CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Functional genomic insights into Floricoccus penangensis ML061-4 isolated from leaf surface of Assam tea
Authors	Rungsirivanich, Patthanasak;Parlindungan, Elvina;Mahony, Jennifer;Supandee, Witsanu;Thongwai, Narumol;van Sinderen, Douwe
Publication date	2025-01-23
Original Citation	Rungsirivanich, P, Parlindungan, E, Mahony, J, Supandee, W, Thongwai, N & van Sinderen, D 2025, 'Functional genomic insights into Floricoccus penangensis ML061-4 isolated from leaf surface of Assam tea', Scientific Reports, vol. 15, 2951, pp. 1-14. https://doi.org/10.1038/s41598-025-86602-x
Type of publication	Article (peer-reviewed)
Link to publisher's version	https://www.scopus.com/pages/publications/85216717775 - 10.1038/s41598-025-86602-x
Rights	© 2025, the Author(s). This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. - https://creativecommons.org/licenses/by/4.0/
Download date	2026-11-15 16:16:14
Item downloaded from	https://hdl.handle.net/10468/19496

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

scientific reports

Check for updates

OPEN

Functional genomic insights into *Floricoccus penangensis* ML061-4 isolated from leaf surface of Assam tea

Patthanasak Rungsirivanich[1,2,3], Elvina Parlindungan[3,4], Jennifer Mahony[3,4], Witsanu Supandee[5], Narumol Thongwai[1]✉ & Douwe van Sinderen[3,4]✉

Floricoccus penangensis **ML061-4 was originally isolated from the leaf surface of an Assam tea plant (*Camellia sinensis* var. *assamica*) from Northern Thailand. To assess the functions encoded by the *F. penangensis* ML061-4 genome, gene identification and annotation were undertaken by in silico analysis. The complete genome of *F. penangensis* ML061-4 consists of single chromosome of 2,159,127 base pairs, containing a GC content of 33.2% and encompassing 2049 predicted protein-encoding genes. A total of 1195 genes (58.0%) in the *F. penangensis* ML061-4 genome have assignable functions based on BlastKOALA analysis. Furthermore, 1235 genes (59.9%) were classified into six KEGG functional categories with 187 associated pathways, while 1419 genes (68.8%) were assigned a putative function by the Clusters of Orthologous Groups (COGs) database. The ML061-4 genome was evaluated for genes associated with complex carbohydrate metabolism, bacterial adhesion, virulence factors, pathogenicity, bacteriophages, antiviral defence systems as well as toxin- and antibiotic-resistance associated genes, and genes involved in toxin production, secondary metabolite biosynthesis and xenobiotics biodegradation. The obtained results support the notion of *F. penangensis* ML061-4 being safe for biotechnological and food industry purposes. This is the first report outlining functional genomic insights regarding a member of the genus *Floricoccus*.**

Floricoccus penangensis, a lactic acid bacterium, was first described by Chuah et al.[1]. It has been classified belonging to the family *Streptococcaceae* originally isolated from flowers of *Durio zibethinus*. The genus *Floricoccus* was explicitly separated from the genera *Lactococcus* and *Streptococcus*, with which it exhibits 94–99% and 49–76% sequence identity, as based on 16S rRNA and the housekeeping *rpoB* genes, respectively[1,2]. *Floricoccus* members have not only been found in various environmental locations, such as plants, flowers and foods[1,2], but have also been associated with potential eukaryotic hosts, such as animals (as a component of the fecal/gut microbiome)[3] and humans (respiratory secretions)[4]. Despite their frequent isolation, the precise ecological roles and functions of members of the genus *Floricoccus* are still largely unknown.

Various members of the genera *Lactococcus* and *Streptococcus*, which are phylogenetically closely related to *Floricoccus*, are extensively used for the production of fermented foods, such as yoghurt, cheese, kefir and lactic butter[5,6]. For example, *Lactococcus raffinolactis* is widely used as a starter culture in fish, meat, vegetables and milk fermentations[7], while *Streptococcus thermophilus* strains are employed in the production of yogurt and certain cheeses and have also been reported to exhibit probiotic potential[8]. However, there have been reports on lactococcal and streptococcal isolates acting as pathogenic bacteria or opportunistic pathogens in humans and animals. For instance, *Streptococcus suis* can act as the causative agent of arthritis, meningitis and septicemia[9], while *Lactococcus lactis* has been reported to very occasionally cause endocarditis and liver abscesses in humans. Moreover, *Lactococcus garvieae* is described as a fish pathogen, whereas *Lactococcus petauri* is associated with abscesses in animals so also exerts pathogenic potential[10].

F. penangensis strain ML061-4 is a Gram-positive, coccus-shaped, non-endospore forming, non-motile and catalase-negative bacterium. This strain has been reported to exhibit properties such as acid-base tolerant, ability to aggregate and adhere to human epithelial cells, which may be desirable in a probiotic exploitation context[2].

[1]Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand. [2]Community Development Department, Ministry of Interior, Bangkok 10210, Thailand. [3]School of Microbiology, University College Cork, Cork T12 YT20, Ireland. [4]APC Microbiome Ireland, University College Cork, Cork T12 TP07, Ireland. [5]Engineering Science Classroom, King Mongkut's University of Technology Thonburi, Bangkok 10150, Thailand. ✉email: nthongw@hotmail.com; d.vansinderen@ucc.ie

In addition, *F. penangensis* ML061-4 may act as a biological control agent in Assam tea (*Camellia sinensis* var. *assamica*) cultivation and during tea leaf fermentation through its ability to produce organic acids as a result of fermentation of a variety of carbohydrates, such as glucose, fructose, mannose, *N*-acetyl-glucosamine, arbutin, aesculin, maltose, sucrose, trehalose, starch and glycogen[1,11].

Although *Floricoccus* has been described as a novel genus, including species such as *F. penangensis* and *F. tropicus*, and being related to streptococci and lactococci, floricoccal genomes have not been extensively analyzed. The genome sequence of *F. penangensis* JCM 31735^T was previously reported by Chuah et al.[1]. using an Illumina sequencing platform. Illumina sequencing allows low-cost and rapid genome analysis, however, this sequencing technology typically results in the generation of multiple contigs, thus not leading to a fully assembled, high quality genome sequence since it is based on the assembly of short reads[12,13]. A recent study by Rungsirivanich et al.[14]. described a completed genome sequence of *F. penangensis* ML061-4 using a combination of Illumina and Pacific Biosciences (PacBio) sequencing platforms. PacBio sequencing generates long reads and facilitates assembly of reads into fully circular chromosomes (and plasmids where relevant)[15].

In silico analysis has been applied for gene predictions to better understand the genetic and metabolic potential as well as to predict functional characteristics of strains[16,17]. In the present study, we analyze and characterize the genome of *F. penangensis* ML061-4 and perform an extensive gene identification and annotation effort, combined with comparative genome analysis including genomes from floricoccal, lactococcal and streptococcal groups in order to validate the functionality and corroborate safety aspects of *F. penangensis* ML061-4.

Results and discussion

General genome features of *F. penangensis* ML061-4

In a previous study, the complete genome of *F. penangensis* ML061-4 was reported to constitute a single contig of 2,159,127 base pairs with a 33.2% GC content, and to encompass 2,134 predicted genes comprising 19 rRNAs, 63 tRNAs, 2 ncRNAs, 16 pseudogenes and 2,049 protein-encoding genes[14]. The genome size of *F. penangensis* ML061-4 is within the range of those determined for three other *Floricoccus* species (2.05–2.16 Mbp), or to those of *Lactococcus* (1.96–2.42 Mbp) and *Streptococcus* (2.15–2.47 Mbp) species (Table 1). Genomes belonging to members of the floricoccal group exhibit a lower % G + C content (approximately 33%) when compared to

Feature	*F. penangensis* ML061-4	*F. penangensis* JCM 31735^T	*F. tropicus* JCM 31733^T	*L. plantarum* DSM 20686^T	*L. lactis* DSM 20481^T	*L. taiwanensis* NBRC 109049^T	*S. salivarius* NCTC 8618^T	*S. hillyeri* DSM 107591^T	*S. ovuberis* CCUG 69612^T
Size (bp)	2,159,127	2,049,722	2,157,756	1,989,191	2,421,471	1,956,764	2,188,923	2,147,614	2,467,419
ANI values[a]	Not relevant	97.90	93.70	69.98	69.84	69.34	70.16	69.95	68.84
G + C content (mol%)	33.2	33.1	33.0	36.8	35.1	38.7	40.1	39.5	42.8
No. of plasmids	0	NA	NA	NA	0	NA	0	NA	NA
Genes (total)	2134	2091	2243	1914	2356	1908	2055	2079	2353
CDSs (total)	2049	2032	2179	1877	2268	1846	1965	2020	2304
No. of rRNAs	19	5	6	3	19	7	18	4	5
No. of tRNAs	63	51	55	30	65	51	68	50	40
No. ncRNAs	2	3	3	4	4	4	4	4	4
Pseudogene	16	25	16	44	12	21	58	63	143
Contigs	1	31	26	43	1	17	1	83	48
Origin	Fresh Assam tea leaf	Flower (*Hibiscus rosa-sinensis* L.)	Flower (*Durio zibethinus* L.)	Frozen peas	Water in the drain pit of a kitchen sink	Fresh cummingcordia	Oral cavity	Tracheal (*Equus ferus caballus*)	Subcutaneous abscess in the udder of a sheep
Sequencing Technology	PacBio RSII	Illumina MiSeq	Illumina MiSeq	Illumina HiSeq	ABI 3730xl	Illumina NovaSeq	Sanger dideoxy sequencing	Illumina	Illumina NovaSeq
Assembly level	Complete	Draft	Draft	Draft	Complete	Draft	Complete	Draft	Draft
Genbank accession number	CP075561	NZ_MKIQ00000000	NZ_MKIR00000000	NZ_JXJX00000000	NC_020450	NZ_BNDT00000000	NZ_CP009913	NZ_RCVM00000000	NZ_JAAXPR000000000

Table 1. Genome features of *F. penangensis* ML061-4 compared with *Floricoccus* spp., *Lactococcus* spp. and *Streptococcus* spp. All data obtained from NCBI (https://www.ncbi.nlm.nih.gov/) except for strain ML061-4, [a]Average nucleotide identity (ANI) values obtained from EZBioCloud database (https://www.ezbiocloud.net/tools/ani). NA, not available. CDSs, coding sequences.

those of lactococci (35–38%) and streptococci (39–42%). To validate genetic relatedness among members of these three genera, the average nucleotide identity (ANI) value of strain ML061-4 was determined and compared to type strains representing members of the genera *Floricoccus*, *Lactococcus* and *Streptococcus*. Strain ML061-4 exhibits ANI values of 97.90, 93.70, 69.98 and 70.16% when compared with *F. penangensis* JCM 31735^T, *Floricoccus tropicus* JCM 31733^T, *Lactococcus plantarum* DSM 20686^T and *Streptococcus salivarius* NCTC 8618^T, respectively (Table 1). Congruently, the 16S rRNA gene sequence of strain ML061-4 displayed 100.0 and 99.8% identity to that of *F. penangensis* JCM 31735^T and *F. tropicus* JCM 31733^T, respectively, as previously described by Rungsirivanich et al.[2].

Genetic fingerprint and whole-genome phylogenetic analysis

DNA fingerprint $(GTG)_5$ profiles showed that *F. penangensis* ML061-4 is closely related to the genera *Streptococcus* and *Lactococcus*. The DNA fingerprint of strain ML061-4 included a band with a size of approximately 1,000 bp which it has in common with the fingerprint obtained for *S. thermophilus* Rico66 and *S. thermophilus* Brie28. The same fingerprinting experiment of *F. penangensis* ML061-4 also generated bands with sizes of ~3,500 and ~4,000 bp, which corresponds to results obtained for *L. lactis* Tempeh6L and *L. lactis* RMN5F, respectively. The obtained results also support the phylogenetic closeness between members of the genera *Streptococcus* and *Lactococcus* as illustrated by the generation of very similarly sized products of 2,100, 2,400 and 2,500 bp. In contrast, and as expected based on its distant phylogenomic relationship, the genetic fingerprinting profiles obtained *Bacillus* species were clearly distinct from each other and from those obtained for floricoccal, lactococcal and streptococcal members (Fig. 1).

A phylogenomic tree was constructed using complete genome sequences of *Lactococcus* spp. (8 genomes) and *Streptococcus* spp. (7 genomes) as well as that of *F. penangensis* ML061-4. The genome sequence of *Weissella koreensis* KACC 15510 was used as outgroup. The obtained phylogenomic tree showed that *F. penangensis* ML061-4 genome is clearly distinguished from genomes belonging to members of the genera *Lactococcus* and *Streptococcus*. Average nucleotide identity (ANI) between the genome of *F. penangensis* ML061-4 and those of selected species of *Lactococcus* and *Streptococcus* was determined. Strain ML061-4 was shown to elicit ANI values of 70.09, 69.43 and 69.90% when compared with *L. cremoris* FM-YL12, *L. protaetiae* KACC 19320 and *S. pneumoniae* PJ755/1, respectively. Furthermore, *F. penangensis* ML061-4 exhibits ANI value of 67.05% when compared with *W. koreensis* KACC 15510. The phylogenetic tree analysis of the whole genome is presented in Fig. 2.

Genome annotation of *F. penangensis* ML061-4

Functional assignment of the gene content of *F. penangensis* ML061-4 was possible for 1,195 genes (58.0%) involving 187 pathways as predicted by BlastKOALA; these annotations were shown to relate to genetic information processing (15.5%, 319 genes), environmental information processing (6.8%, 141 genes), signaling and cellular processes (6.6%, 136 genes), carbohydrate metabolism (6.5%, 134 genes), amino acid metabolism (4.0%, 82 genes), nucleotide metabolism (2.6%, 54 genes), metabolism of cofactors and vitamins (1.8%, 38 genes), lipid metabolism (1.6%, 33 genes), energy metabolism (1.4%, 28 genes), glycan biosynthesis and metabolism (1.3%, 26 genes), human diseases (0.1%, 3 genes) and unclassified (9.9%, 201 genes) (Fig. 3; Supplementary Table S1). These metabolic pathways may allow strain ML061-4 to survive under nutrient poor conditions[18].

A total of 1,235 genes (59.9%) harboured by the *F. penangensis* ML061-4 genome were classified into six KEGG functional categories (Fig. 4; Supplementary Table S2). KEGG functional annotation identified genes associated with metabolism (34.4%), genetic information processing (8.2%), environmental information processing (8.2%), cellular processes (3.3%), organismal systems (1.9%) and human diseases (3.9%). Among these, 199 genes (9.7%) were predicted to be associated with carbohydrate metabolism. Biosynthesis of substrates and products is necessary for body structure, microbial survival, replication and evolution[19]. Although gene annotation of *F. penangensis* ML061-4 revealed a number of genes functionally associated with human diseases, this does not mean that the strain is a pathogen.

Metabolic pathways associated with carbohydrate utilization in *F. penangensis* ML061-4 are listed in Supplementary Table S3. The ML061-4 genome is predicted to encode a fructose-bisphosphate aldolase (locus tag number KIW23_01985), a key enzyme of the Embden-Meyerhof (glycolysis) pathway, and an essential enzyme to generate pyruvate from glucose prior to conversion into L-lactic acid via L-lactate dehydrogenase (locus tag number KIW23_05220). No D-lactate dehydrogenase-encoding gene, which would be expected to be responsible for D-lactic acid production, was identified in the ML061-4 genome, in agreement with Rungsirivanich et al.[2]. , who reported that this strain produces a high proportion of L-lactic acid (94.6%).

In vitro carbohydrate fermentation abilities by *F. penangensis* ML061-4 have previously been reported, showing that this strain is able to ferment various monosaccharides (e.g., glucose, fructose and mannose), disaccharides (e.g., lactose, maltose, sucrose and trehalose), oligosaccharide (raffinose), polysaccharide (starch) and a sugar alcohol (mannitol)[1,2]. These results are consistent with KEGG functional prediction of the *F. penangensis* ML061-4 genome. The genome of ML061-4 was also shown to encompass genes required for various phosphoenolpyruvate: sugar phosphotransferase systems (PEP-PTSs), which are known to be involved in the transport and metabolism of carbon sources such as fructose, galactose, glucose, lactose and sucrose[20,21]. The *ptsP* and *ptsH* genes of *F. penangensis* ML061-4 encode the general PTS proteins EI and Hpr, respectively, while genes encoding putative substrate-specific permeases (EII) were also identified, for example *treP* (positions 1,580,662–1,582,716) responsible for the production of EIIBC^{trehalose}, and *manA* (positions 1,613,369–1,614,310) encoding a putative EIIABCD$^{[mannose}$[22].

Furthermore, our functional prediction efforts assigned 1,419 genes (68.8%) to a putative function by the COG database. We found that 16% of the genes identified on the ML061-4 genome was classified into translation, ribosomal structure, and biogenesis (J) category, 11.7% was associated with amino acid transport

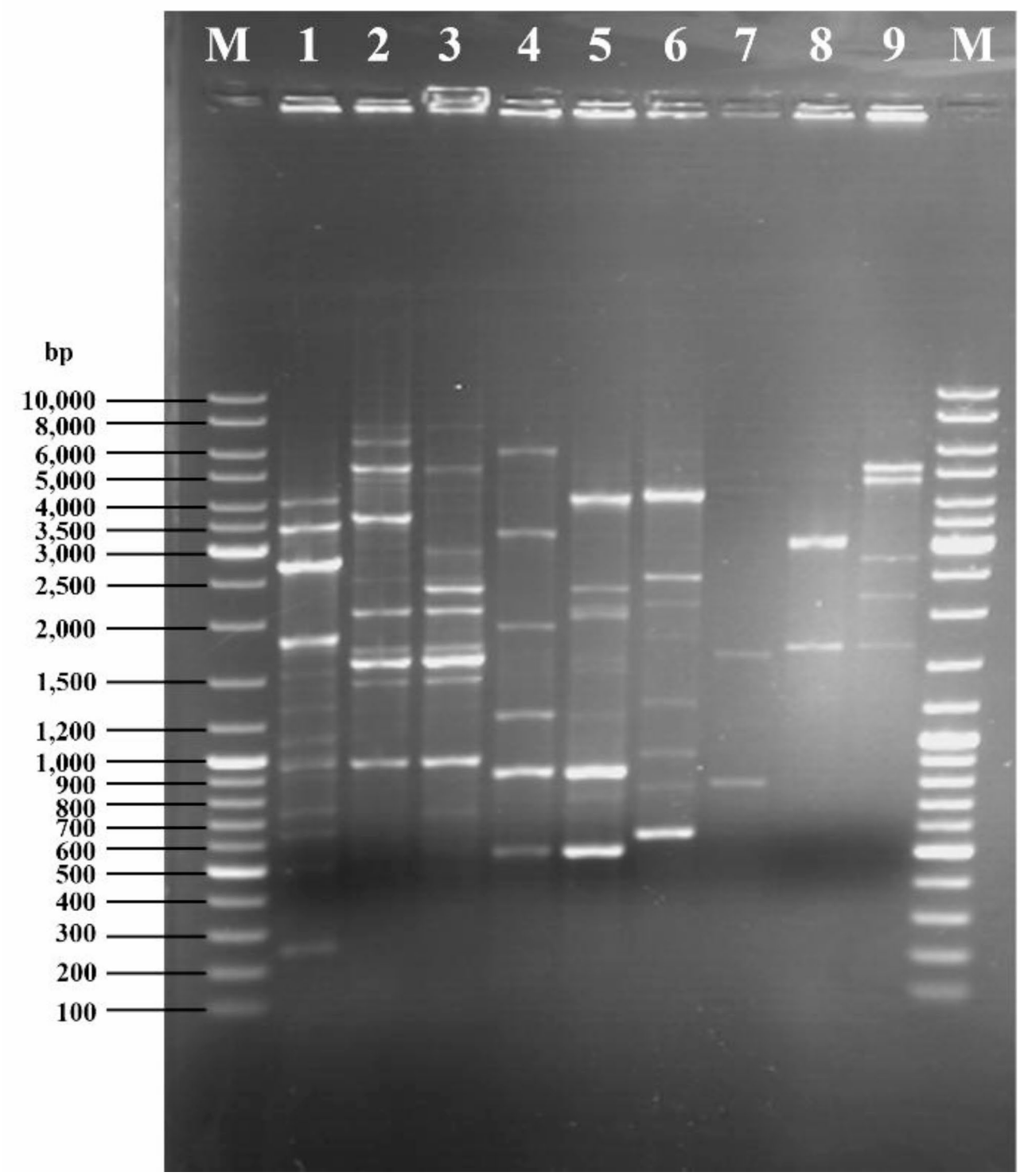


Fig. 1. (GTG)5-PCR DNA fingerprints of *Floricoccus penangensis* ML061-4 compared with *Streptococcus* spp., *Lactococcus* spp. and *Bacillus* spp. Lane M : GeneRuler DNA Ladder Mix, Lane 1 : *F. penangensis* ML061-4, Lane 2 : *S. thermophilus* Rico66, Lane 3 : *S. thermophilus* Brie28, Lane 4 : *L. lactis* Tempeh6L, Lane 5 : *L. lactis* RMN5F, Lane 6 : *B. velezensis* ML122-2, Lane 7 : *B. licheniformis* ML075-1, Lane 8 : *B. cereus* TISTR 687, Lane 9 : *B. subtilis* NCTC 10073.

and metabolism (category E), and 8.5% was involved in carbohydrate transport and metabolism (category G) (Fig. 5; Supplementary Table S4). Comparison of COG functional classification of genes encoded by *F. penangensis* ML061-4, *F. penangensis* JCM 31735^T, *F. tropicus* JCM 31733^T, *L. plantarum* DSM 20686^T, *L. lactis* subsp. *lactis* DSM 20481^T, *S. hillyeri* DSM 107591^T and *S. ovuberis* CCUG 69612^T revealed the close phylogenetic

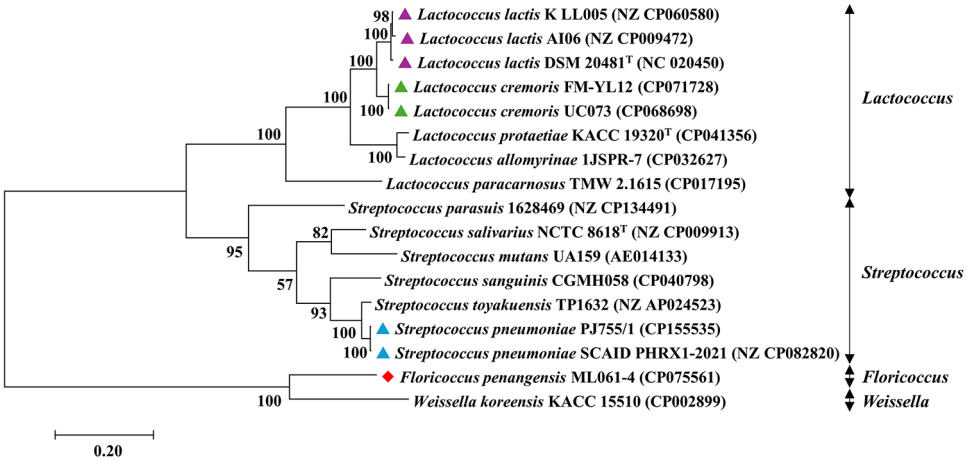


Fig. 2. Whole genome sequencing (WGS) phylogenetic tree of *Floricoccus penangensis* ML061-4 (red diamond) and the members of the genera *Lactococcus* and *Streptococcus*. The phylogenetic tree consisting of 17 strains was constructed using the Maximum Likelihood method. Bootstrap values were generated using 1,000 replications. The WGS of *Weissella koreensis* KACC 15510 was used as outgroup. Different symbols were used to highlight different species: a purple triangle corresponds to *Lactococcus lactis*, a green triangle indicates *Lactococcus cremoris* and a blue triangle reflects *Streptococcus pneumoniae*.

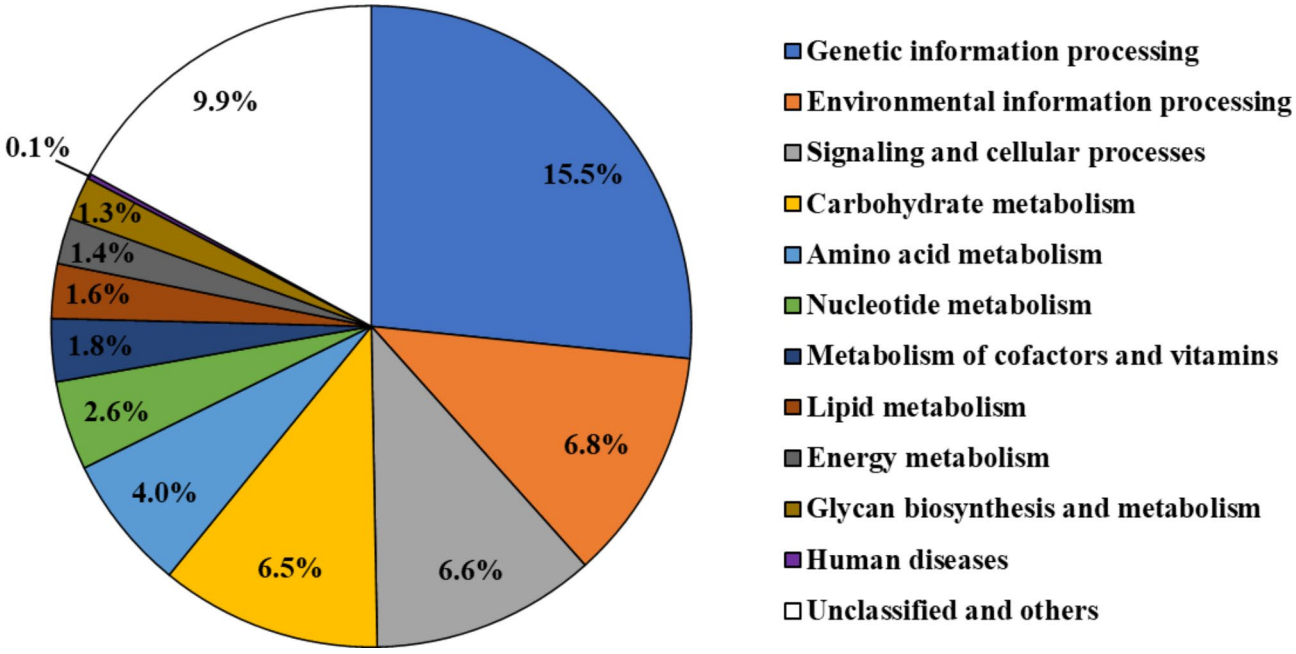


Fig. 3. Gene functional categories of *Floricoccus penangensis* ML061-4 genome predicted by BlastKOALA.

relationship between the genera *Floricoccus*, *Lactococcus* and *Streptococcus*. Meanwhile, *Bacillus subtilis* Bbv57, an outgroup, presented higher putative function genes except category J which related to floricoccal, lactococcal and streptococcal members, and disappeared with general function prediction (R) category. Between 80 and 170 genes of the genera *Floricoccus*, *Lactococcus* and *Streptococcus* were assigned as function unknown (category S), a much lower number compared to for example *Bacillus subtilis* strain Bbv57 which had 1,086 genes assigned to this category (Fig. 6; Supplementary Table S5 – S10). This phenomenon is likely due to differences in genome size (flori-, lacto- and streptococcus members possess genomes of approximately 2 Mbp in size (Table 1), whereas bacilli elicit genome sizes that are larger than 4 Mbp)[23,24].

Virulence factors, toxin genes and bacterial pathogenicity in *F. penangensis* ML061-4

To identify possible virulence factors and toxins that may be encoded by the ML061-4 genome, a similarity search was performed employing the virulence factors of pathogenic bacteria database (VFDB). Gene clusters with nucleotide sequence similarity values below 80% and E-value above $1e^{-10}$ were deemed non-significant[25].

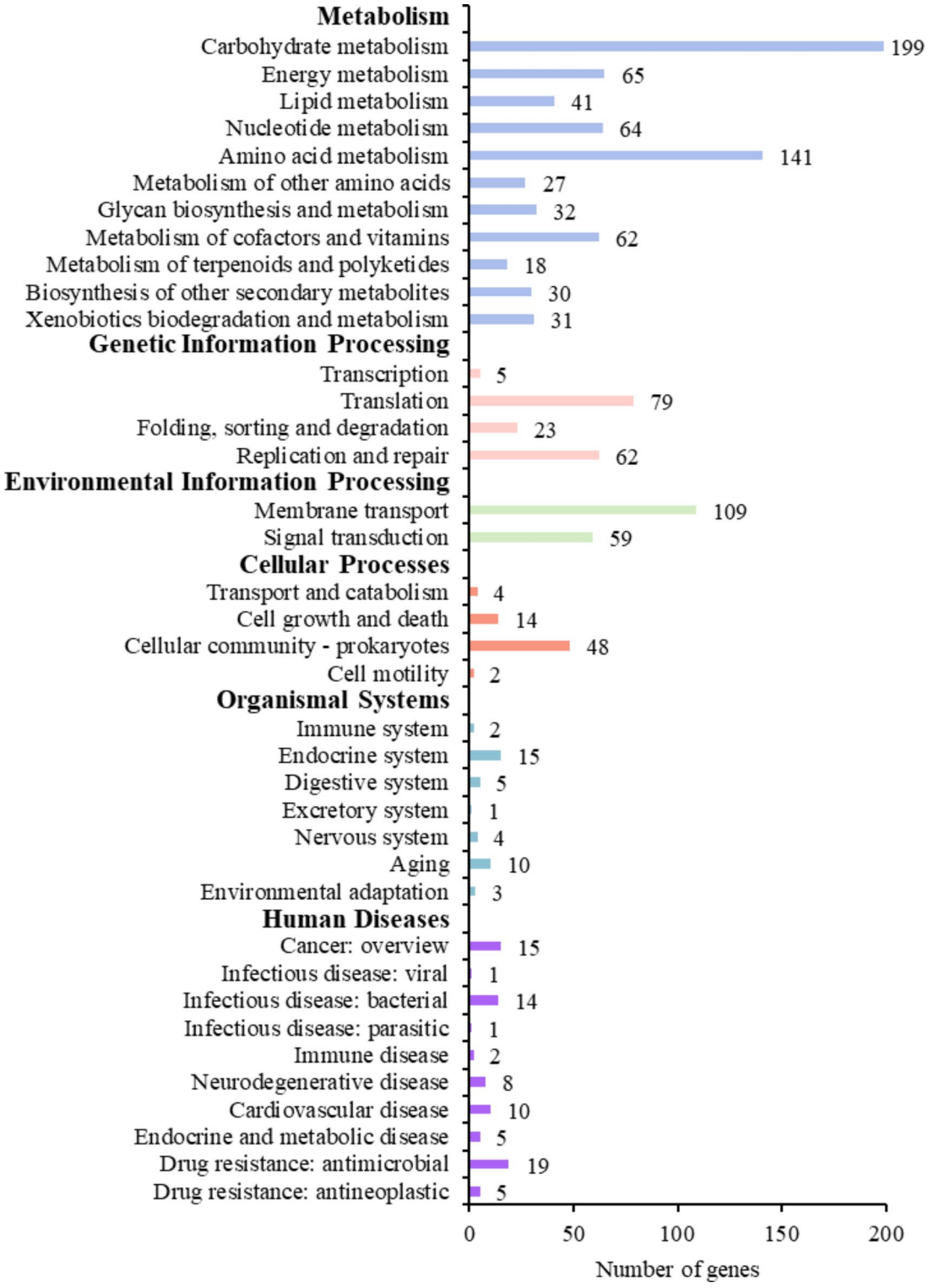


Fig. 4. KEGG functional annotation of genes encoded by *Floricoccus penangensis* ML061-4 which is separated in six groups including metabolism (blue), genetic information processing (pink), environmental information processing (green), cellular processes (orange), organismal systems (turquoise) and human diseases (purple).

Using these criteria no homologs corresponding to toxin production were found in the genome of ML061-4. The analysis revealed genes associated with virulence factors including *clpC*, *cpsI*, *lap*, *clpP*, *htpB*, *glf*, *tufA*, *groEL*, *hasC* and *clpE* with a nucleotide identity level of 79–90% (Table 2). These genes, many of which represent cellular housekeeping functions, are not necessarily linked directly to virulence such as *clpC* and *htpB* involved in ATP binding, while *cpsI* and *glf* encode UDP-galactopyranose mutase involved in cell wall biosynthesis[26].

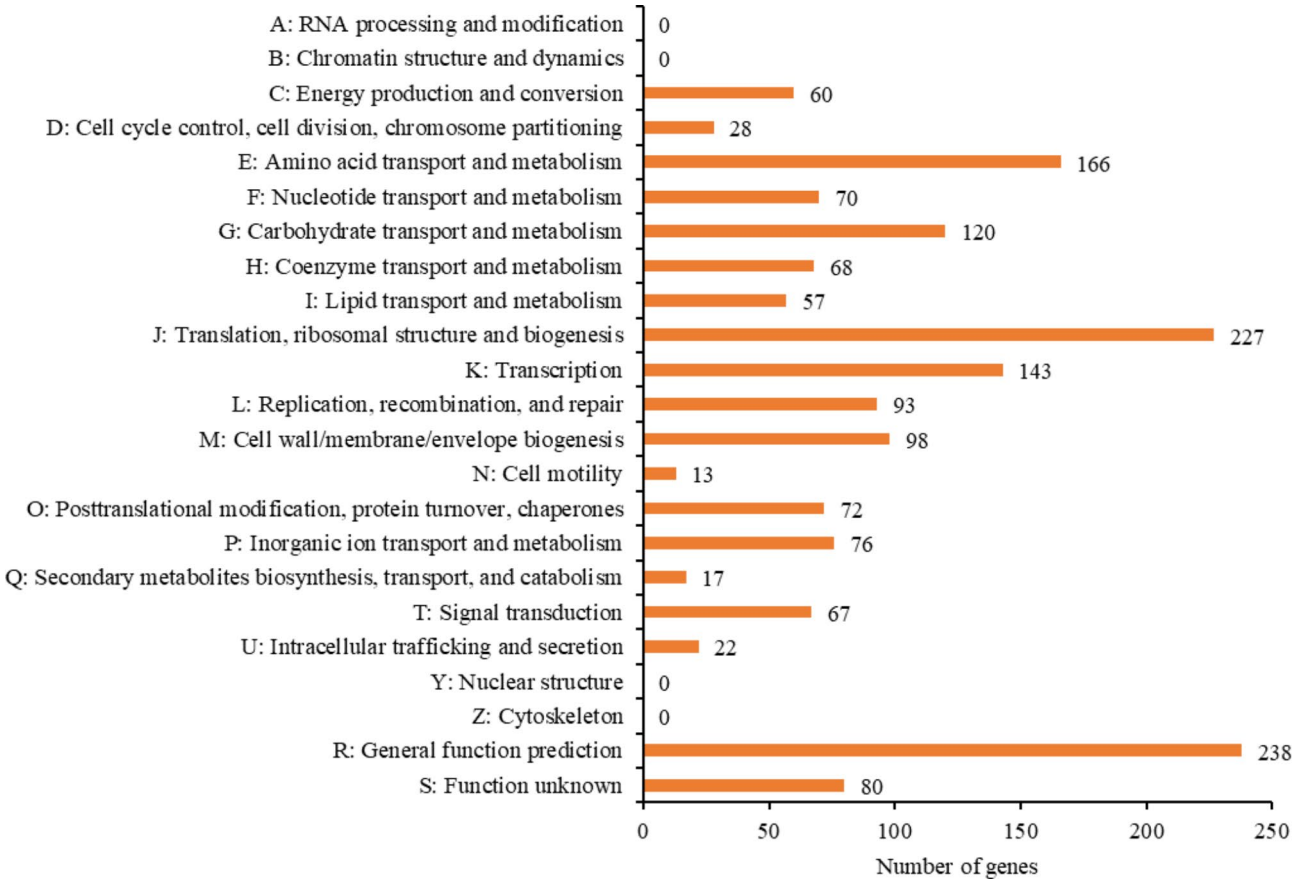


Fig. 5. COG functional annotation of genes encoded by *Floricoccus penangensis* ML061-4. Protein-coding sequences are classified according to COG functional categories. Alphabets A-Z exhibits COG functions. A: RNA processing and modification, B: Chromatin structure and dynamics, C: Energy production and conversion, D: Cell cycle control, cell division, chromosome, E: Amino acid transport and metabolism, F: Nucleotide transport and metabolism, I: Lipid transport and metabolism, J: Translation, ribosomal structure and biogenesis, K: Transcription, L: Replication, recombination, and repair, M: Cell wall/membrane/envelope biogenesis, N: Cell motility, O: Posttranslational modification, protein turnover, chaperones, P: Inorganic ion transport and metabolism, Q: Secondary metabolites biosynthesis, transport, and catabolism, R: General function prediction, S: Function Unknown, T: Signal transduction, U: Intracellular trafficking and secretion, Y: Nuclear structure, Z: Cytoskeleton.

Several genes were considered as conserved genes involved in stress responses (*clpC*, *clpP*, *clpE*, *tufA*, *htpB* and *groEL*). Furthermore, PathogenFinder software was applied in order to estimate the bacterial pathogenicity of *F. penangensis* ML061-4. Two genes were predicted to belong to pathogenic protein families including 30S ribosomal protein S19 (*rpsS* gene) and 30S ribosomal protein S21 (*rpsU* gene) with 100.0 and 93.1% identity (Supplementary Table S11). *rpsS* and *rpsU* genes have been described to be associated with rRNA binding, structural constituent of ribosome and tRNA binding (Uniport ID: P0A7U3 and P68679)[27].

Molecular mechanisms associated with adhesion

Four protein-encoding sequences in the ML061-4 genome were predicted to be involved in adhesion based on the VFDB database, including homologs of the *Listeria* adhesion protein (LAP; *lap* gene), 60 K heat shock protein (Hsp60; *htpB* gene), elongation factor Tu (*tufA* gene) and chaperonin (*groEL* gene) with 79–90% nucleotide identity (Table 2). It should be noted that these genes are widely considered to represent housekeeping genes and their physiological role in adhesion has been controversial[28–30]. Previous studies by Burkholder and Bhunia[31] and Drolia et al.[32]. revealed the roles of LAP and Hsp60 which have been associated with bacterial adhesion ability of *Escherichia coli* and *Listeria monocytogenes* to Caco-2 cells[33], while chaperonin GroEL and elongation factor Tu were described to participate in mucus adhesion[16,34]. Obtained data corresponding to in vitro experiments by Rungsirivanich et al.[2]. showed that *F. penangensis* ML061-4 possesses adhesion ability to an African green monkey kidney cell line (Vero cell) similar to *Lactobacillus acidophilus* and *Lactiplantibacillus plantarum* strains. The adhesion of microbes to host mucus and epithelial cells is the first step of human colonization and infection[35,36]. Bacterial adhesion ability is considered to be an essential trait to colonize and possible invasion of a host in case of pathogenic bacteria. Conversely, candidate probiotic strains with adhesion features may not only act as competitive exclusion to pathogens, yet may also positively modulate the immune response[37].

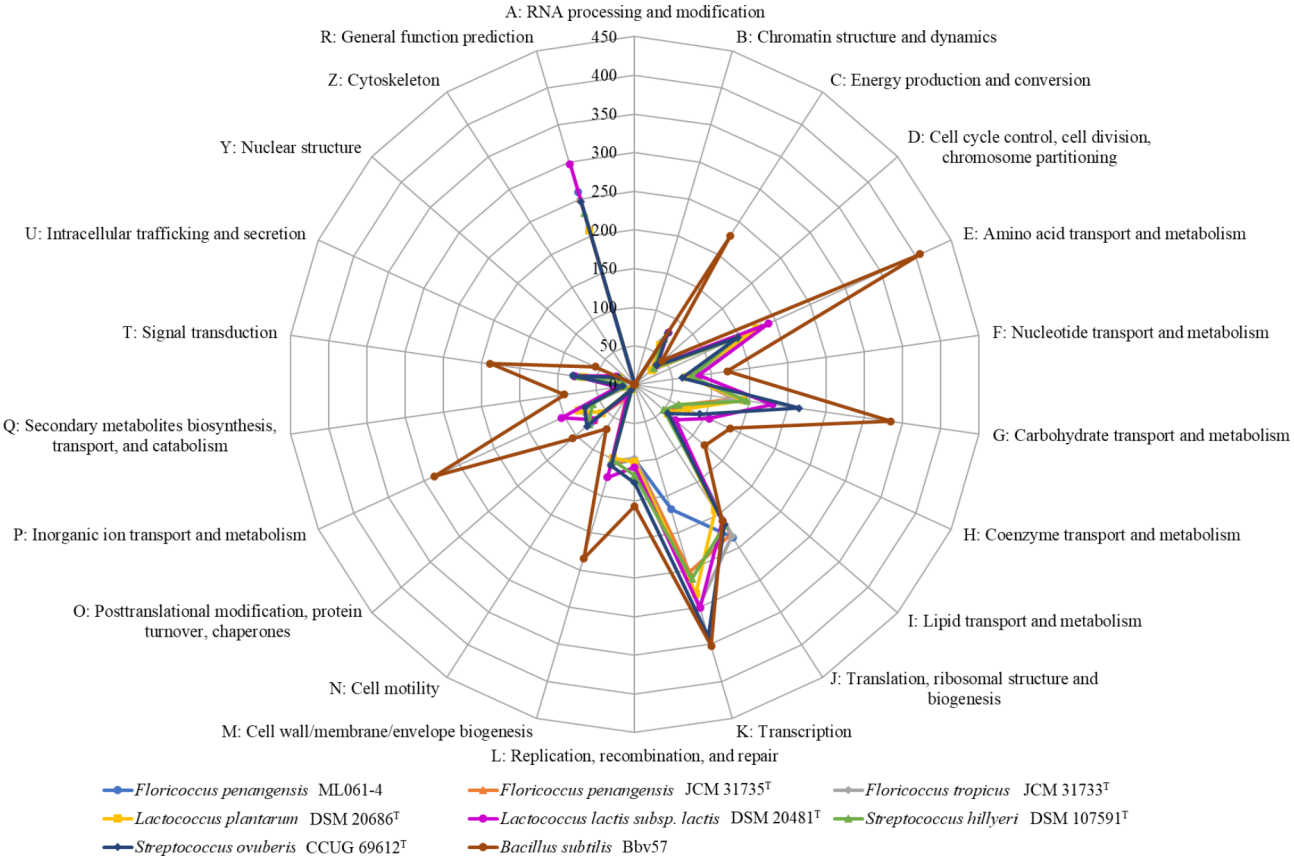


Fig. 6. COG functional classification of genes encoded by *Floricoccus penangensis* ML061-4 compared to *F. penangensis* JCM 31735^T, *F. tropicus* JCM 31733^T, *L. plantarum* DSM 20686^T, *L. lactis* subsp. *lactis* DSM 20481^T, *S. hillyeri* DSM 107591^T and *S. ovuberis* CCUG 69612^T. COG functional classification of genes encoded by *Bacillus subtilis* Bbv57, as previous described by Thiruvengadam et al.[24]. , is used as an outgroup. Between 80 and 170 genes of the genera *Floricoccus*, *Lactococcus* and *Streptococcus* were assigned as function unknown (category S), whereas *Bacillus subtilis* had 1,086 genes assigned to category S.

VFDB Gene ID	Gene name	Product	Function	Genome location	Nucleotide identity (%)	E-value	NCBI reference sequence number
VFG000079	*clpC*	Endopeptidase Clp ATP-binding chain C	Stress survival	301,586 – 301,743	84	8e-25	NP_463763
VFG002182	*cpsI*	UDP-galactopyranose mutase	Immune modulation	183,518 – 183,732	80	3e-18	WP_002376666
VFG006717	*lap*	*Listeria* adhesion protein Lap	Adherence	1,998,874- 1,998,938	90	3e-12	NP_465159
VFG000077	*clpP*	ATP-dependent Clp protease proteolytic subunit	Stress survival	290,069–290,161	86	3e-12	NP_465991
VFG001855	*htpB*	Hsp60, 60 K heat shock protein HtpB	Adherence	1,720,195- 1,720,355	81	3e-12	WP_197535493
VFG001967	*glf*	UDP-galactopyranose mutase	Immune modulation	183,518 – 183,603	86	2e-10	YP_002344822
VFG046465	*tufA*	Elongation factor Tu	Adherence	640,452–640,588	81	7e-10	WP_003028672
VFG012095	*groEL*	Chaperonin GroEL	Adherence	1,721,173- 1,721,373	79	7e-10	WP_003435012
VFG000964	*hasC*	UTP-glucose-1-phosphate uridylyltransferase HasC	Immune modulation	1,810,158- 1,810,261	82	6e-07	WP_010922799
VFG000080	*clpE*	ATP-dependent protease	Stress survival	1,801,366- 1,801,457	83	6e-07	NP_464522

Table 2. Prediction of genes related to virulence factor in *F. penangensis* ML061-4 using VFDB database.

Bacteriophages, antibiotic and antiviral defenses

To predict prophage elements carried by the *F. penangensis* ML061-4 genome, in silico analysis using PHASTEST software was performed. One region with homology to *Streptococcus* phage PH15 were identified in the genome of *F. penangensis* ML061-4. This 16.1 Kbp DNA region with a 32.85% GC content was predicted to represent an incomplete prophage (Table 3; Supplementary Fig. S1 and S2). Of note, a prophage-encompassing region homologous to the complete *Enterococcus* phage phiFL1A was predicted to be present in the *F. penangensis* JCM

31735^T genome, while two prophage regions related to *Lactococcus* phage 50101 and *Bacillus* phage BCJA1c were predicted to be harbored by the genome of *F. tropicus* JCM 31733^T. Furthermore, various predicted (partial) prophage-encompassing regions were predicted in the *L. lactis* DSM 20481^T, *S. salivarius* NCTC 8618^T and *S. hillyeri* DSM 107591^T genomes, while no prophages or remnants thereof appeared to be present in the genomes of *L. plantarum* DSM 20686^T and *S. ovuberis* CCUG 69612^T (Table 3). None of these predicted prophage regions appeared to be intact; however, the relatedness of these regions to phages of diverse bacterial species provides insights (especially regarding phage origins and evolution) into the co-existence and interactions with other organisms in its environmental niche. Phages can infect bacteria and integrate their genomes into the bacterial genome in order to allow phage multiplication. Such integrated phage genomes or prophages may in turn be activated, resulting in phage particle production and host death[38]. Prophage regions are commonly present in bacterial genomes and have been reported to have many beneficial distinguishing qualities for bacteria such as bacterial adaption in a new environment as well as promoting bacterial resistance[26,39].

To identify genes involved in antibiotic resistance, the complete genome of ML061-4 was determined using CARD database based on RGI criteria with perfect and strict hits. The results showed that the ML061-4 genome did not contain any obvious antibiotic resistance genes. Nevertheless, the KEGG ontology of ML061-4 revealed genes involved in drug resistance pathways including beta-lactam, vancomycin and cationic antimicrobial peptide. A previous study by Rungsirivanich et al.[2]. has shown that *F. penangensis* ML061-4 is susceptible to various tested antibiotics such as ampicillin, cefotaxime, gentamicin, meropenem and tetracycline. The absence of antibiotic resistance genes is suggested to be a desirable property as it eliminates the risk of horizontal transfer to other microorganisms leading to antibiotic drug resistance[40,41]. The genome of *F. penangensis* ML061-4 was shown to lack any CRISPR-Cas systems. CRISPR-Cas systems are found in approximately 40% of assessed bacteria and are known to provide immunity against exogenous nucleic acids, especially bacteriophages[42]. Antiphage systems of strain ML061-4 were predicted by the PADLOC analysis. The results revealed that the genome of *F. penangensis* ML061-4 contains 5 antiphage systems (DMS_other, RM_type_IV, RM_type_IIG, retron_III-A and PD-T4-6) and 2 phage defense candidate systems (PDC-S30 and PDC-M01). Antiphage systems are multigene (located in operons) or single genes. DNA modification systems (DMS) and restriction-modification systems (RM), retrons have been found to play an important role in bacterial antiviral defense[43,44]. The predicted antiviral defence systems of *F. penangensis* ML061-4 are listed in Table 4, Supplementary Fig. S3 and Supplementary Table S12.

Secondary metabolite biosynthesis and xenobiotics biodegradation

Secondary metabolic potential of strain ML061-4 was examined using antiSMASH, yet did not produce any credible candidates. The antiSMASH software is widely used to predict microbial secondary metabolites for biosynthetic gene clusters particularly antimicrobial, non-ribosomally synthesized peptides (NRPs), polyketides (PKs) and bioactive compounds[45]. Conversely, KEGG ontology of strain ML061-4 revealed 48 genes involved in 20 metabolic pathways for terpenoids and polyketides, and secondary metabolite biosynthesis pathways (e.g., antibiotics, insect hormone and plant secondary metabolites) and 31 genes associated with 11 pathways of xenobiotic biodegradation (e.g., organic compounds and drugs) (Supplementary Table S13). Xenobiotics are chemical substances that are not naturally produced or found in organisms or the environment[46]. In general, xenobiotic refers to environmental pollutants resulting in industry and agriculture such as pesticides, polycyclic aromatic hydrocarbons, pharmaceutical active compounds and phenolic compounds[47]. Xenobiotic biodegradation ability may assist a microorganism to survive in particular polluted environments[48]. KEGG annotation revealed gene clusters involved in xenobiotic biodegradation pathway in ML061-4 genome comprising benzoate, aminobenzoate, chloroalkane, chloroalkene, chlorocyclohexane, chlorobenzene, xylene,

Strain	Number of phage found	Location	Length (Kbp)	Completeness	GC (%)	Number of phage gene found	Predicted phage	NCBI reference sequence number
F. penangensis ML061-4	1	86,797–102,892	16.1	Incomplete	32.85	14	*Streptococcus* phage PH15	NC_010945
F. penangensis JCM 31735^T	1	936,118–1,318,224	48.1	Intact	33.11	24	*Enterococcus* phage phiFL1A	NC_013646
F. tropicus JCM 31733^T	2	703,783–1,097,746	35.7	Intact	33.36	40	*Lactococcus* phage 50101	NC_031040
		1,423,859–1,539,006	44.3	Intact	32.69	40	*Bacillus* phage BCJA1c	NC_006557
L. plantarum DSM 20686^T	Not detected	-	-	-	-	-	-	-
L. lactis DSM 20481^T	2	1,706,646–1,765,449	58.8	Intact	35.40	53	*Lactococcus* phage PLgT-1	NC_031016
		1,951,310–1,976,355	25.0	Incomplete	33.98	14	*Lactococcus* phage bIL312	NC_002671
L. taiwanensis NBRC 109049^T	Not detected	-	-	-	-	-	-	-
S. salivarius NCTC 8618^T	1	666,756–715,649	48.9	Intact	42.02	52	*Streptococcus* phage 5093	NC_012753
S. hillyeri DSM 107591^T	2	1,236,759–1,307,534	38.8	Intact	40.28	47	*Streptococcus* phage 20617	NC_023503
		1,648,481–1,689,670	37.9	Intact	39.78	44	*Streptococcus* phage PH10	NC_012756
S. ovuberis CCUG 69612^T	Not detected	-	-	-	-	-	-	-

Table 3. Gene cluster identification of prophages in *F. penangensis* ML061-4 genome compared with *Floricoccus* spp., *Lactococcus* spp. and *Streptococcus* spp. using PHASTEST.

System	Protein name	Genome location	Strand	Relative position
DMS_other	Control_protein	108,671–108,886	Forward	99
DMS_other	DrmMII	108,971–110,209	Forward	100
RM_type_IV	mREase_IV	403,002–405,881	Forward	370
RM_type_IIG	REase_MTase_IIG	713,307–716,267	Forward	646
DMS_other	MTase_II	716,260–717,891	Forward	647
retron_III-A	PRTase_III-A	909,798–911,384	Forward	848
retron_III-A	RT_III-A	911,645–912,646	Forward	850
PDC-S30	PDC-S30	1,106,895–1,107,746	Forward	1042
PD-T4-6	PD-T4-6	1,583,628–1,585,646	Reverse	1507
PDC-M01	PDC-M01B	1,793,808–1,795,058	Reverse	1718
PDC-M01	PDC-M01A	1,795,003–1,797,120	Reverse	1719

Table 4. Antiviral defence systems detected in *F. penangensis* ML061-4 genome using PADLOC analysis.

styrene, dioxin and naphthalene (Supplementary Table 13). Therefore, secondary metabolites and substances related to xenobiotic biodegradation produced by *F. penangensis* ML061-4 should be further elucidated.

Materials and methods

Bacterial strain and growth conditions

F. penangensis ML061-4 (GenBank accession no. MH050697) was isolated from the leaf surface of an Assam tea plant (*Camellia sinensis* var. *assamica*) in the Sakat sub-district, Pua district, Nan province, Thailand at an altitude of 1,038 m above sea level (19°15'53.62"N, 101°0'30.22"E)[2]. The *F. penangensis* ML061-4 was cultivated in M17 (Oxoid™, Basingstoke, England) supplemented with 0.5% (w/v) glucose (GM17) broth and incubated at 30 °C for 24 h, as previously reported by Rungsirivanich et al.[23].

Genome sequencing and assembly

Whole genome sequencing and assembly of strain Ml061-4 was accomplished as previously described by Rungsirivanich et al.[14]. The genome sequence was deposited in GenBank under accession number CP075561 and Sequence Read Archive (SRA) number SRP399917.

PCR fingerprinting using the $(GTG)_5$ primer

$(GTG)_5$-PCR DNA fingerprint analysis of *F. penangensis* ML061-4 was performed according to the protocol of Parlindungan et al.[49]. *Streptococcus thermophilus* Rico66, *S. thermophilus* Brie28, *Lactococcus lactis* Tempeh6L and *L. lactis* RMN5F as previously described by Parlindungan et al.[50] , *Bacillus velezensis* ML122-2, *B. licheniformis* ML075-1, *B. cereus* TISTR 687 and *B. subtilis* NCTC 10073 as previously described by Rungsirivanich et al.[23]. were used for comparison purposes. PCR amplifications were performed using *Taq* DNA polymerase mastermix (Qiagen, Manchester, UK) with the single oligonucleotide primer $(GTG)_5$, 5′-GTGGTGGTGGTGGTG-3′. The PCR reaction mixtures were initially denatured at 95 °C for 7 min, followed by 30 cycles of denaturation at 90 °C for 30 s, annealing at 40 °C for 1 min, extension at 65 °C for 8 min and a final extension step at 65 °C for 16 min using an Applied Biosystems™ 2720 Thermal Cycler (Thermo Fisher, CA, USA). Amplicons were subjected to electrophoresis on a 1% (w/v) agarose gel consisting of ethidium bromide for 1 h at 110 V 100 mA in 1X Tris-acetate-EDTA (TAE) buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8.0). GeneRuler DNA Ladder Mix (Thermo Fisher, CA, USA) was used as a marker. The GTG PCR profiles were visualized under ultraviolet light by gel documentation system, and gel images were captured using GeneSnap software (Syngene, MD, USA).

Phylogenomic tree analysis

A phylogenomic tree based on whole genome sequences of strain ML061-4 and various members of the genera *Lactococcus* and *Streptococcus* was constructed according to the method of L'Huillier et al.[51]. A total of 16 complete genome sequences were retrieved from the National Centre for Biotechnology Information (NCBI) including *Lactococcus lactis* K_LL005 (GenBank accession no. NZ_CP060580), *Lactococcus lactis* AI06 (GenBank accession no. NZ_CP009472), *Lactococcus lactis* DSM 20481^T (GenBank accession no. NC_020450), *Lactococcus cremoris* FM-YL12 (GenBank accession no. CP071728), *Lactococcus cremoris* UC073 (GenBank accession no. CP068698), *Lactococcus protaetiae* KACC 19320^T (GenBank accession no. CP041356), *Lactococcus allomyrinae* 1JSPR-7 (GenBank accession no. CP032627), *Lactococcus paracarnosus* TMW 2.1615 (GenBank accession no. CP017195), *Streptococcus parasuis* 1628469 (GenBank accession no. NZ_CP134491), *Streptococcus salivarius* NCTC 8618^T (GenBank accession no. NZ_CP009913), *Streptococcus mutans* UA159 (GenBank accession no. AE014133), *Streptococcus sanguinis* CGMH058 (GenBank accession no. CP040798), *Streptococcus toyakuensis* TP1632 (GenBank accession no. NZ_AP024523), *Streptococcus pneumoniae* PJ755/1 (GenBank accession no. CP155535), *Streptococcus pneumoniae* SCAID PHRX1-2021 (GenBank accession no. NZ_CP082820) and *Weissella koreensis* KACC 15510 (GenBank accession no. CP002899). The kSNP 4.0 was applied to perform alignment-free whole genome phylogenetic analysis by identifying single nucleotide polymorphisms (SNPs) across the genome sequences[52]. The obtained SNP-matrix results were imported into MEGA 11 for phylogenetic/phylogenomic tree construction[53]. The phylogenomic tree was analyzed by a Maximum Likelihood method[54] with bootstrap values of 1000 replicates[55].

Genome annotation

Automatic annotation of predicted open reading frames (ORFs) was performed using a combination of Prodigal (version 2.6.3) and BLASTp (version 2.2.26) sequence alignments[56] to assign annotation, using an E-value cut-off of 0.0001 for hits exhibiting at least 50% similarity across at least 50% of the sequence length) against a non-redundant protein database provided by the NCBI portal. Functional prediction of genes and proteins was integrated using the Clusters of Orthologous Groups (COGs)[57] and protein family (Pfam)[58], respectively, as previously described by Martín et al.[59]. Ribosomal RNA (rRNA) and transfer RNA (tRNA) genes were detected using RNAmmer version 1.2[60] and tRNA-scanSE version 2.0[61], respectively.

The Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology annotation was undertaken[62] by the KAAS online server version 2.1[63] using the bi-directional best hit (BBH) method (https://www.genome.jp/kegg/kaas/). Genome functional characterization was further assigned by BlastKEGG orthology and links annotation (BlastKOALA; https://www.kegg.jp/blastkoala/)[64]. The functional analysis of singleton genes was classified into categories of the COG using the Batch Web CD-Search Tool (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi)[65]. The virulence factor search in the ML061-4 genome was conducted using the virulence factors of pathogenic bacteria database (VFDB) (http://www.mgc.ac.cn/VFs/)[66], and bacterial pathogenicity elements were predicted using PathogenFinder version 1.1 (https://cge.cbs.dtu.dk/services/PathogenFinder/)[67]. Moreover, the presence of lysogenic bacteriophages within the genome was predicted using PHASTEST (https://phastest.ca/)[68], and putative virulence factors and antibiotic resistance genes were identified by CARD database (https://card.mcmaster.ca/)[69]. Genes involved in secondary metabolite biosynthesis were predicted using antiSMASH (https://antismash-db.secondarymetabolites.org/)[45]. The presence of *c*lustered *r*egularly *i*nterspaced *s*hort *p*alindromic *r*epeats (CRISPR) systems was investigated using CRISPRfinder (https://crispr.i2bc.paris-saclay.fr/Server/)[70]. Antiviral defence systems were identified using PADLOC analysis (https://padloc.otago.ac.nz)[71].

Genbank accession numbers and comparative genome analysis

Genome information of *Floricoccus penangensis* JCM 31735^T (GenBank accession no. NZ_MKIQ00000000), *Floricoccus tropicus* JCM 31733^T (GenBank accession no. NZ_MKIR00000000), *Lactococcus plantarum* DSM 20686^T (GenBank accession no. NZ_JXJX00000000), *Lactococcus lactis* DSM 20481^T (GenBank accession no. NC_020450), *Lactococcus taiwanensis* NBRC 109049^T (GenBank accession no. NZ_BNDT00000000), *Streptococcus salivarius* NCTC 8618^T (GenBank accession no. NZ_CP009913), *Streptococcus hillyeri* DSM 107591^T (GenBank accession no. NZ_RCVM00000000) and *Streptococcus ovuberis* CCUG 69612^T (GenBank accession no. NZ_JAAXPR000000000) was retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/) for comparative genomic analysis. The genome identity assessments were performed by average nucleotide identity (ANI) determination using EzBioCloud databases (https://www.ezbiocloud.net/tools/ani)[72].

Conclusions

This study reports on a functional genomic analysis of a member of the genus *Floricoccus*, and in particular compares its genome to those belonging to closely related members of the genera *Lactococcus* and *Streptococcus*. Comparative genomic analysis showed that the G + C content of the genus *Floricoccus* is explicitly distinguished from *Lactococcus* and *Streptococcus* genera. In silico analysis exhibited that *F. penangensis* ML061-4 encodes enzymes for the metabolism of various carbohydrates. Genome sequence analysis exposed a high level of genetic variation among members of the *Floricoccus*, *Lactococcus* and *Streptococcus* genera. Strain ML061-4 contains genes associated with both probiotic and pathogenic properties, and future research efforts will assist in elucidating the biological functionality, safety and potential biotechnological applications of strains of this species.

Data availability

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ncbi.nlm.nih.gov/genbank/, CP075561, NZ_MKIQ00000000, NZ_MKIR00000000, NZ_JXJX00000000, NC_020450, NZ_BNDT00000000, NZ_CP009913, NZ_RCVM00000000, NZ_JAAXPR000000000, NZ_CP060580, NZ_CP009472, NC_020450, CP071728, CP068698, CP041356, CP032627, CP017195, NZ_CP134491, NZ_CP009913, AE014133, CP040798, NZ_AP024523, CP155535, NZ_CP082820, CP002899.

Received: 30 June 2024; Accepted: 13 January 2025
Published online: 23 January 2025

References

1. Chuah, L. O. et al. *Floricoccus tropicus* gen. nov., sp. nov. and *Floricoccus penangensis* sp. nov. isolated from fresh flowers of durian tree and hibiscus. *Int. J. Syst. Evol. Microbiol.* **67**, 4979–4985. https://doi.org/10.1099/ijsem.0.002386 (2017).
2. Rungsirivanich, P., Inta, A., Tragoolpua, Y. & Thongwai, N. Partial *rpoB* gene sequencing identification and probiotic potential of *Floricoccus penangensis* ML061-4 isolated from Assam tea (*Camellia sinensis* var. *assamica*). *Sci. Rep.* **9**, 16561. https://doi.org/10.1038/s41598-019-52979-9 (2019).
3. Steiner, K., Laroche, O., Walker, S. P. & Symonds, J. E. Effects of water temperature on the gut microbiome and physiology of Chinook salmon (*Oncorhynchus tshawytscha*) reared in a freshwater recirculating system. *Aquaculture* **560**, 738529. https://doi.org/10.1016/j.aquaculture.2022.738529 (2022).
4. Vasiljevs, S., Witney, A. A. & Baines, D. L. The presence of cystic fibrosis-related diabetes modifies the sputum microbiome in cystic fibrosis disease. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **326**, L125–L134. https://doi.org/10.1152/ajplung.00219.2023 (2024).
5. Song, A. A., In, L. L. A., Lim, S. H. E. & Rahim, R. A. A review on *Lactococcus lactis*: From food to factory. *Microb. Cell. Fact.* **16**, 55. https://doi.org/10.1186/s12934-017-0669-x (2017).

6. Li, W. et al. Fermentation characteristics of *Lactococcus lactis* subsp. *lactis* isolated from naturally fermented dairy products and screening of potential starter isolates. *Front. Microbiol.* 11, 1794, (2020). https://doi.org/10.3389/fmicb.2020.01794
7. Jung, M. Y., Lee, C., Seo, M. J., Roh, S. W. & Lee, S. H. Characterization of a potential probiotic bacterium *Lactococcus raffinolactis* WiKim0068 isolated from fermented vegetable using genomic and *in vitro* analyses. *BMC Microbiol.* **20**, 136. https://doi.org/10.1186/s12866-020-01820-9 (2020).
8. Martinović, A., Cocuzzi, R., Arioli, S. & Mora, D. *Streptococcus thermophilus*: To survive, or not to survive the gastrointestinal tract, that is the question! *Nutrients* **12**, 2175. https://doi.org/10.3390/nu12082175 (2020).
9. Wei, W., Qiao, Z., Qin, D. & Lan, Y. Acute multiple brain infarctions associated with *Streptococcus suis* infection: A case report. *BMC Infect. Dis.* **24**, 447. https://doi.org/10.1186/s12879-024-09318-9 (2024).
10. Lahlou, W. et al. *Lactococcus lactis* endocarditis and liver abscess in an immunocompetent patient: A case report and review of the literature. *J. Med. Case Rep.* **17**, 115. https://doi.org/10.1186/s13256-022-03676-1 (2023).
11. Rungsirivanich, P. et al. Culturable bacterial community on leaves of Assam tea (*Camellia sinensis* var. *assamica*) in Thailand and human probiotic potential of isolated *Bacillus* spp. *Microorganisms* **8**, 1585. https://doi.org/10.3390/microorganisms8101585 (2020).
12. Yang, L. A., Chang, Y. J., Chen, S. H., Lin, C. Y. & Ho, J. M. SQUAT: a sequencing quality assessment tool for data quality assessments of genome assemblies. *BMC Genom.* **19**, 238. https://doi.org/10.1186/s12864-019-5445-3 (2019).
13. Zhang, P. et al. Comparison of de novo assembly strategies for bacterial genomes. *Int. J. Mol. Sci.* **22**, 7668. https://doi.org/10.3390/ijms22147668 (2021).
14. Rungsirivanich, P. et al. Complete genome sequence of *Floricoccus penangensis* ML061-4 isolated from Assam tea leaf [*Camellia sinensis* var. *assamica* (J.W.Mast.) Kitam]. *J. Genomics*. **11**, 37–39. https://doi.org/10.7150/jgen.83521 (2023).
15. Rhoads, A. & Au, K. F. PacBio sequencing and its applications. *Genom Proteom. Bioinform*. **13**, 278–289. https://doi.org/10.1016/j.gpb.2015.08.002 (2015).
16. Abriouel, H. et al. Insight into potential probiotic markers predicted in *Lactobacillus pentosus* MP-10 genome sequence. *Front. Microbiol.* **8**, 891. https://doi.org/10.3389/fmicb.2017.00891 (2017).
17. Zheng, J. et al. *In vitro* probiotic characteristics and whole genome sequence analysis of *Lactobacillus* strains isolated from cattle-yak milk. *Biology* **11**(44). https://doi.org/10.3390/biology11010044 (2022).
18. Zhang, W. et al. Complete genome sequencing and comparative genome characterization of *Lactobacillus johnsonii* ZLJ010, a potential probiotic with health-promoting properties. *Front. Genet.* **10**, 812. https://doi.org/10.3389/fgene.2019.00812 (2019).
19. Khatiebi, S. et al. Shotgun metagenomic analyses of microbial assemblages in the aquatic ecosystem of Winam Gulf of Lake Victoria, Kenya reveals multiclass pollution. *Biomed. Res. Int.* **2023** (3724531). https://doi.org/10.1155/2023/3724531 (2023).
20. Saier, M. H. J. The bacterial phosphotransferase system: new frontiers 50 years after its discovery. *J. Mol. Microbiol. Biotechnol.* **25**, 73–78. https://doi.org/10.1159/000381215 (2015).
21. Ou, Y. J. et al. Complete genome insights into *Lactococcus petauri* CF11 isolated from a healthy human gut using second- and third-generation sequencing. *Front. Genet.* **11**, 119. https://doi.org/10.3389/fgene.2020.00119 (2020).
22. Jamal, Z. et al. Distribution and functions of phosphotransferase system genes in the genome of the lactic acid bacterium *Oenococcus oeni*. *Appl. Environ. Microbiol.* **79**, 3371–3379. https://doi.org/10.1128/AEM.00380-13 (2013).
23. Rungsirivanich, P. et al. Simultaneous production of multiple antimicrobial compounds by *Bacillus velezensis* ML122-2 isolated from Assam tea leaf [*Camellia sinensis* var. *assamica* (J.W.Mast.) Kitam]. *Front. Microbiol.* **12**, 789362. https://doi.org/10.3389/fmicb.2021.789362 (2021).
24. Thiruvengadam, R. et al. Complete genome sequence analysis of *Bacillus subtilis* Bbv57, a promising biocontrol agent against phytopathogens. *Int. J. Mol. Sci.* **23**, 9732. https://doi.org/10.3390/ijms23179732 (2022).
25. Chokesajjawatee, N. et al. Safety Assessment of a *Nham* starter culture *Lactobacillus plantarum* BCC9546 via whole-genome analysis. *Sci. Rep.* **10**, 10241. https://doi.org/10.1038/s41598-020-66857-2 (2020).
26. Panthee, S., Paudel, A., Blom, J., Hamamoto, H. & Sekimizu, K. Complete genome sequence of *Weissella hellenica* 0916-4-2 and its comparative genomic analysis. *Front. Microbiol.* **10**, 1619. https://doi.org/10.3389/fmicb.2019.01619 (2019).
27. Ban, N. et al. A new system for naming ribosomal proteins. *Curr. Opin. Struct. Biol.* **24**, 165–169. https://doi.org/10.1016/j.sbi.2014.01.002 (2014).
28. Shi, J. et al. Characterization and function analysis of Hsp60 and Hsp10 under different acute stresses in black tiger shrimp, *Penaeus monodon*. *Cell. Stress Chaperones*. **21**, 295–312. https://doi.org/10.1007/s12192-015-0660-6 (2016).
29. Xie, J. et al. Identification and evaluation of suitable reference genes for RT-qPCR analysis in *Hippodamia variegata* (Coleoptera: Coccinellidae) under different biotic and abiotic conditions. *Front. Physiol.* **12**, 669510. https://doi.org/10.3389/fphys.2021.669510 (2021).
30. Tsuchihashi, H. et al. Genetic diversity of *Lactobacillus delbrueckii* isolated from raw milk in Hokkaido, Japan. *J. Dairy. Sci.* **105**, 2082–2093. https://doi.org/10.3168/jds.2021-21135 (2022).
31. Burkholder, K. M. & Bhunia, A. K. *Listeria monocytogenes* uses *Listeria* adhesion protein (LAP) to promote bacterial transepithelial translocation and induces expression of LAP receptor Hsp60. *Infect. Immun.* **78**, 5062–5073. https://doi.org/10.1128/IAI.00516-10 (2010).
32. Drolia, R., Tenguria, S., Durkes, A. C., Turner, J. R. & Bhunia, A. K. *Listeria* adhesion protein induces intestinal epithelial barrier dysfunction for bacterial translocation. *Cell. Host Microbe*. **23**, 470–484e7. https://doi.org/10.1016/j.chom.2018.03.004 (2018).
33. Kim, K. P. et al. Adhesion characteristics of Listeria adhesion protein (LAP)-expressing *Escherichia coli* to Caco-2 cells and of recombinant LAP to eukaryotic receptor Hsp60 as examined in a surface plasmon resonance sensor. *FEMS Microbiol. Lett.* **256**, 324–332. https://doi.org/10.1111/j.1574-6968.2006.00140.x (2006).
34. Moraes, C. T. et al. Flagellin and GroEL mediates *in vitro* binding of an atypical enteropathogenic *Escherichia coli* to cellular fibronectin. *BMC Microbiol.* **15**, 278. https://doi.org/10.1186/s12866-015-0612-4 (2015).
35. Monteagudo-Mera, A., Rastall, R. A., Gibson, G. R., Charalampopoulos, D. & Chatzifragkou, A. Adhesion mechanisms mediated by probiotics and prebiotics and their potential impact on human health. *Appl. Microbiol. Biotechnol.* **103**, 6463–6472. https://doi.org/10.1007/s00253-019-09978-7 (2019).
36. Leshem, A., Liwinski, T. & Elinav, E. Immune-Microbiota interplay and colonization resistance in infection. *Mol. Cell.* **78**, 597–613. https://doi.org/10.1016/j.molcel.2020.03.001 (2020).
37. Van Zyl, W. F., Deane, S. M. & Dicks, L. M. T. Molecular insights into probiotic mechanisms of action employed against intestinal pathogenic bacteria. *Gut Microbes*. **12**, 1831339. https://doi.org/10.1080/19490976.2020.1831339 (2020).
38. Gonçalves-Oliveira, J. et al. Exploring the diversity and evolutionary strategies of prophages in Hyphomicrobiales, comparing animal-associated with non-animal-associated bacteria. *BMC Microbiol.* **24**, 159. https://doi.org/10.1186/s12866-024-03315-3 (2024).
39. Kozinski, A. W., Doermann, A. H. & Kozinski, P. B. Absence of interparental recombination in multiplicity reconstitution from incomplete bacteriophage T4 genomes. *J. Virol.* **18**, 873–884. https://doi.org/10.1128/JVI.18.3.873-884.1976 (1976).
40. Ramirez, J. et al. Antibiotics as major disruptors of gut microbiota. *Front. Cell. Infect. Microbiol.* **10**, 572912. https://doi.org/10.3389/fcimb.2020.572912 (2020).
41. Hyun, J. et al. Faecal microbiota transplantation reduces amounts of antibiotic resistance genes in patients with multidrug-resistant organisms. *Antimicrob. Resist. Infect. Control.* **11**, 20. https://doi.org/10.1186/s13756-022-01064-4 (2022).
42. Burstein, D. et al. Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nat. Commun.* **7**, 10613. https://doi.org/10.1038/ncomms10613 (2016).

43. Botelho, J. Defense systems are pervasive across chromosomally integrated mobile genetic elements and are inversely correlated to virulence and antimicrobial resistance. *Nucleic Acids Res.* **51**, 4385–4397. https://doi.org/10.1093/nar/gkad282 (2023).
44. Burke, K. A., Urick, C. D., Mzhavia, N., Nikolich, M. P. & Filippov, A. A. Correlation of *Pseudomonas aeruginosa* phage resistance with the numbers and types of antiphage systems. *Int. J. Mol. Sci.* **25**, 1424. https://doi.org/10.3390/ijms25031424 (2024).
45. Blin, K. et al. AntiSMASH 6.0: Improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **49**, W29–W35. https://doi.org/10.1093/nar/gkab335 (2021).
46. Štefanac, T., Grgas, D. & Landeka Dragičević, T. Xenobiotics-division and methods of detection: a review. *J. Xenobiot.* **11**, 130–141. https://doi.org/10.3390/jox11040009 (2021).
47. Mishra, S. et al. Recent advanced technologies for the characterization of xenobiotic-degrading microorganisms and microbial communities. *Front. Bioeng. Biotechnol.* **9**, 632059. https://doi.org/10.3389/fbioe.2021.632059 (2021).
48. Zhu, F. et al. Metagenomic analysis exploring microbial assemblages and functional genes potentially involved in di (2-ethylhexyl) phthalate degradation in soil. *Sci. Total Environ.* **715**, 137037. https://doi.org/10.1016/j.scitotenv.2020.137037 (2020).
49. Parlindungan, E., Lugli, G. A., Ventura, M., van Sinderen, D. & Mahony, J. Lactic acid bacteria diversity and characterization of probiotic candidates in fermented meats. *Foods* **10**, 1519. https://doi.org/10.3390/foods10071519 (2021).
50. Parlindungan, E. et al. Dairy streptococcal cell wall and exopolysaccharide genome diversity. *Microb. Genom.* **8**, 000803. https://doi.org/10.1099/mgen.0.000803 (2022).
51. L'Huillier, A. G. et al. Laboratory testing and phylogenetic analysis during a mumps outbreak in Ontario, Canada. *Virol. J.* **15**, 98. https://doi.org/10.1186/s12985-018-0996-5 (2018).
52. Hall, B. G. & Nisbet, J. Building phylogenetic trees from genome sequences with kSNP4. *Mol. Biol. Evol.* **40**, msad235. https://doi.org/10.1093/molbev/msad235 (2023).
53. Tamura, K., Stecher, G. & Kumar, S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38**, 3022–3027. https://doi.org/10.1093/molbev/msab120 (2021).
54. Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**, 512–526. https://doi.org/10.1093/oxfordjournals.molbev.a040023 (1993).
55. Felsenstein, J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**, 783–791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x (1985).
56. Strepis, N. et al. Genome-guided analysis allows the identification of novel physiological traits in *Trichococcus* species. *BMC Genom.* **21**, 24. https://doi.org/10.1186/s12864-019-6410-x (2020).
57. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**, 33–36. https://doi.org/10.1093/nar/28.1.33 (2000).
58. Tatusov, R. L., Koonin, E. V. & Lipman, D. J. A genomic perspective on protein families. *Science* **278**, 631–637. https://doi.org/10.1126/science.278.5338.631 (1997).
59. Martín, R. et al. The infant-derived *Bifidobacterium bifidum* strain CNCM I-4319 strengthens gut functionality. *Microorganisms* **8**, 1313. https://doi.org/10.3390/microorganisms8091313 (2020).
60. Lagesen, K. et al. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108. https://doi.org/10.1093/nar/gkm160 (2007).
61. Lowe, T. M. & Chan, P. P. tRNAscan-SE On-line: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* **44**, W54–W57. https://doi.org/10.1093/nar/gkw413 (2016).
62. Kanehisa, M. & Goto, S. K. E. G. G. Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. https://doi.org/10.1093/nar/28.1.27 (2000).
63. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**, W182–W185. https://doi.org/10.1093/nar/gkm321 (2007).
64. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726–731. https://doi.org/10.1016/j.jmb.2015.11.006 (2016).
65. Galperin, M. Y. et al. COG database update: Focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* **49**, D274–D281. https://doi.org/10.1093/nar/gkaa1018 (2021).
66. Liu, B., Zheng, D. D., Zhou, S. Y., Chen, L. H. & Yang, J. VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res.* **50**, D912–D917. https://doi.org/10.1093/nar/gkab1107 (2022).
67. Cosentino, S., Voldby Larsen, M., Møller Aarestrup, F. & Lund, O. PathogenFinder - distinguishing friend from foe using bacterial whole genome sequence data. *PLoS One.* **8**, e77302. https://doi.org/10.1371/journal.pone.0077302 (2013).
68. Wishart, D. S. et al. PHASTEST: Faster than PHASTER, better than PHAST. *Nucleic Acids Res.* **51**, W443–W450. https://doi.org/10.1093/nar/gkad382 (2023).
69. McArthur, A. G. et al. The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.* **57**, 3348–3357. https://doi.org/10.1128/AAC.00419-13 (2013).
70. Grissa, I., Vergnaud, G. & Pourcel, C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* **35**, W52–W57. https://doi.org/10.1093/nar/gkm360 (2007).
71. Payne, L. J. et al. PADLOC: A web server for the identification of antiviral defence systems in microbial genomes. *Nucleic Acids Res.* **50**, W541–W550. https://doi.org/10.1093/nar/gkac400 (2022).
72. Yoon, S. H., Ha, S. M., Lim, J. M., Kwon, S. J. & Chun, J. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek.* **110**, 1281–1286. https://doi.org/10.1007/s10482-017-0844-4 (2017).

Acknowledgements

This research was funded by the National Research Council of Thailand (grant number PHD60I0089); the Biology Department, Faculty of Science, Chiang Mai University, Thailand and the APC Microbiome Ireland, University College Cork, Ireland through the financial support of Science Foundation Ireland (SFI) through the Irish Government's National Development Plan (Grant numbers SFI/12/RC/2273a and SFI/12/RC/2273b).

Author contributions

P.R., E.P., J.M., N.T., and D.vS. designed the experiments. P.R., W.S. and E.P. analyzed the data. P.R., J.M., N.T., and D.vS. investigated the data. N.T. and D.vS. acquired the funding. P.R. and N.T. prepared the original draft. P.R., J.M., N.T., and D.vS. reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-86602-x.

Correspondence and requests for materials should be addressed to N.T. or D.S.

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

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2025