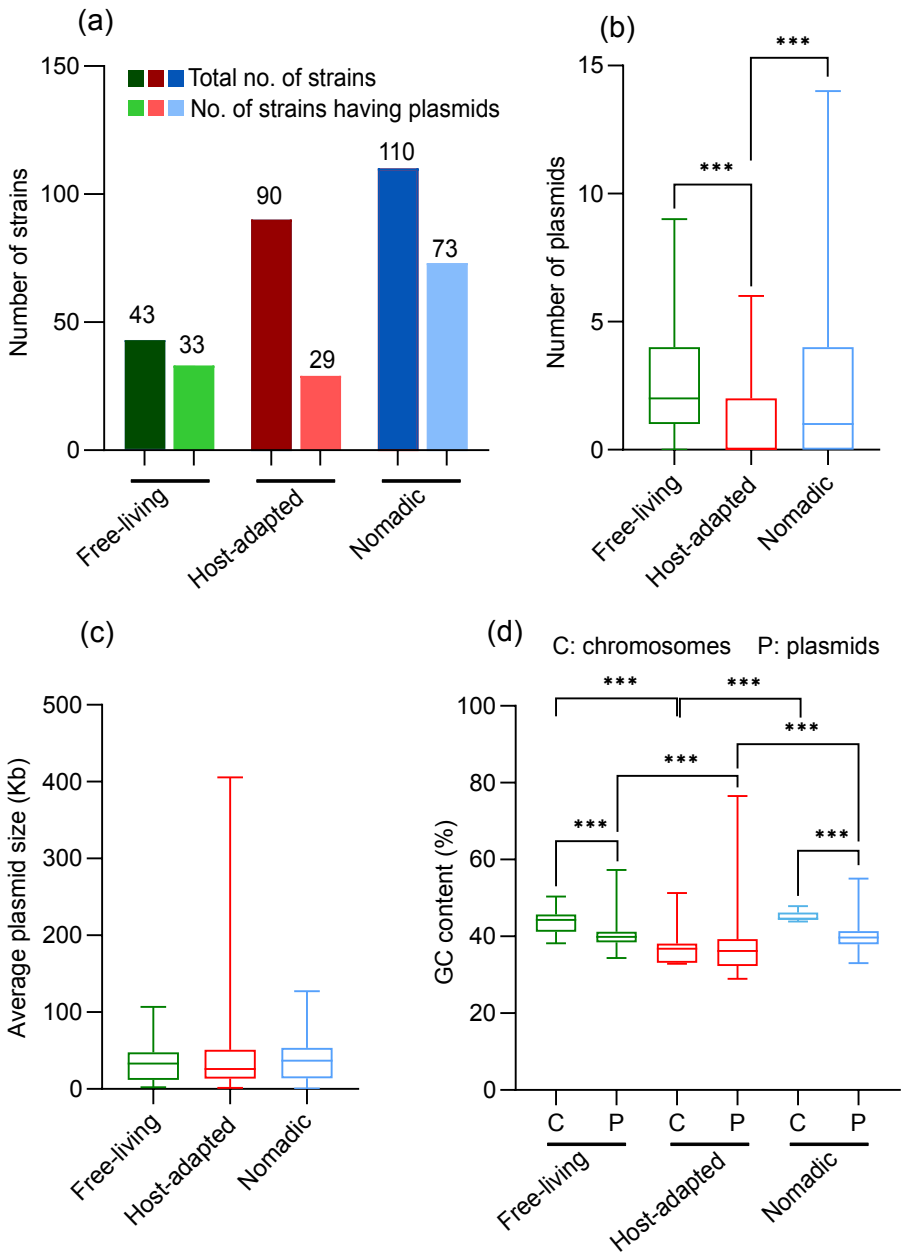

Figure 1: General features of the plasmids in the *Lactobacillaceae* species from various habitats indicating the abundance of plasmids across habitats (a), the number of plasmids per strain (b), average size of the plasmids (c), and GC content (d). The statistical analyses were performed to compare the data across various habitats using Kruskal-Wallis test with Dunn's post hoc test (***, $p<0.001$; **, $p< 0.01$; *, $p< 0.05$).

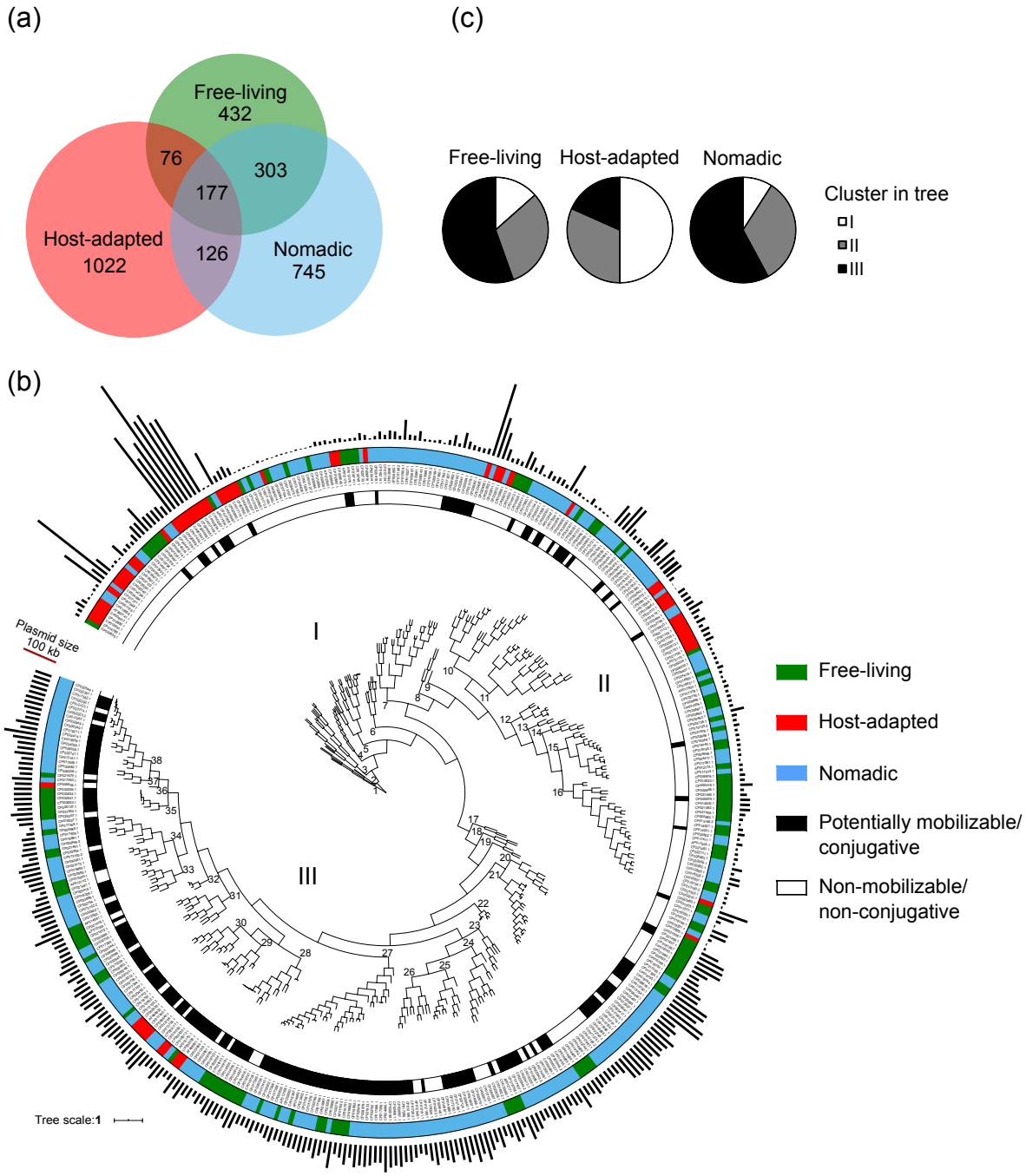

Figure 2: Hierarchical clustering (HCL) analysis of the *Lactobacillaceae* plasmids. (a) Sharing of protein families encoded by plasmids as identified by markov clustering (MCL) across lactobacilli from various habitats. (b) HCL tree displaying the clustering of the plasmids based on their protein family composition. Black bars on the outermost periphery represent plasmid genome size. The plasmids having mobilization/conjugation genes were identified based the eggNOG annotations detailed in section 2.2.1 (c) Pie chart representing the proportion of the plasmids from each habitat under clusters I, II, and III of the tree shown in panel (b).

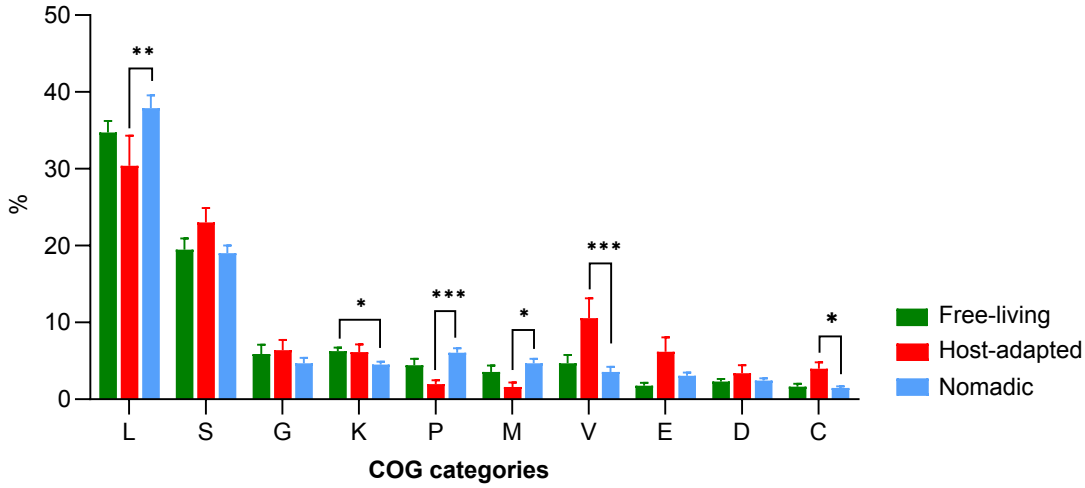

Figure 3: Average proportion of each of the top-ten COG categories in the *Lactobacillaceae* plasmids per strain from various habitats. L, Replication, repair and recombination; S, Function unknown; G, Carbohydrate transport and metabolism; K, Transcription; P, Inorganic ion transport and metabolism;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

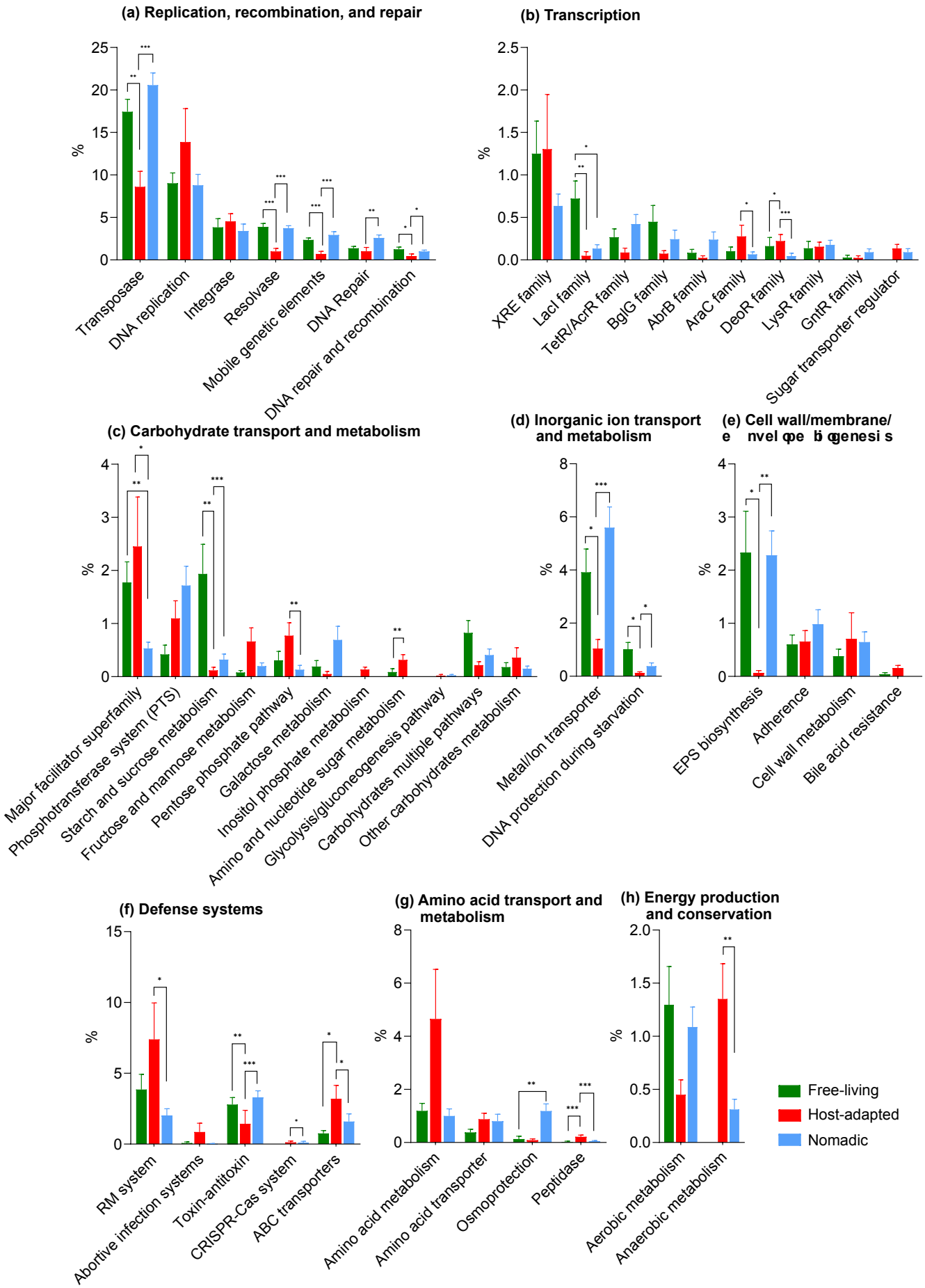

Figure 4: Average proportion of each of the COG subcategories in the *Lactobacillaceae* plasmids per strain from various habitats. The statistical analyses were performed to compare the data across various habitat using the Kruskal-Wallis test with Dunn's post hoc test (***, $p<0.001$; **, $p<0.01$; *, $p< 0.05$). The error bars represent standard error of measurement.

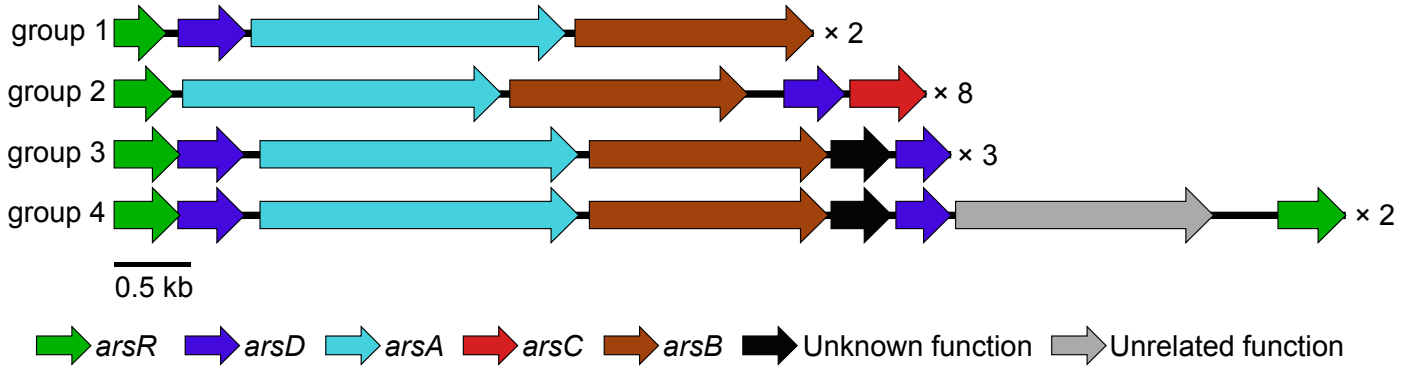

Figure 5: Arsenic resistance gene clusters identified in 15 *Lactobacillaceae* plasmids. Numbers on the right side of the clusters indicate the number of plasmids showing that types of arrangement of the cluster. *arsR*, a regulator; *arsA*, ATPase component; *arsB*, secondary transporter; *arsD*, transcriptional 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 the supplementary Table S9.

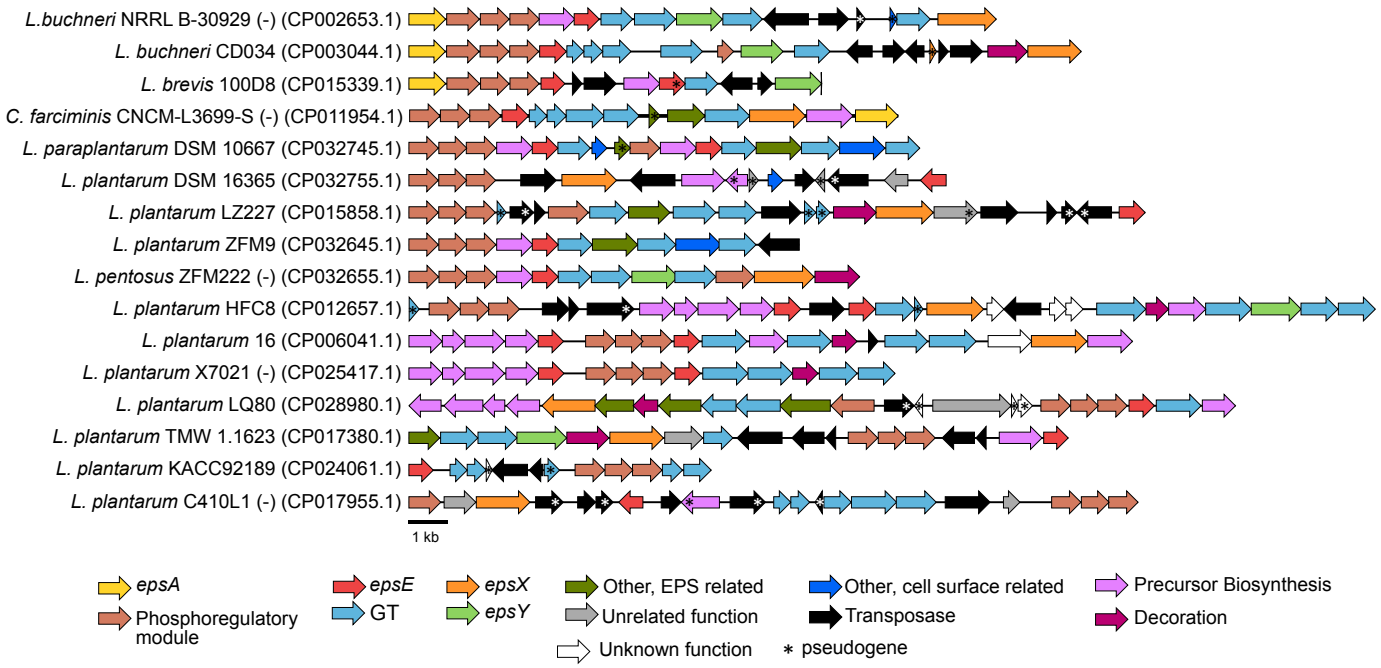

Figure 6: EPS gene clusters identified in 16 *Lactobacillaceae* plasmids. The clusters encoded by the negative strand are represented by the negative sign in the parentheses after the strain name. The Genbank accession number of the plasmid has been mentioned in the parentheses after the strain name.

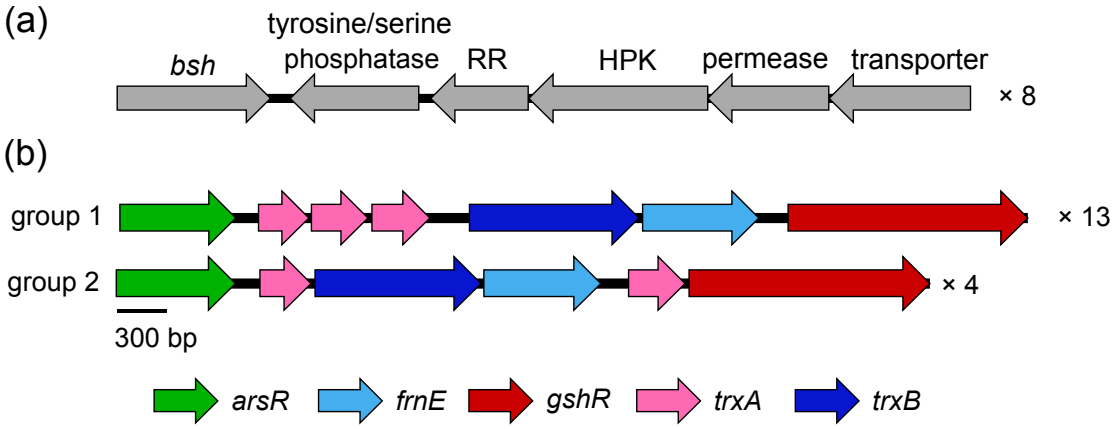

Figure 7: Organization of the (a) bile and (b) oxidative stress resistance genes in *Lactobacillacea*e plasmids. Numbers on the right side of the clusters indicate the number of plasmids showing that types of arrangement of the cluster. The bile resistance genes (a) were *bsh*, choloylglycine hydrolase; 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 AT Pase 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).