Comparative genomics and evolutionary analysis of *Lactococcus garvieae* isolated from human endocarditis

Carlos Francés-Cuesta[1,*], Iván Ansari[1], José Francisco Fernández-Garayzábal[2,3], Alicia Gibello[2]† and Fernando González-Candelas[1,*]

Abstract

Lactococcus garvieae is a well-known pathogen of fish, but is rarely involved in infections in humans and other mammals. In humans, the main clinical manifestation of *L. garvieae* infections is endocarditis usually related to the ingestion of contaminated food, such as undercooked fish and shellfish. This study presents the first complete genomic sequence of a clinical *L. garvieae* strain isolated from a patient with endocarditis and its comparative analysis with other genomes. This human isolate contains a circular chromosome of 2099060 bp and one plasmid of 50557 bp. In comparison with other fully sequenced *L. garvieae* strains, the chromosomal DNA of *L. garvieae* Lg-Granada carries a low proportion of insertion sequence elements and a higher number of putative prophages. Our results show that, in general, *L. garvieae* is a highly recombinogenic species with an open pangenome in which almost 30% of its genome has undergone horizontal transfers. Within the genus *Lactococcus*, *L. lactis* is the main donor of genetic components to *L. garvieae* but, taking Lg-Granada as a representative, this bacterium tends to import more genes from *Bacilli* taxa than from other *Lactococcus* species.

DATA SUMMARY

Lg-Granada genome sequencing data have been deposited at NCBI database under project PRJNA763925 with accession numbers CP084377.1 and CP084378.1 for the chromosome and plasmid, respectively. All supporting data are provided in supplementary data files.

INTRODUCTION

The genus *Lactococcus* comprises 21 species, with *Lactococcus lactis* and *Lactococcus garvieae* being the most well known. *L. lactis* is considered safe for humans and animals and it is commonly used in the dairy industry [1]. However, *L. lactis* has also been associated sporadically with bovine mastitis [2, 3] and human infections [4, 5]. In contrast, *L. garvieae* is an important pathogen for freshwater and marine fishes and shellfish [6] and it is also commonly isolated from infections in ruminants, pigs and humans [3–14].

The role of *L. garvieae* as an emerging human pathogen has gained relevance recently, being associated with different clinical manifestations, such as urinary infections, meningitis, bacteraemia, peritonitis, liver abscess, osteomyelitis, endophthalmitis

Impact Statement

Understanding how a bacterial species becomes pathogenic for humans is one of the possibilities offered by the analysis of complete genome sequences. Here, we have studied the genome of a clinical isolate of a rare human pathogen, *Lactococcus garvieae*, obtained from a deceased individual. Its sequence differs from that of other human clinical isolates and includes 64 genes not detected previously in this species. Some of these are apparently related to mobile genetic elements and this observation prompted us to analyse the role of horizontal gene transfer in the genome evolution of this species. We have studied gene transfers at three different levels: among strains of *L. garvieae*, from other *Lactococcus* species into *L. garvieae* and from other species of the class *Bacilli* into the target species. Our results show that recombination and horizontal gene transfer have played and still are playing a major role in the evolution of *L. garvieae*: almost 30% of the genes shared by all the strains of this species have been involved in transfer from other strains and species from the same genus. We have identified another human pathogen, *L. lactis*, as the main source of genes transferred into the *L. garvieae* genome.

and, most notably, endocarditis, which represents more than 50% of the total human infections associated with this pathogen [7, 8, 11–14]. Apparently, the source of human infections is the ingestion of contaminated food, mainly raw fish, seafood or crude dairy products, or contact with infected animals [11–13]. However, not all the *L. garvieae* strains are pathogenic to humans and some strains have been reported as members of the autochthonous microbiota of dairy products, contributing to their organoleptic properties [15–18].

Whole-genome sequencing (WGS) is gaining importance in the analysis of bacterial pathogens to identify virulence factors and genes involved in their adaptation to the host. WGS studies allow comparative genomic analyses of bacterial populations, providing new insights into their genetic diversity and evolution. The genome sequences of 43 *L. garvieae* strains isolated from different sources were available at NCBI as of January 2021 (https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/699/). Only nine genomes have been closed and none of them was isolated from a human clinical source. The two available genomes from human clinical isolates – strains 21881 and Lg-ilsanpaik-gs201105 – are not completely closed.

In this study, we present the first completely closed genome of a human clinical isolate of *L. garvieae* – strain Lg-Granada – to assess its genomic features, metabolic diversity and phylogenetic relationships, and to explore the relevance of recombination and horizontal gene transfer (HGT) at the intra- and inter-species (genus and class) levels.

METHODS

Bacterial characterization, DNA extraction and high-throughput sequencing

L. garvieae strain Lg-Granada was isolated in pure culture from a fatal case of infective endocarditis [19]. API 50CH and API ZYM strips (bioMérieux) were used to determine the biochemical profile of Lg-Granada according to the manufacturer's instructions.

The strain was cultured using log-phase cultures (OD$_{600}$, ~1) in BHI (Panreac AppliChem), and DNA was extracted using the Blood and Cell Culture DNA Maxi Kit (Qiagen) according to the manufacturer's specifications. Quality control was assessed and concentration of the DNA was measured by both photometric (NanoDrop; Thermo Fisher Scientific) and fluorometric (PicoGreen; Promega) methods. The DNA had a good quality, with a 260/280 ratio of 1.8, a 260/230 ratio of 2.2 and a concentration of 462.6 ng µl^{-1}.

The genomic DNA was sheared and size selected to produce ~15–25 kb insert-size libraries using BluePippin (Sage Science) according to the manufacturer's size-selection system protocol. The selected fragments were sequenced on a PacBio RS II platform (Pacific Biosciences) using a P6-C4 polymerase-chemistry combination, and data acquisition time of ≥4 h.

Genome assembly and annotation

Quality checking of the reads and *de novo* assembly were performed using HGAP 3.0 (SMRT analysis v.2.3.0) [20] with default parameters and the minimum seed read length set at 6000 bp. The first step of this approach was preassembly, which generated long and highly accurate sequences by mapping single pass reads to long reads (seeds). In this step, reads were filtered using minimal polymerase read length and minimal subread length of 500 bp, and minimal polymerase read quality of 0.8. The second step of the approach was assembly using the Overlap Layout Consensus (OLC) algorithm implemented in the Celera WGS-assembler v.7.0 [21]. We applied a polishing step using the Quiver algorithm implemented in the Genomic Consensus package (https://github.com/PacificBiosciences/GenomicConsensus) to reduce the remaining indel and base errors in the assembly. Finally, the resulting contigs were circularized using Circlator [22].

The assembled genome sequences were annotated using Prokka v.1.11 [23]. The predicted coding sequences (CDSs) were further annotated using the MvirDB database [24] of virulence factors, the CARD database [25] of antibiotic resistance genes and the UniProtKB/Swiss-Prot database [26] to obtain the associated Gene Ontology (GO) terms.

Characterization of mobile genetic elements

The genomes of strains Lg-Granada, ATCC49156, Lg2, 122061, M14 and DSM 20684 were analysed for the presence of prophage-like elements using PHASTER [27] and PhiSpy [28] tools. To confirm the presence of prophages, the detected regions were manually curated and annotated using BLAST [29].

Manual reannotation of plasmid pLG50 was performed using ORF Finder [30] and the RAST Server [31]. We looked for putative ORFs encoding proteins of more than 30 aa and flanked by an upstream potential Shine–Dalgarno sequence with homology to others. The functions of these ORFs were presumed by comparing deduced protein sequences with the GenBank database using BLAST as well as by identifying conserved protein domains.

Insertion sequences (IS) and transposons were detected using ISfinder [32] and the results were curated using BLAST.

Core genome identification

The core genomes corresponding to three different taxon levels were identified. We downloaded the 23 genomes of *L. garvieae* available at NCBI as of October 2018 (Table S1), six closed genomes of the genus *Lactococcus* (Table S2) and 19 closed genomes of the class *Bacilli* (Table S3). The Lg-Granada strain was included in the three analyses. The average nucleotide identity (ANI) between pairs of *L. garvieae* isolates was calculated using pyani v.0.2.10 [33] to have an estimate of the divergence among the genomes. The orthologous genes (OGs) between the genomes in each taxon were ascertained using the BLAST-based program Proteinortho v.5.16b [34]. Default parameters were used except for the minimum similarity, which was adjusted to 80% for the species level, and 70% for the genus and class levels. The core genome was determined as strict core – common orthologues for all the isolates of the corresponding level – and relaxed core – orthologues shared by ≥80% isolates.

OGs included in the strict and relaxed cores were aligned using MACSE v.1.2 [35], and concatenated using Python scripts. The resulting multiple alignments were used for phylogenetic reconstruction by maximum-likelihood (ML) using the GTR+F+I+G4 nucleotide substitution model and 10 000 ultrafast bootstrap (UFBoot) replicates [36]. The phylogenetic reconstructions were performed using IQ-TREE v.1.6.1 [37]. For the three taxonomic levels, both the strict and relaxed core phylogenetic trees were topologically identical.

Recombination detection

The detection of recombinant genes was based on a phylogenetic incongruence approach [38]. First a likelihood mapping (LM) test [39] was performed for each OG alignment included in the relaxed core in order to determine its phylogenetic signal [40]. This analysis was performed using IQ-TREE v.1.6.1 with 10 000 quartets randomly drawn. Additionally, the proportion of pairwise-informative SNPs of each alignment was also computed. Only the OGs with ≥70% of quartets completely resolved in the LM test and with a proportion of informative sites of at least 10% from the total length of the alignment were eligible to proceed to the next stage of the analysis.

Next, the OGs that met the requirements were tested for topology congruence. To do this, an ML tree was reconstructed for each OG using the same parameters as in the phylogenetic reconstruction of the core genomes. Then, the topologies of each OG and the core genome trees were compared using the Shimodaira–Hasegawa (SH) [41] and Expected-Likelihood Weights (ELW) [42] tests. OGs that rejected the core genome topology in both tests were selected as recombinants. Due to the large number of results at the species level, a correction in SH P-values was applied using the false discovery rate (FDR) method implemented in the p.adjust function in R v.3.6.3 [43].

The recombinant gene trees were visually compared with the core genome tree using Phylo.io [44] in order to see the movement of the genes between strain Lg-Granada and the other species included in the genus and class levels (Fig. S1). If two or more consecutive genes determined as recombinant in or horizontally transferred to the Lg-Granada chromosome had the same source, these genes were included in a single, larger recombination event.

RESULTS AND DISCUSSION

Main features of *L. garvieae* strain Lg-Granada genome

The complete genome of *L. garvieae* strain Lg-Granada has a 2099060 bp chromosome and one plasmid of 50557 bp, named pLG50 (Fig. 1), with average GC content of 38.74 and 33.70%, respectively. Both the length of the chromosome and its GC content are within the range obtained for the species (Table S1). The genome contains a total of 2,164 CDS – 2101 in the chromosome and 63 in the plasmid – and 81 structural RNAs – 16 rRNAs and 65 tRNAs.
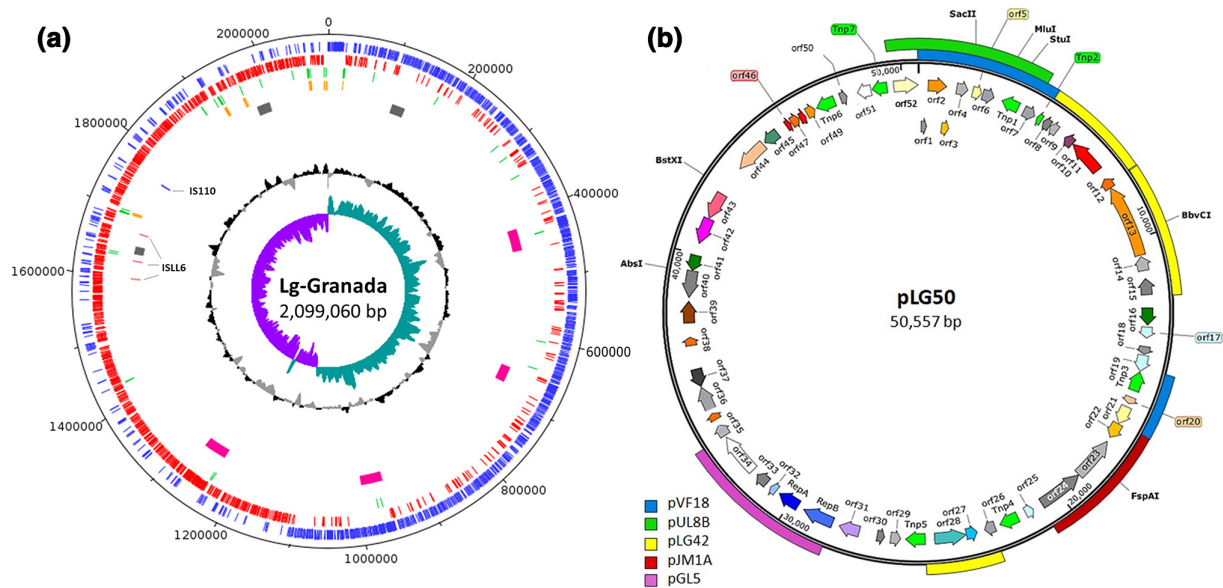
Fig. 1. (a) Chromosome of *L. garvieae* Lg-Granada showing, from outer to inner circles, the forward strain coding sequences (blue), the reverse strain coding sequences (red), the tRNA positions (green), the rRNA positions (orange), the three exclusive regions of Lg-Granada (grey), the phages (pink), the insertion sequences (blue and red, labelled), the GC content (black/grey plot) and the GC skew (purple/teal plot). (b) Map of plasmid pLG50, in which ORFs are indicated by arrows showing the direction of transcription, and *Tnp*'s indicate genes encoding transposases of IS (green). Similarity to other plasmids from lactic acid bacteria are indicated by external-coloured circles (see key). The chromosome was depicted using DNAPlotter v.18.0.2 [73] and plasmid using SnapGene (https://www.snapgene.com/).

Most chromosomal genes encode cytoplasmic and membrane proteins, an important part being those involved in the metabolism of the carbon source and energy. Thus, Lg-Granada is able to metabolize different sugars such as D-ribose, D-glucose, D-fructose, D-mannose, D-galactose, D-mannitol, amygdalin, arbutin, *N*-acetylglucosamine, aesculin, salicin and cellobiose when tested with the API 50CH strips.

The chromosome of Lg-Granada has several regions that seem to be totally or partially absent in the genome of other *L. garvieae* strains with a closed genome available – ATCC 49156, Lg2, 122061, M14, and DSM 20684. The most remarkable is a region of 27324 bp inserted at position 1507043 – ATCC49156 strain coordinates – absent in the other strains. This region contains some genes (loci _02048 to _02073) encoding conjugative transfer proteins, a sortase A and a new LPXTG protein (Table S4), with a genetic organization resembling those of some conjugative plasmids of other Gram-positive bacteria, such as pCW3 from *Clostridium perfringens* [45]. The existence of a neighbouring gene (locus _02042) encoding an integrase could be also indicative of a vestigial integrative conjugative element (ICE) in the Lg-Granada chromosome. LPXTG proteins are surface proteins anchored to the cell-wall peptidoglycan by sortases, often being essential for the infectivity and survival of Gram-positive bacteria in the host, contributing to their pathogenicity [46–48].

There is a second large region of 22620 bp (loci _00111 to _00130) in the Lg-Granada chromosome inserted at position 120,172 – ATCC49156 coordinates – which is absent from strains ATCC49156, Lg2 and DSM 20684. In addition, loci _00118 to _00121, and _00130 (Table S4) are also absent from strain 12061. This region contains genes involved in the synthesis of cell surface components, such as exopolysaccharides and teichoic acids (TAs) – genes *tagD*, *tagG* and *tagH*. Additionally, there is a 3429 bp fragment in the Lg-Granada genome that contains three genes of the same operon (loci _01425 to _01427) encoding conserved proteins of unknown function and the protein TagF involved in the polymerization of the main chain of TA (Table S4). These genes are inserted between the extra copies of genes *tagD* (locus_01424) and *tagG* (locus_01428), belonging to the *tag* operon. The function of all these genes has not been examined directly but *L. garvieae* mutants in the biosynthesis of TAs showed a lower proliferation rate in rainbow trout than normal pathogenic strains [49, 50]. In Gram-positive bacteria, TAs are involved in the activity of cell-wall hydrolases, the regulation of cell-wall elongation and cell division, transport processes, resistance to antimicrobial peptides and lysozymes, homoeostasis, and the interaction with host factors [51]. Therefore, these genes could provide new features to the bacterial surface relevant for bacterial infection.

Another region of 13610 bp (loci _01702 to _01712) inserted at position 1407979 – ATCC49156 coordinates – is absent in strain Lg2 and partially absent in strain 12061. This region contains a gene cluster encoding enzymes involved in the metabolism of aromatic amino acids, with three genes exclusive of strain Lg-Granada (loci _01710 to _01712, Table S4).

When compared with other strains of the species used in this study (Table S1), Lg-Granada has 64 unique genes, some of which encode phage proteins, and others are involved in metabolic and transport processes (Table S5). If we analyse the genes that Lg-Granada shares exclusively with the other strains according to their isolation source, we can see that it shares 34 genes with one or more strains of animal origin (some phage proteins, or proteins with catalytic activity), 21 genes with strains isolated from food (some proteins involved in metabolism, but also plasmid proteins involved in pathogenesis and defence against other bacteria), one gene of unknown function with a strain isolated from the soil, and one gene of unknown function with the strains of human origin (Table S6). Thus, Lg-Granada seems to share more genes with strains of animal and food origin than with strains isolated from human hosts.

Mobile genetic elements: plasmids

Accessory genetic elements – plasmids, prophages, transposons and islands – are known to frequently encode virulence factors. Plasmid pLG50 contains 63 CDS, with seven IS elements belonging to IS6 family transposases (Tnp1 to Tnp7; Fig. 1b). Predicted functions were assigned to 38 CDS (the most representative ones are detailed in Table S7). The pLG50 plasmid apparently replicates by a theta-replication mechanism, which is common to large and low-copy-number plasmids. The replication module carries three clustered genes (*orf31*, *repB* and *repA*) that are transcribed in the same direction. The gene *orf31* (locus _02218) encodes a protein which contains the XRE-family HTH domain, sharing 99.18% identity in 99% of the amino acids with transcriptional regulator proteins. Downstream to this gene is an AT-rich intergenic sequence followed by three directly repeated sequences of 22 bp – DR, iterons – that are thought to interact directly with the Rep proteins to initiate replication (Fig. S2). Before the iterons, there are two inverted repeat sequences of 5–12 bp, designated as IRa and IRb. that usually overlap the putative promoter region of the *repB* gene encoding the replication initiator protein [52]. The gene *repA* encodes a partitioning protein involved in the correct distribution of large and low-copy-number plasmids into daughter cells at cell division [53, 54]. These plasmid replication genes shared more than 95% identity with homologous genes in the pGL5 plasmid of *L. garvieae* strain 21881 [54]. Mobility is one of the most important features of plasmid molecules, but neither relaxases nor proteins involved in plasmid mobilization were found in pLG50.

pLG50 carries a putative operon (*orf1* to *orf4*) with 99.9% identity to the homologous genes (*orf18* to *orf21*) of plasmid pVF18 involved in Co^{2+} uptake [55]. The *orf2*, *orf3* and *orf4* genes also exhibit a high identity (98.9–100%; Table S7) to the Fe^{2+} export FetABC transporter present in related bacteria, which indicates that these proteins might be involved in the transport of both compounds. The genes *orf11* and *orf13* (loci _02194 and _02196, respectively) encode extracellular cysteine proteases (Table S7), which are considered to be virulence factors in *Staphylococcus* and *Giardia* [56, 57]. Both genes are located in different operons, because there is a putative terminator sequence (CCACTGTCTCAGTGG) downstream of *orf11*. These two genes share an identical DNA sequence of 820 nt that suggests a common origin. In *Staphylococcus*, cysteine proteases have been demonstrated *in vitro* as virulence factors to prevent neutrophil recruitment induced by CXCR2 ligands [56]. Cysteine proteases are also virulence factors in *Giardia*, involved in destruction of the intestinal epithelial barrier and chemokine degradation [57].

The sequence of pLG50 indicates a complex plasmid structure organized in functional modules or cassettes of different origins, mainly from plasmids of *L. garvieae* and *L. lactis* (Fig. 1b). Several sequences of pLG50 exhibited high identity values (>90%) with sequences of plasmid pGL5 from the clinical strain *L. garvieae* 21881, and with plasmid sequences of *L. garvieae* strains isolated from dairy products, such as pLG42 of *L. garvieae* IPLA 31405. Moreover, sequences of pLG50 exhibit also significant identity to different plasmids of *L. lactis*, such as pVF18 (99.83% identity), pUL8B (97.68% identity) and pJM1A (99.12% identity). As in other plasmids of *Lactococcus* [52], pLG50 has two genes (*orf42* and *orf43*) encoding an abortive bacterial infection (Abi) system, to avoid phage propagation in Lg-Granada. Other genes involved in the defence mechanism found in pLG50 were *orf29,* which encodes a DNA-cytosine methylase, and *orf46* to *orf49,* which encode two bacteriocins with some similarity to Garvieacin Q [58] and their corresponding immunity proteins (Table S7).

Mobile genetic elements: prophages

Four prophages (P1Lg-Granada to P4Lg-Granada) were found on the chromosome (Fig. S3), with a GC content ranging from 34.72 to 36.46%, a lower value than that of the whole chromosome (38.74%). Their length, integration *att* sites and CDS numbers are detailed in Table S8. CDS with predicted function are shown in Table S9. These prophages belong to the order *Caudovirales*, family *Siphoviridae,* the most abundant phages described in *L. garvieae* [59, 60].

P1Lg-Granada was integrated into a transfer-messenger RNA and shares 95.75% identity in a length coverage of 58% with phage PLgT-1 of strain Lg2 (accession number KU892558). However, P1Lg-Granada contains only a tRNA-Thr while PLgT-1 contains a tRNA-Lys and a tRNA-Met. P1Lg-Granada appears to be a defective prophage because it does not have a gene encoding the endolysin in the lysis module (Table S9). A search in the GenBank database also revealed the presence of similar sequences between P1Lg-Granada and the temperate phage of strain UNIUD074 (accession number NZ_AFHF01000007), with 93.73% identity in a length coverage of 52%. P1Lg-Granada also showed high identity with fragments of *L. lactis* phages, such as TP12 (accession number AY766464), with 95.3% identity in a length coverage of 42%. Moreover, three ORFs of P1Lg-Granada (*orf 33, 39* and *40*) showed high similarity – 98.3 and 96.2% overlap – with chromosomal sequences from *L. lactis* subsp. *lactis* (Table S9).

P2Lg-Granada was integrated in the 3′-end of the genes involved in the putrescine/spermidine transporter. It appears to be a defective prophage because the gene coding for integrase was not in the *att* sequences. It contains one tRNA-Trp. P2Lg-Granada shared 97.76% identity with length coverage of 59% with the phage 2 of strain M14 (Fig. S3). Furthermore, several genes of this phage had high similarity (>98%) with PLg-TB25 (accession number KX833905; Table S9).

P3Lg-Granada was integrated near a gene encoding subunit II of the maltose/glucose phosphotransferase ABC transport system. It has 89.02% identity with length coverage of 7% with the phage PLg-TB25, mainly with genes involved in cell lysis (Table S9). Nucleotide identity values lower than 20% were also detected with phage 1 of strain IPLA 31405 (accession number NZ_AKFO000017.1), and with phage 1 of strain M14 (Fig. S3). These low identities suggest the novelty of P3Lg-Granada. In fact, the highest DNA identity was found with phage vB_BthS-HD29phi of *Bacillu*s (accession number MN065183), with 73.55% identity over 15% of the coverage length.

P4Lg-Granada was integrated near the genes coding for proteins involved in sulphur metabolism, and contains a tRNA-Ser. No gene encoding endolysin was found; thus, it could be a defective phage. It is most closely related to the bIL286 and proPhi2 phages of *L. lactis* (accession numbers AF323669 and MN534316, respectively), with 93% identity over 45% of the coverage length. Regions of similarity included the modules involved in head and tail structural components and assembly (Table S9). However, two genes of P4Lg-Granada (*orf53* and *orf59*) showed >97% identity over 100% of the coverage length with the homologous genes in *L. garvieae* Plg-TB25 phage (Table S9).

Overall, the four prophages of Lg-Granada shared significant identities with fragments of phages from other strains of *L. garvieae*, mainly isolated from milk or dairy products. The Lg-Granada prophages showed also a high similarity with *L. lactis* phage*s*. In this sense, Lg-Granada prophages had GC content lower than that of their host, being closer to that exhibited by *L. lactis*, which might indicate they originated from this bacterial species.

Mobile genetic elements: insertion sequences

Overall, 12 copies of IS from seven distinct IS elements belonging to four different families – IS30, IS3, IS6 and IS110 – were identified in the Lg-Granada strain. Most of these IS elements have been found previously in *L. garvieae* and *L. lactis* [61]. Seven IS elements (transposases) were identified in plasmid pLG50 (Fig. 1b), one assigned to the family IS30 and six assigned to the family IS6.

Two different IS types – ISLL6, which is member of the IS3 family, and IS110 – were located on chromosome. Lg-Granada carried three copies of ISLL6 (loci _01645, _01686 and _01738), and two copies of the IS110 family (loci _01839 and _01840). ISLL6 is widely distributed in *L. garvieae*, while IS110 is not frequently detected [61]. These IS were located on a region ranging from position 1594039 to 1763734, suggesting a high chromosomal instability in this area.

The low number of IS in the chromosome of Lg-Granada differentiates this strain from other *L. garvieae* strains that harbour more IS in their chromosomes. Strains Lg2 and ATCC 49156 thus contain 23 ISLL6 and a few other IS, strain 122061 contains 26 IS (17 belonging to ISLL6), strain DSM 20684 contains 26 IS (12 IS110 and other IS belonging to IS3, IS982, and IS256), and strain M14 contains four copies of the IS6 family, 25 copies of IS3, and a variety of copies belonging to IS4, IS30, IS110 and IS981 families [61].

The presence of IS in bacteria has an important impact on genome structure and function. IS expansion has been commonly observed in bacteria with recently adopted fastidious, host-restricted lifestyles, previous to their genome reduction. Thus, IS expansion is considered an early step in the evolutionary process of adaptation of pathogenic bacteria to their hosts [62]. Obligate pathogens or endosymbionts usually show low IS and a significant genome reduction in comparison with their ancestral bacteria because in the nutritionally rich environment of the host, many genes of free-living bacteria are not essential, and deletion of superfluous genes could increase fitness [62, 63]. The genome size of Lg-Granada is slightly larger than that of other pathogenic strains of *L. garvieae* (Table S1). These data, together with its low IS content compared with other pathogenic strains, suggests that Lg-Granada, despite being isolated from a case of infective endocarditis, would not be adapted to a pathogenic lifestyle. In fact, Lg-Granada affected a 68-year-old patient with an extensive history of heart disease and several other predisposing factors [19].

Phylogenetics, pangenome and core genome of *L. garvieae*

The phylogenetic tree obtained with the 24 strains of *L. garvieae* reflects the genetic heterogeneity reported for this pathogen [64], grouping the strains in four different clusters (Fig. 2). The Lg-Granada strain was close to the food-borne strains Tac2, UBA5784, UBA11300 and IPLA 31405, with which it exhibited ANI values above 99% (Table S10). Curiously, Lg-Granada was located in a different cluster than the other strains isolated from human infections – 21881 and Lg-ilsanpaik-gs201105 – with which it shares ANI values >94%. The two clusters that included human clinical strains also included strains from infected fish and food, supporting the genetic relatedness between food and human *L. garvieae* isolates previously deduced by multilocus sequence typing [65]. Strains A1 and DCC43 were clearly distant from the other *L. garvieae* strains, with which they share ANI values ≤82.03%. These values were lower than the proposed and generally accepted ANI cut-off value of 95–96% for species
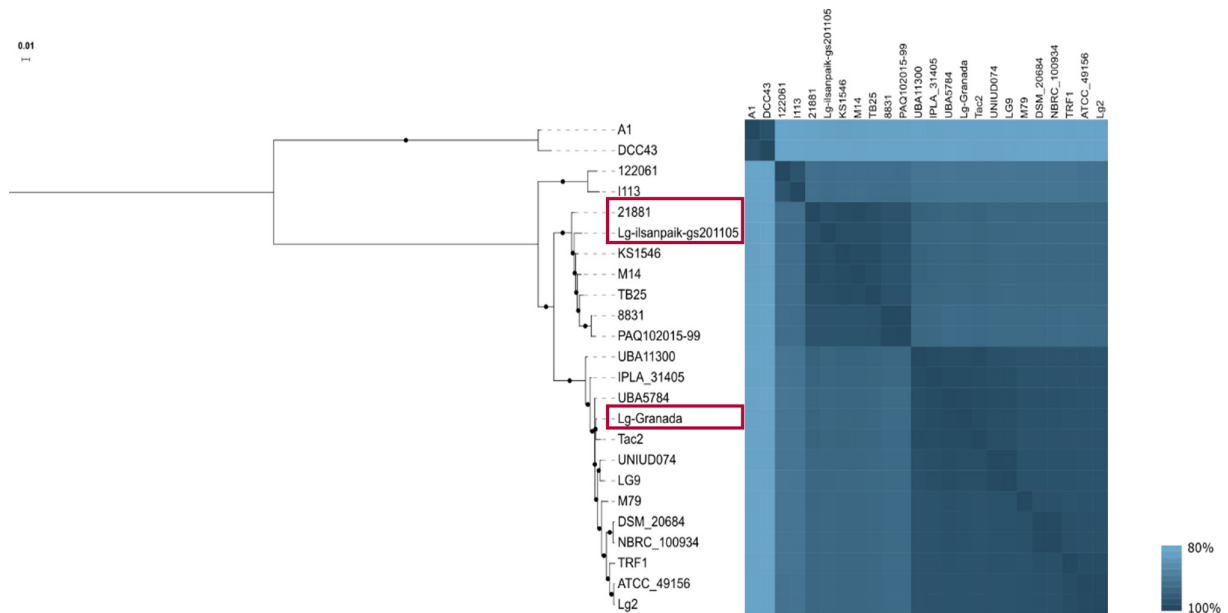
Fig. 2. Phylogenetic tree showing the relationships between *L. garvieae* strains used in this study. Black dots within branches indicate a bootstrap support value ≥90%. The associated heat-map represents pairwise ANI values. Lg-Granada and the other strains isolated from clinical samples are highlighted. The figure was created using iToL v.4 [74].

delineation [66], therefore suggesting that strains A1 and DCC43 should not be classified within *L. garvieae*. In fact, the average ANI value for all 24 strains was 93.14% but, after removal of these two strains, the average value increased to 95.6%. Further taxonomic studies including all *Lactococcus* species should be performed to determine if both strains represent a novel species of *Lactococcus* or are identified as one of the currently recognized species of the genus. However, these analyses are beyond the objective of the present study.

The average number of CDS in the 24 *L. garvieae* genomes is 1975 (ranging from 1781 to 2167; Table S1). The strict core genome has 1157 genes, spanning 1098087 bp of which 277438 are variant positions. That is, the 24 strains used in this study share 58.6% of the CDS over the average value, covering approximately half the length of the *L. garvieae* genome. These values increase when we relax the core, including genes shared by at least 80% of the genomes (19–24 genomes), resulting in 1556 shared genes, spanning 1488588 bp with 391094 variant positions. Our estimate of the strict core genome of *L. garvieae* is lower than that obtained by Ferrario *et al.* [64], who estimated it at 1341 genes. However, our result is consistent with theirs if we consider that twice as many *L. garvieae* genomes have been used in this new work. Another recent study [67] estimated the core genome of the species at 1850 genes based on only eight genomes of *L. garvieae*. This value is inconsistent with our results – around 1350 genes for eight genomes – which may be due to the usage of less stringent parameters in the orthologous identification step (e.g. lower percentages of identity and similarity).

The pangenome of *L. garvieae* reaches 5031 genes. This number is based on analysis of the 24 genomes used in this study, but it would probably be larger if there were more genomes available at the time of the analyses, as suggested by the trend of the rarefaction curve in Fig. 3. In contrast, the core genome appears to reach a plateau at 23 genomes, so the number of core genes for the species seems to be around 1100–1200. Regarding gene frequencies (Fig. 4), unique genes, i.e. the genes present in every single genome, represent 34% of the pangenome (1708 genes), an amount immediately followed by the strict core genes (23% of the pangenome).

Intraspecies recombination in *L. garvieae*

The study of recombination in bacteria has contributed to understanding their evolution, adaptability and pathogenesis [68]. The detection of recombination could facilitate the identification of regions of interest in pathogen genomes [69]. However, despite its important role in the acquisition of virulence and pathogenic genes, recombination has not been well studied in *L. garvieae*. Here, we used the Lg-Granada strain to explore intra- and interspecific HGT of *L. garvieae* at different taxonomic levels.

A total of 739 OGs from the relaxed core passed the LM test, and 698 of them also had an adequate proportion of informative sites to be subjected to topological congruence tests. Finally, of the 1556 genes in the relaxed core genome, 592 OGs were statistically significant in rejecting the reference tree topology for both SH and ELW tests, after applying the FDR correction. Hence, we
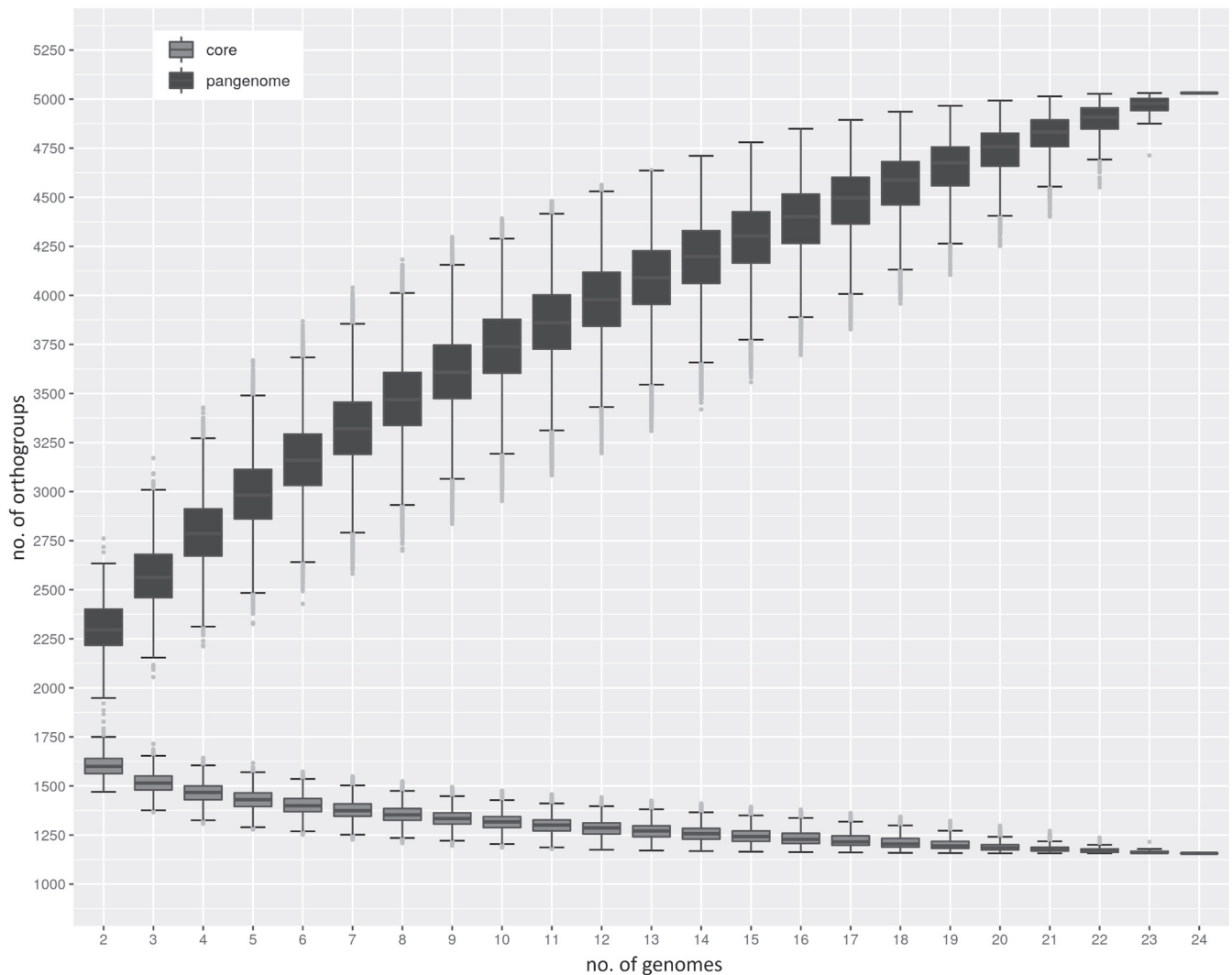
Fig. 3. Rarefaction curve for the pangenome (purple) and the genes in the strict core (green) given a number of genomes of *L. garvieae* used. The figure was created using ggplot2 v.3.2.0 [75].

identified them as putative recombinant genes, representing 38% of the relaxed core (Table S11). Most of these putative recombinant OGs were singletons, but there were recombination events that included a variable number of genes. So, 57 recombination events included two genes, 13 included three genes, 12 included four genes, three included five genes, and a single event included 10 genes (Table S12). These results revealed that *L. garvieae* is a highly recombinogenic species although there are differences among the four clusters, the cluster that includes Lg-Granada being the most promiscuous one (Fig. S4). If the average number of CDS in the genome of the species is 1975 – based on the 24 strains used in this work – this means that almost 30% of a typical genome might be involved in within-species horizontal transfers. A similar study focused on recombination in *Streptococcus*, a genus genetically related to *Lactococcus*, revealed that 35% of the *Streptococcus pyogenes* genome is recombinant, a value similar to that obtained for *L. garvieae* [70].

Compared to the gene content on the chromosome of Lg-Granada, intraspecific recombinant genes mostly encode membrane and cytoplasmic proteins, with catalytic and binding activity, which are involved in metabolic and transport processes. There is also a large proportion of recombinant genes involved in response to external stimuli, cell division and pathogenesis (Fig. S5).

Recombination in the genus *Lactococcus*

The Lg-Granada strain was compared with six genomes of three *Lactococcus* species, *L. lactis*, *L. raffinolactis* and *L. piscium*, which are frequently isolated from fish or dairy products (Table S2). The phylogenetic reconstruction shows that *L. garvieae* shares an ancestor with *L. lactis*, *L. piscium* and *L. raffinolactis* being more distant (Fig. 5a).
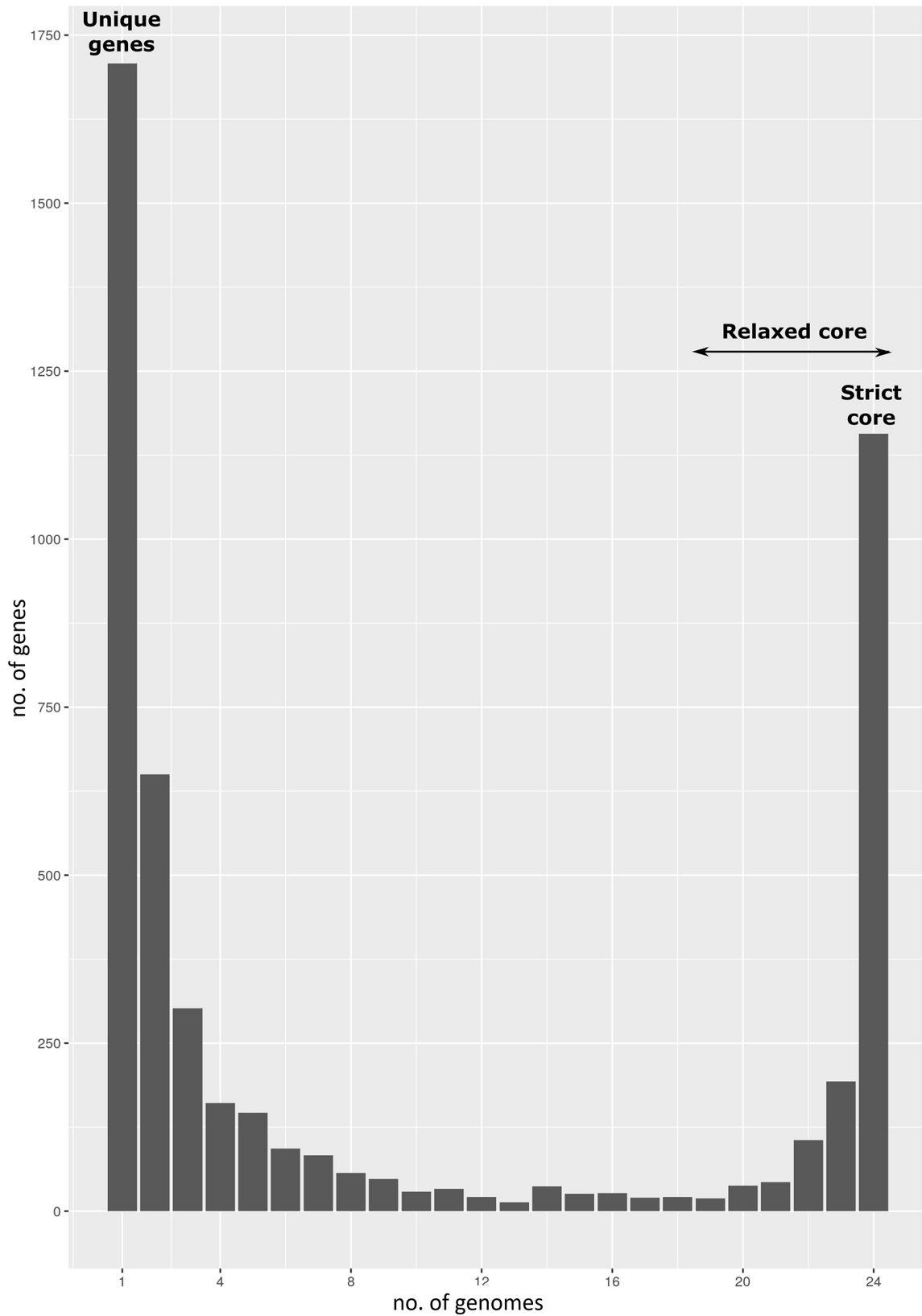
Fig. 4. Distribution of gene frequencies in the 24 genomes of *L. garvieae* included in this analysis. Unique genes and both strict and relaxed cores are marked. The figure was created using ggplot2 v.3.2.0 [75].

(a)

(b)
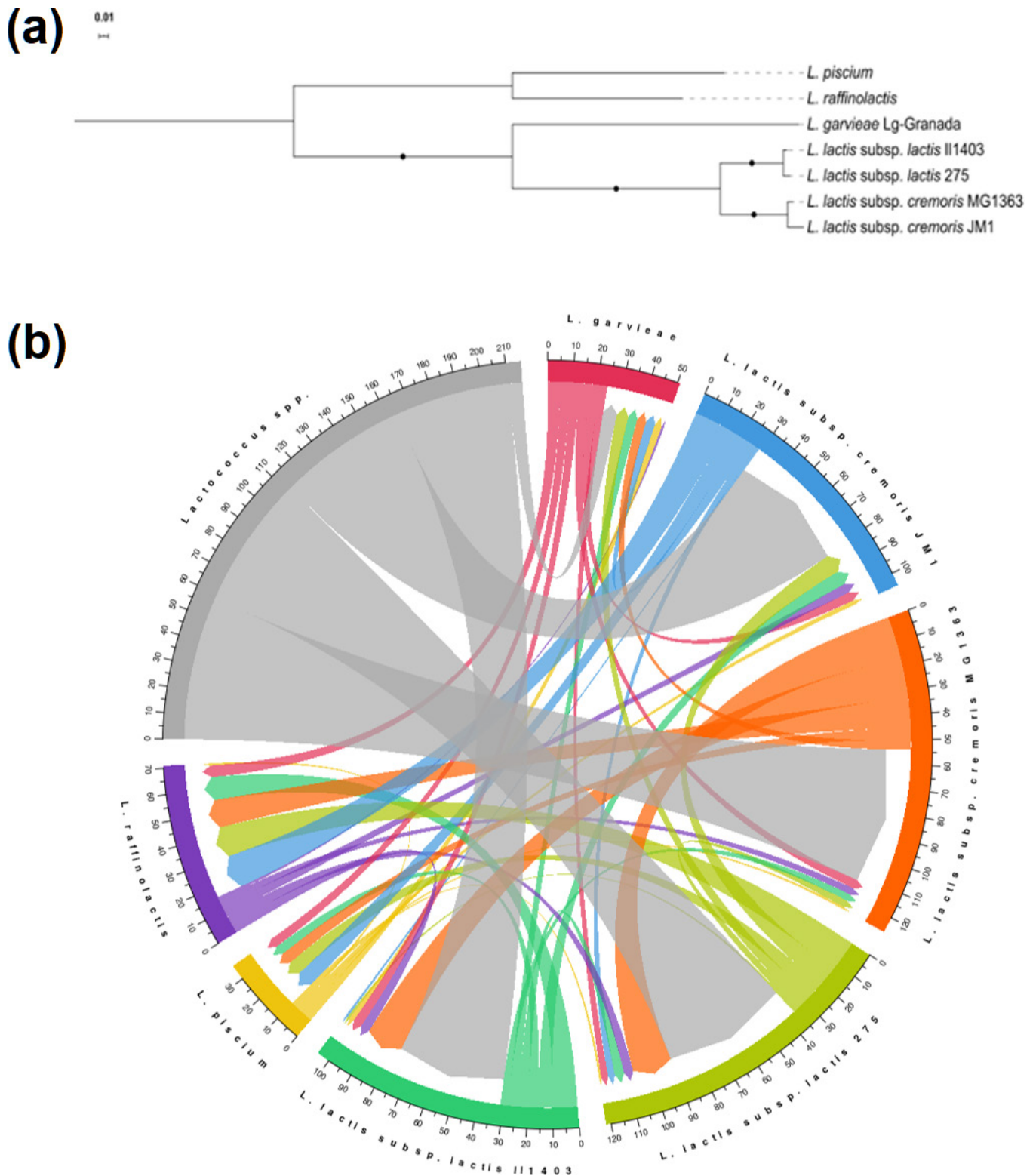
Fig. 5. (a) Phylogenetic tree of *Lactococcus* strains used in this study. Black dots within branches indicate a bootstrap support value ≥90%. (b) Summary of gene movements between the species of *Lactococcus* used in this study. The phylogenetic tree was depicted using iTol v.4 [74], and the chord diagram was created using circlize v.0.4.7 [76].
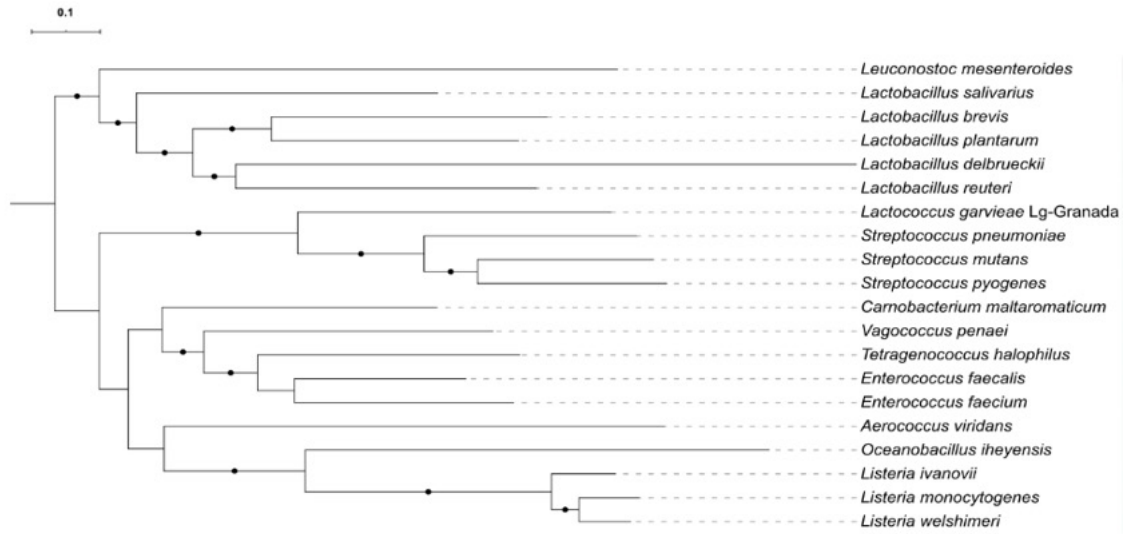
Fig. 6. (a) Phylogenetic tree of *Bacilli* species used in this study. Black dots within branches indicate a bootstrap support value ≥90%. (b) Summary of gene movements between the different genera of the class *Bacilli* used in this study. The phylogenetic tree was depicted using iToL v.4 [74], and the chord diagram was created using circlize v.0.4.7 [76].

The strict core genome of the genus, based on these seven genomes, encompassed 924 genes, spanning 927 996 bp with 468 861 variant positions. This value was consistent with the 949-genes core calculated by Ferrario *et al.* [64]. The relaxed core (five to seven genomes) encompassed 1462 genes, spanning 1 434 816 bp with 723 149 variant positions. In total, 103 OGs were eligible for topological congruence testing, 97 of which were recombinant (Table S13), but only 27 of them implied movement between Lg-Granada and the other species. This result revealed low recombination between Lg-Granada and the other species of the genus (1.8% of the relaxed core). Two recombination events including two genes and another one including four genes were detected (Table S14). The other 19 recombinant OGs were singletons.

Lg-Granada showed a balance in the number of genes donated to and received from other species in the genus (Fig. 5b). This does not occur in the other species of *Lactococcus*, which receive more than they donate. The largest flow of gene movements occurs among the *L. lactis* subspecies, being the main donors among them and to the other species of the genus (Fig. 5b). The gene transfer events from *L. lactis* to *L. garvieae* have been also demonstrated in the genes harboured in the plasmid pLG50 and prophages, as noted above.

Again, when compared with the gene proportion in the Lg-Granada strain chromosome, interspecific recombinant genes at the genus level mostly encode membrane and macromolecular complex proteins – cytoplasmic proteins were also an important fraction, although they were not proportionally higher than the content of the Lg-Granada chromosome. The most abundant categories of recombinant genes included genes involved in transport, transcription regulation and cell division. The proportion of genes involved in homeostatic processes was also remarkable (Fig. S5).

Recombination between *L. garvieae* and other species of the class *Bacilli*

The genome of strain Lg-Granada, as representative of *L. garvieae*, was compared with 19 other genomes from the class *Bacilli*, grouped in eight genera of the order *Lactobacillales* – *Aerococcus, Carnobacterium, Enterococcus, Tetragenococcus, Vagococcus, Lactobacillus, Leuconostoc* and *Streptococcus* – and two genera of the *Bacillales* order – *Oceanobacillus* and *Listeria* (Table S3). The phylogenetic tree shows a close relationship between *L. garvieae* and *Streptococcus*, sharing a common ancestor with other species, some of clinical relevance, such as *Enterococcus faecalis, Enterococcus faecium* or *Listeria monocytogenes* (Fig. 6a).

The strict core genome of the class *Bacilli*, based on our 20 genomes, has 409 genes, spanning 466 956 bp with 344 201 variant positions. The relaxed core (16–20 genomes) included 775 genes, spanning 881 175 bp with 661 535 variant positions. A total of 144 OGs were eligible for topological congruence testing, 135 of which were recombinant (Table S15), but only 34 of them implied movement between Lg-Granada and the other species. All the recombinant OGs were singletons.

Lg-Granada is less a donor than a recipient of genes in relation to other genera of the class *Bacilli* (Fig. 6b). The main recipient from Lg-Granada genes was *Lactobacillus*. The main donor of genes to Lg-Granada was *Streptococcus*, followed by *Aerococcus* and *Enterococcus*. Genera not included in the analysis – marked as external, but probably belonging to the class *Bacilli* – were the main donors to *Streptococcus, Vagococcus, Enterococcus, Lactobacillus* and *Listeria*.

In this case, when compared to the Lg-Granada chromosome, the recombinant genes have a higher proportion of genes that encode cytoplasm, membrane and macromolecular complex proteins, but also ribosomal, chromosomal and extracellular proteins. Most of these proteins have catalytic and binding activity, transporters, structural activity, or translation regulation and they are involved in metabolic processes, transport, translation, response to stimulus, cellular organization, cell division, protein folding, homeostatic processes and growth (Fig. S5).

In general, at the interspecies level (both genus and class) *L. garvieae* tends to import more genes than to export them (see Tables S16 and S17 for further details). Many of the genes detected as recombinant encode membrane proteins, many of them with transporter function or surface structure. This has also been observed in other studies, both in Gram-positive and in Gram-negative bacteria [70, 71]. Many of these proteins will be on the cell surface, so it is highly likely that they are involved in the interaction of *L. garvieae* Lg-Granada with the environment, either the natural environment or a host, and this medium exerts a selective pressure that favours both genetic exchange and fixation of these genes in the species which share the same ecological niche. In fact, the maintenance of large, non-transmissible plasmids, such as pLG50, despite their high cost to the bacteria, depends on the positive selection of genes with a high adaptive value: virulence genes, bacteriocins, immunity proteins, etc. [54, 72].

This work provides evidence that the genomic plasticity and evolution of *L. garvieae* is supported by the pervasive changes in its pangenome through HGT events from phylogenetically related and distant bacteria. The results broaden our understanding of the evolution of this emerging human pathogen. In addition, this work analyses in depth the genome of a new strain from a clinical source, contributing to increased knowledge of this species.

References

1. EFSA Panel on Biological Hazards (BIOHAZ), Koutsoumanis K, Allende A, Alvarez-Ordóñez A, Bolton D, *et al*. Update of the list of QPS-recommended biological agents intentionally added to food or feed as notified to EFSA 11: suitability of taxonomic units notified to EFSA until September 2019. *EFSA J* 2020;18:e05965.

2. Plumed-Ferrer C, Uusikylä K, Korhonen J, von Wright A. Characterization of *Lactococcus lactis* isolates from bovine mastitis. *Vet Microbiol* 2013;167:592–599.

3. Plumed-Ferrer C, Barberio A, Franklin-Guild R, Werner B, McDonough P, *et al*. Antimicrobial susceptibilities and random amplified polymorphic DNA-PCR fingerprint characterization of *Lactococcus lactis* ssp. *lactis* and *Lactococcus garvieae* isolated from bovine intramammary infections. *J Dairy Sci* 2015;98:6216–6225.

4. Chen F, Zhang Z, Chen J. Infective endocarditis caused by *Lactococcus lactis* subsp. *lactis* and *Pediococcus pentosaceus*: A case report and literature review. *Medicine (Baltimore)* 2018;97:e13658.

5. Shimizu A, Hase R, Suzuki D, Toguchi A, Otsuka Y, *et al*. *Lactococcus lactis* cholangitis and bacteremia identified by MALDI-TOF mass spectrometry: A case report and review of the literature on *Lactococcus lactis* infection. *J Infect Chemother* 2019;25:141–146.

6. Meyburgh CM, Bragg RR, Boucher CE. *Lactococcus garvieae*: an emerging bacterial pathogen of fish. *Dis Aquat Organ* 2017;123:67–79.

7. Lee JY, Hyun M, Kim HA, Ryu SY. Infectious spondylitis and septicemia due to *Lactococcus garvieae*: a literature review of non-endocarditis cases. *Infect Chemother* 2020;52.

8. Wang C-YC, Shie H-S, Chen S-C, Huang J-P, Hsieh I-C, *et al*. *Lactococcus garvieae* infections in humans: possible association with aquaculture outbreaks. *Int J Clin Pract* 2007;61:68–73.

9. Teixeira LM, Merquior VL, Vianni MC, Carvalho MG, Fracalanzza SE, *et al*. Phenotypic and genotypic characterization of atypical *Lactococcus garvieae* strains isolated from water buffalos with subclinical mastitis and confirmation of *L. garvieae* as a senior subjective synonym of *Enterococcus seriolicida*. *Int J Syst Bacteriol* 1996;46:664–668.

10. Tejedor JL, Vela AI, Gibello A, Casamayor A, Domínguez L, *et al*. A genetic comparison of pig, cow and trout isolates of *Lactococcus garvieae* by PFGE analysis. *Lett Appl Microbiol* 2011;53:614–619.

11. Kim JH, Go J, Cho CR, Kim JI, Lee MS, *et al*. First report of human acute acalculous cholecystitis caused by the fish pathogen *Lactococcus garvieae*. *J Clin Microbiol* 2013;51:712–714.

12. Gibello A, Galán-Sánchez F, Blanco MM, Rodríguez-Iglesias M, Domínguez L, *et al*. The zoonotic potential of *Lactococcus garvieae*: An overview on microbiology, epidemiology, virulence factors and relationship with its presence in foods. *Res Vet Sci* 2016;109:59–70.

13. Rösch RM, Buschmann K, Brendel L, Schwanz T, Vahl C-F. *Lactococcus garvieae* endocarditis in a prosthetic aortic valve: a case report and literature review. *J Investig Med High Impact Case Rep* 2019;7:2324709619832052.

14. Malek A, De la Hoz A, Gomez-Villegas SI, Nowbakht C, Arias CA. *Lactococcus garvieae*, an unusual pathogen in infective endocarditis: case report and review of the literature. *BMC Infect Dis* 2019;19:301.

15. Fortina MG, Ricci G, Foschino R, Picozzi C, Dolci P, *et al*. Phenotypic typing, technological properties and safety aspects of *Lactococcus garvieae* strains from dairy environments. *J Appl Microbiol* 2007;103:445–453.

16. El-Baradei G, Delacroix-Buchet A, Ogier JC. Bacterial biodiversity of traditional Zabady fermented milk. *Int J Food Microbiol* 2008;121:295–301.

17. Alegría A, Alvarez-Martín P, Sacristán N, Fernández E, Delgado S, *et al*. Diversity and evolution of the microbial populations during manufacture and ripening of Casín, a traditional Spanish, starter-free cheese made from cow's milk. *Int J Food Microbiol* 2009;136:44–51.

18. Flórez AB, Mayo B. The plasmid complement of the cheese isolate *Lactococcus garvieae* IPLA 31405 revealed adaptation to the dairy environment. *PLoS One* 2015;10:e0126101.

19. Heras Cañas V, Pérez Ramirez MD, Bermudez Jiménez F, Rojo Martin MD, Miranda Casas C, *et al*. *Lactococcus garvieae* endocarditis in a native valve identified by MALDI-TOF MS and PCR-based 16s rRNA in Spain: A case report. *New Microbes New Infect* 2015;5:13–15.

20. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, *et al*. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 2013;10:563–569.

21. Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, *et al*. A whole-genome assembly of Drosophila. *Science* 2000;287:2196–2204.

22. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, *et al*. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 2015;16:294.

23. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.

24. Zhou CE, Smith J, Lam M, Zemla A, Dyer MD, *et al*. MvirDB--a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. *Nucleic Acids Res* 2007;35:D391-4.

25. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, *et al*. The comprehensive antibiotic resistance database. *Antimicrob Agents Chemother* 2013;57:3348–3357.

26. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. UniProtKB/Swiss-Prot. *Methods Mol Biol* 2007;406:89–112.

27. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, *et al*. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 2016;44:W16-21.

28. Akhter S, Aziz RK, Edwards RA. PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res* 2012;40:e126.

29. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, *et al*. BLAST+: architecture and applications. *BMC Bioinformatics* 2009;10:421.

30. Rombel IT, Sykes KF, Rayner S, Johnston SA. ORF-FINDER: a vector for high-throughput gene identification. *Gene* 2002;282:33–41.

31. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, *et al*. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 2008;9:75.

32. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* 2006;34:D32-6.

33. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods* 2016;8:12–24.

34. Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, *et al*. Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics* 2011;12:124.

35. Ranwez V, Harispe S, Delsuc F, Douzery EJP. MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS One* 2011;6:e22594.

36. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* 2013;30:1188–1195.

37. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;32:268–274.

38. Pla-Díaz M, Sánchez-Busó L, Giacani L, Šmajs D, Bosshard PP, *et al*. Evolutionary processes in the emergence and recent spread of the syphilis agent, *Treponema pallidum*. *Mol Biol Evol* 2021:msab318.

39. Strimmer K, von Haeseler A. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc Natl Acad Sci U S A* 1997;94:6815–6819.

40. Münkemüller T, Lavergne S, Bzeznik B, Dray S, Jombart T, *et al*. How to measure and test phylogenetic signal. *Methods in Ecology and Evolution* 2012;3:743–756.

41. Shimodaira H, Hasegawa M. Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. *Molecular Biology and Evolution* 1999;16:1114–1116.

42. Strimmer K, Rambaut A. Inferring confidence sets of possibly misspecified gene trees. *Proc Biol Sci* 2002;269:137–142.

43. R Core Team. R: a language and environment for statistical computing. 2019.

44. Robinson O, Dylus D, Dessimoz C. Phylo.io: interactive viewing and comparison of large phylogenetic trees on the web. *Mol Biol Evol* 2016;33:2163–2166.

45. Bannam TL, Teng WL, Bulach D, Lyras D, Rood JI. Functional identification of conjugation and replication regions of the tetracycline resistance plasmid pCW3 from *Clostridium perfringens*. *J Bacteriol* 2006;188:4942–4951.

46. Mazmanian SK, Liu G, Jensen ER, Lenoy E, Schneewind O. *Staphylococcus aureus* sortase mutants defective in the display of surface proteins and in the pathogenesis of animal infections. *Proceedings of the National Academy of Sciences* 2000;97:5510–5515.

47. Löfling J, Vimberg V, Battig P, Henriques-Normark B. Cellular interactions by LPxTG-anchored pneumococcal adhesins and their streptococcal homologues. *Cellular Microbiology* 2011;13:186–197.

48. Morita H, Toh H, Oshima K, Yoshizaki M, Kawanishi M, *et al*. Complete genome sequence and comparative analysis of the fish pathogen *Lactococcus garvieae*. *PLoS ONE* 2011;6:e23184.

49. Menéndez A, Fernández L, Reimundo P, Guijarro JA. Genes required for *Lactococcus garvieae* survival in a fish host. *Microbiology* 2007;153:3286–3294.

50. Reimundo P, Menéndez A, Méndez J, Pérez-Pascual D, Navais R, *et al*. dltA gene mutation in the teichoic acids alanylation system of *Lactococcus garvieae* results in diminished proliferation in its natural host. *Veterinary Microbiology* 2010;143:434–439.

51. Denapaite D, Brückner R, Hakenbeck R, Vollmer W. Biosynthesis of teichoic acids in *Streptococcus pneumoniae* and closely related species: lessons from genomes. *Microb Drug Resist* 2012;18:344–358.

52. Boucher I, Emond E, Parrot M, Moineau S. DNA sequence analysis of three *Lactococcus lactis* plasmids encoding phage resistance mechanisms. *J Dairy Sci* 2001;84:1610–1620.

53. Mills S, McAuliffe OE, Coffey A, Fitzgerald GF, Ross RP. Plasmids of lactococci - genetic accessories or genetic necessities? *FEMS Microbiol Rev* 2006;30:243–273.

54. Aguado-Urda M, Gibello A, Blanco MM, López-Campos GH, Cutuli MT, *et al*. Characterization of plasmids in a human clinical strain of *Lactococcus garvieae*. *PLoS One* 2012;7:e40119.

55. Fallico V, McAuliffe O, Fitzgerald GF, Ross RP. Plasmids of raw milk cheese isolate *Lactococcus lactis* subsp. *lactis* biovar diacetylactis DPC3901 suggest a plant-based origin for the strain. *Appl Environ Microbiol* 2011;77:6451–6462.

56. Laarman AJ, Mijnheer G, Mootz JM, van Rooijen WJM, Ruyken M, *et al*. *Staphylococcus aureus* Staphopain A inhibits CXCR2-dependent neutrophil activation and chemotaxis. *EMBO J* 2012;31:3607–3619.

57. Liu J, Ma'ayeh S, Peirasmaki D, Lundström-Stadelmann B, Hellman L, *et al*. Secreted Giardia intestinalis cysteine proteases disrupt intestinal epithelial cell junctional complexes and degrade chemokines. *Virulence* 2018;9:879–894.

58. Tosukhowong A, Zendo T, Visessanguan W, Roytrakul S, Pumpuang L, *et al*. Garvieacin Q, a novel class II bacteriocin from *Lactococcus garvieae* BCC 43578. *Appl Environ Microbiol* 2012;78:1619–1623.

59. Hoai TD, Nishiki I, Yoshida T. Properties and genomic analysis of *Lactococcus garvieae* lysogenic bacteriophage PLgT-1, a new member of Siphoviridae, with homology to *Lactococcus lactis* phages. *Virus Res* 2016;222:13–23.

60. Eraclio G, Fortina MG, Labrie SJ, Tremblay DM, Moineau S. Characterization of prophages of *Lactococcus garvieae*. *Sci Rep* 2017;7:1856.

61. Eraclio G, Ricci G, Fortina MG. Insertion sequence elements in *Lactococcus garvieae*. *Gene* 2015;555:291–296.

62. Siguier P, Gourbeyre E, Chandler M. Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol Rev* 2014;38:865–891.

63. Bobay L-M, Ochman H. The evolution of bacterial genome architecture. *Front Genet* 2017;8:72.

64. Ferrario C, Ricci G, Milani C, Lugli GA, Ventura M, *et al*. *Lactococcus garvieae*: where is it from? A first approach to explore the evolutionary history of this emerging pathogen. *PLoS One* 2013;8:e84796.

65. Reguera-Brito M, Galán-Sánchez F, Blanco MM, Rodríguez-Iglesias M, Domínguez L, *et al*. Genetic analysis of human clinical isolates of *Lactococcus garvieae*: Relatedness with isolates from foods. *Infect Genet Evol* 2016;37:185–191.

66. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* 2009;106:19126–19131.

67. Shahi N, Mallik SK. Emerging bacterial fish pathogen *Lactococcus garvieae* RTCLI04, isolated from rainbow trout (*Oncorhynchus mykiss*): Genomic features and comparative genomics. *Microb Pathog* 2020;147:104368.

68. Gyles C, Boerlin P. Horizontally transferred genetic elements and their role in pathogenesis of bacterial disease. *Vet Pathol* 2014;51:328–340.

69. Awadalla P. The evolutionary genomics of pathogen recombination. *Nat Rev Genet* 2003;4:50–60.

70. Lefébure T, Stanhope MJ. Evolution of the core and pan-genome of Streptococcus: positive selection, recombination, and genome composition. *Genome Biol* 2007;8:R71.

71. Petersen L, Bollback JP, Dimmic M, Hubisz M, Nielsen R. Genes under positive selection in *Escherichia coli*. *Genome Res* 2007;17:1336–1343.

72. Nishiki I, Oinaka D, Iwasaki Y, Yasuike M, Nakamura Y, *et al*. Complete genome sequence of nonagglutinating *Lactococcus garvieae* strain 122061 isolated from yellowtail in Japan. *Genome Announc* 2016;4:00592–16.

73. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 2009;25:119–120.

74. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 2019;47:W256–W259.

75. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag, 2016. https://ggplot2.tidyverse.org/

76. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular visualization in R. *Bioinformatics* 2014;30:2811–2812.