Comparative Genome Analysis of *Lactobacillus rhamnosus* Clinical Isolates from Initial Stages of Dental Pulp Infection: Identification of a New Exopolysaccharide Cluster

Mangala A. Nadkarni1,2*, Zhiliang Chen3, Marc R. Wilkins3, Neil Hunter1,2

1 Institute of Dental Research, Westmead Centre for Oral Health and Westmead Millennium Institute, Westmead, New South Wales, Australia, 2 Faculty of Dentistry, The University of Sydney, Sydney, New South Wales, Australia, 3 Systems Biology Initiative, School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, New South Wales, Australia

**Abstract**

The human oral microbiome has a major role in oral diseases including dental caries. Our studies on progression of caries infection through dentin and more recently, the invasion of vital dental pulp, detected *Lactobacillus rhamnosus* in the initial stages of infection of vital pulp tissue. In this study employing current high-throughput next generation sequencing technology we sought to obtain insight into genomic traits of tissue invasive *L. rhamnosus*, to recognise biomarkers that could provide an understanding of pathogenic potential of lactobacilli, generally regarded as safe. Roche GS FLX+ technology was used to generate whole genome sequences of two clinical isolates of *L. rhamnosus* infecting vital pulp. Detailed genome-wide comparison of the genetic profiles of tissue invasive *L. rhamnosus* with probiotic *L. rhamnosus* was performed to test the hypothesis that specific strains of *L. rhamnosus* possessing a unique gene complement are selected for the capacity to invade vital pulp tissue. Analysis identified 264 and 258 genes respectively, from dental pulp-invasive *L. rhamnosus* strains LRHMDP2 and LRHMDP3 isolated from two different subjects that were not present in the reference probiotic *L. rhamnosus* strain ATCC 53103 (GG). Distinct genome signatures identified included the presence of a modified exopolysaccharide cluster, a characteristic confirmed in a further six clinical isolates. Additional features of LRHMDP2 and LRHMDP3 were altered transcriptional regulators from RpoN, NtrC, ArsR and zinc-binding Cro/CI families, as well as changes in the two-component sensor kinase response regulator and ABC transporters for ferric iron. Both clinical isolates of *L. rhamnosus* contained a single SpaFED cluster, as in *L. rhamnosus* Lc705, instead of the two Spa clusters (SpaCBA and SpaFED) identified in *L. rhamnosus* ATCC 53103 (GG). Genomic distance analysis and SNP divergence confirmed a close relationship of the clinical isolates but segregation from the reference probiotic *L. rhamnosus* strain ATCC 53103 (GG).

**Citation:** Nadkarni MA, Chen Z, Wilkins MR, Hunter N (2014) Comparative Genome Analysis of *Lactobacillus rhamnosus* Clinical Isolates from Initial Stages of Dental Pulp Infection: Identification of a New Exopolysaccharide Cluster. PLoS ONE 9(3): e90643. doi:10.1371/journal.pone.0090643

**Editor:** Christiane Forestier, Université d’Auvergne Clermont 1, France

**Received:** May 30, 2013; **Accepted:** February 3, 2014; **Published:** March 14, 2014

**Copyright:** © 2014 Nadkarni et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This study was supported partially by NIDCR grant R01 DE015272-07, by the Westmead Centre for Oral Health, by the Australian Government Super Science Scheme, the New South Wales State Government Science Leveraging Fund (SLF) and by the University of New South Wales. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: mangala.nadkarni@sydney.edu.au

**Introduction**

The human oral microbiome has a major causal association with dental caries. Organic acids produced by microbial hydrolysis of dietary sugars cause demineralization of hydroxyapatite crystals leading to dental caries. The Human Microe Identification Microarray (HOMIM) for identification of named and unnamed taxa associated with oral infections [1], in combination with pyrosequencing, has enabled the complex uncultured microbial ecosystem of the oral cavity to be more precisely defined [2–5]. Other techniques such as scanning electron microscopy (SEM), fluorescence in-situ hybridization (FISH) and confocal microscopy using fluorochrome-labelled 16S rRNA probes, have given insights into the community structure and spatial distribution of polymicrobial consortia in infected tissues [6,7]. More recently, these analyses have provided evidence indicating that a restricted group of lactobacilli, particularly *Lactobacillus rhamnosus*, are implicated in the initial stages of infection of vital pulp tissue [6].

While many current techniques have allowed partial profiling of the human microbiome, important questions regarding functional attributes, metabolic specialization and virulence traits can only be answered by comprehensive post-genomic analysis of the ecosystems of the human body [8]. Current high-throughput Next Generation Sequencing has the power to provide detailed knowledge of the diverse array of genes that influence biochemical and physiological functions responsible for host-microbe associations. This enables the identification of bacterial traits that contribute to adhesion, invasion and evasion of host defences. In the present study, whole genome sequencing of two clinical isolates of *L. rhamnosus* using Roche GS FLX+ technology combined with detailed genome-wide analysis of the genetic profiles of tissue invasive *L. rhamnosus* compared with probiotic *L. rhamnosus*, was undertaken. The hypothesis to be tested was that invasive *L.
**Results and Discussion**

**Genome sequencing**

*L. rhamnosus* clinical isolates (LRHMDP2 and LRHMDP3) were sourced from infected dental pulps from carious teeth categorised as representing the initial stages of infection of vital pulp tissue from two different subjects [6] (see Methods section for detail description for selection criteria; Tables S1, S2a, S2b). These isolates were speciated by PCR analysis of 16S rRNA gene sequences using *Lactobacillus* genus-specific primers as described in an earlier report [10]. LRHMDP2 and LRHMDP3 16S rRNA gene sequences were 99% homologous to *L. rhamnosus*. Genomic DNA from both the clinical isolates was sequenced by a whole genome shotgun strategy using Roche GS FLX+ pyrosequencing at the Ramaciotti Centre for Gene Function Analysis, University of New South Wales (see Methods section for detailed sequencing protocol).

**Genome properties**

A total of 77,329 reads from *L. rhamnosus* LRHMDP2 and 75,985 reads from *L. rhamnosus* LRHMDP3 were generated to reach a depth of ~17-fold genome coverage, and assembled by Newbler Assembler 2.6 into 51 and 47 contigs longer than 200 base pairs, respectively (Table 1, [9]). Gene definition and annotation was performed by merging the result from the RAST (Rapid Annotation using Subsystem Technology) server and tRNAscan-SE. This showed that sequences of the two strains, *L. rhamnosus* LRHMDP2 and *L. rhamnosus* LRHMDP3, comprised 2,910 coding sequences (CDSs), 7 rRNA loci and 56 tRNAs. The genomic sequence of *L. rhamnosus* LRHMDP2 was 96.6% similar to *L. rhamnosus* GG and *L. rhamnosus* Lc705 (Table 1, Figure 1 [9]). By pair-wise genome comparisons Lukjancenko et al. have reported 93.3% similarity between two *L. rhamnosus* strains [11].

---

**Table 1. General Genome sequencing features and annotation of *L. rhamnosus* clinical isolates and probiotic reference strains.**

| Organism | Total Reads | Genome coverage | Number of contigs | Number of coding sequences (CDS) | rRNA operons | tRNA genes | GC content (GC%) | Hypothetical proteins (total) |
|----------|-------------|-----------------|------------------|----------------------------------|--------------|------------|------------------|-------------------------------|
| LRHMDP2  | 77,329      | 17 X 51         | 2,910            | 2,910                            | 7            | 56         | 46.6             | 815                           |
| LRHMDP3  | 75,985      | 17 X 47         | 2,925            | 2,925                            | 6            | 57         | 46.6             | 819                           |
| GG        | -           | -               | -                | 3,014                            | 5            | 57         | 47               | 734                           |
| Lc705     | -           | -               | -                | 3,031                            | 5            | 47         | 47               | 760                           |

*a* Using Roche GS(FLX+). Pyrosequencing based on results merged from the RAST(Rapid Annotation using Subsystem Technology) server and tRNAscan-SE.

*b* Using Newbler Assembler 2.6. These isolates were sequenced by a whole genome shotgun strategy using Roche GS FLX+ pyrosequencing at the Ramaciotti Centre for Gene Function Analysis, University of New South Wales (see Methods section for detailed sequencing protocol).
Comparative Genome Analysis of Tissue Invasive Lactobacillus rhamnosus

Accession: NC_013198
Length: 3,010,111 bp

A
Lactobacillus rhamnosus GG complete genome

Accession: NC_013198
Length: 3,010,111 bp

B
Lactobacillus rhamnosus GG complete genome
Comparative Genome Analysis of Tissue Invasive Lactobacillus rhamnosus

Genome annotation using the RAST server assigns functions to genes, enables assembly of metabolic pathways and makes connections into sub-systems represented in the genome to achieve comparative analysis with annotated genomes maintained in the SEED (http://www.theseed.org/) [12]. Biological roles could be assigned to 72.0% of CDSs for the genomic sequences of L. rhamnosus LRHMDP2 and L. rhamnosus LRHMDP3. A comparative annotation summary indicated 264 of the 2,910 coding sequences from L. rhamnosus strain LRHMDP2 and 258 of the 2,925 coding sequences from L. rhamnosus strain LRHMDP3 (~9% of coding sequences) in the clinical isolates were newly identified genes (Table 1, Table 2) where homologues could not be detected in the probiotic reference strain of L. rhamnosus ATCC 55103 (GG). New genes are defined as no BLAST hits found or the normalised BLAST score less than 2 where normalised BLAST score is calculated as BLAST score divided by alignment length. Some 9% of coding sequences attributed to genes from the probiotic L. rhamnosus strain GG as well as the other probiotic L. rhamnosus strain Lc705 were not detected in either clinical isolate (Table 3).

Comparisons based on sub-system category distribution highlighted distinctive features that discriminated the two clinical isolates of L. rhamnosus, (LRHMDP2 and LRHMDP3) and L. rhamnosus GG. Comparative genomic analysis between L. rhamnosus GG and the two strains, L. rhamnosus LRHMDP2 and L. rhamnosus LRHMDP3, based on 27 sub-system category distributions, showed differences in 6 distribution categories, namely; co-factors, vitamins, prosthetic groups and pigments; cell wall and capsule; virulence, disease and defence traits; phage, prophage, transposable elements, plasmids, iron acquisition and metabolism and carbohydrates (data not shown).

Inter-strain divergence between L. rhamnosus strains LRHMDP2 and LRHMDP3

Although the probiotic L. rhamnosus strains GG and Lc705 were shown to exhibit 331 and 383 strain-specific genes respectively [13], the inter-strain difference between clinical isolates L. rhamnosus LRHMDP2 and L. rhamnosus LRHMDP3 was limited. Analysis revealed 26 strain-specific genes for L. rhamnosus LRHMDP2 and 28 strain-specific genes for L. rhamnosus LRHMDP3 (Table S6a, S6b). Many of the strain-specific genes encoded hypothetical proteins. CRISPRs and CRISPR-associated (cas) genes, in addition to genes coding for hypothetical proteins (Table S3). Similarly, genes present in the other L. rhamnosus probiotic strain, L. rhamnosus Lc705 that were absent in both clinical isolate assemblies were mainly assigned to transposases (IS111A, IS1328, IS1533, IS116, IS110, IS902, IS4, IS1480, IS30, IS5, tsp4), phage-related products, expolysaccharide (eps) clusters; also fructose-specific, sorbose-specific, galactitol-specific, lactose/cellobiose-specific and mannose-specific PTS systems. Also noted as absent from the clinical isolates were other carbohydrate metabolism genes and the type I restriction modification system in addition to genes encoding hypothetical proteins (Table S4).

The clinical isolates, L. rhamnosus LRHMDP2 and L. rhamnosus LRHMDP3 were also distinct from the other L. rhamnosus probiotic strain Lc705 as no plasmid genes could be identified in either as compared to L. rhamnosus Lc705 that harbours a 64.5 kb plasmid [13]. However, the two clinical isolates from infected dental pulp resembled L. rhamnosus Lc705 and other lactobacilli such as L. gasseri, L. johnsonii and L. iners in the absence of CRISPRs and cas detected in L. rhamnosus GG [13,14]. CRISPRs and cas comprise a distinct defence system present in ~47% bacterial genomes (http://crispr.u-psud.fr/crispr/). This system provides resistance against foreign genetic elements and bacteriophage predation by inactivating incoming DNA. Similarly, both clinical isolates resembled L. rhamnosus Lc705 in the absence of the spaCB gene cluster and presence of a homologous spaFED cluster also present in L. rhamnosus GG. In Gram-positive bacteria the three subunits of the spaCB gene cluster are flanked by transposable elements and in L. rhamnosus strain GG this cluster is located on a genomic island GGISL2 [13]. Absence of the spaCB gene cluster in the clinical isolates could be related to the paucity in these strains of transposable elements that regulate horizontal gene transfer.

Both clinical isolates were characterised by a modified eps cluster, a two-component sensor kinase, a response regulator and ABC transporters for ferric iron, RNA polymerase sigma 54 factor KpoN, MutR, NtrC, ArsR and zine-binding Cio/Ci family transcriptional regulators, ThiJ/PpiI family protein and PTS system mannose, galactitol, mannitol, cellobiose and beta-glucoside-specific II components in addition to genes encoding hypothetical proteins (Table S3a, S3b). The majority of the coding sequences for newly identified genes had closest homology with genes in bacteria from the phylum Firmicutes.

Table 2. Number of new genes identified in the assemblies of clinical isolates, L. rhamnosus LRHMDP2 and L. rhamnosus LRHMDP3.

| Assembly Name | Number of genes |
|---------------|-----------------|
| L. rhamnosus LRHMDP2 | 264* |
| L. rhamnosus LRHMDP3 | 258* |

*Based on comparative studies with reference probiotic L. rhamnosus strains ATCC53103 (GG) and Lc705.
*Inclusive of 176 genes coding for hypothetical proteins.
*Inclusive of 166 genes coding for hypothetical proteins.

doi:10.1371/journal.pone.0090643.t002
protein PlnI, membrane-bound protease CAAX family and adhesion exoprotein were unique to _L. rhamnosus_ LRHMDP2. DinG family ATP-dependent helicase YoaA, a putative regulator of the mannose operon ManO, cytochrome d ubiquinol oxidase subunit I, phage lysin, flagellar hook-length control protein FlkK, phosphotransferase system, phosphocarrier protein HPr, phosphoenolpyruvate-protein phosphotransferase of PTS system and transcriptional regulator CtsR, were identified only in _L. rhamnosus_ LRHMDP3. Genome sequencing has identified CtsR (class III stress gene repressor) in other _L. rhamnosus_ strains. However, CtsR is well characterised in _Lactobacillus plantarum_ - found in vegetable, meat and dairy products and also as a natural inhabitant of the mammalian gut. CtsR plays an important role in the response to environmental stress conditions [15]. CtsR is also considered to play an important role in stress resistance and virulence in _Listeria monocytogenes_, a firmicute that causes food-borne illness and is associated with invasive gastrointestinal infection [16].

**Genomic distance and SNP divergence**

The genomic difference between the genomic sequences of two _L. rhamnosus_ clinical isolates and eleven other _L. rhamnosus_ strains from NCBI was analysed. Figure S1 shows a dendrogram that clusters the 15 different _L. rhamnosus_ strains based on full genome nucleotide-level comparisons. The branches of the dendrogram are scaled to show the distance between genomes. The two clinical isolates show small genomic difference, and are found in a different part of the dendrogram compared to the other probiotic _L. rhamnosus_ dairy isolates, intestinal isolates from healthy humans and isolates from feces of healthy humans.

A more specific SNP divergence analysis between the probiotic _L. rhamnosus_ strain GG and the two _L. rhamnosus_ clinical isolates identified 50,459 SNPs with 95% confidence (Table S7). Of these, 6,899 SNPs were found to be located in non-coding regions. Both clinical isolates exhibited 50,370 SNPs in common when compared with GG. Only 30 SNPs were common between _L. rhamnosus_ GG and _L. rhamnosus_ LRHMDP2 compared to _L. rhamnosus_ LMS2-1, which is inclusive of 77 genes coding for hypothetical proteins. Remaining genes of _L. rhamnosus_ GG and _L. rhamnosus_ ATCC 9595 have been characterised in detail [18,21].

The _eps_ cluster from the two clinical isolates, _L. rhamnosus_ LRHMDP2 and _L. rhamnosus_ LRHMDP3, is composed of 15 ORFs (12.44 kb) and 14 ORFs (11.55 kb) respectively, with notable absence of genes (rmlACBD) involved in dTDP-rhamnose biosynthesis (Figure 2, Table 4, Table 5). Genes involved in dTDP-rhamnose biosynthesis were identified in the 16.37 kb _eps_ gene cluster of _L. rhamnosus_ GG consisting of 17 putative ORFs [18] and the 18.7 kb _eps_ gene cluster of _L. rhamnosus_ ATCC 9595 consisting of 18 putative ORFs [21]. However, insertion of the putative transposase DDE domain from _L. casei_ ATCC 334 (62% identity) was found to inactivate _rmlA_ from the _eps_ cluster of _L. rhamnosus_ GG [18]. In both clinical isolates, a degenerate transposase belonging to the TraPh-IS66 superfamily and with 97% identity with the homolog from _L. rhamnosus_ LMS2-1 was identified at a similar location within the _eps_ cluster (Table 4, Table 5). The absence of _rmlACBD_ genes involved in the dTDP-rhamnose biosynthetic enzyme system in the _eps_ cluster from both clinical isolates implies an alternate route for EPS synthesis employed by the clinical isolates as compared to _L. rhamnosus_ GG and _L. rhamnosus_ ATCC 9595 [19]. The _eps_ cluster of both clinical isolates also comprises 3 glycosyltransferases as opposed to 5 glycosyltransferases in _L. rhamnosus_ GG [18] and in _L. rhamnosus_ ATCC 9595 [21]. The presence of family 2 glycosyltransferases showing 50% amino acid identity to a homolog from _Lactobacillus pentosus_ strain IG1 suggests a different pathway for polymerisation of monosaccharide repeating units. Family 2 glycosyltransferases are inverting glycosyltransferases, implying probable β-glycosidic linkage in the repeating unit of EPS (www.cazy.org/glycosyltransferases.html). A near-identical EPS biosynthetic gene cluster in _L. rhamnosus_ strains ATCC 9595 and _L. rhamnosus_ Lc705 produces rhamnose-rich EPS as opposed to the galactose-rich EPS from _L. rhamnosus_ GG. On a broader scale, however, the organisation of the _eps_ gene cluster from all the _L. rhamnosus_ strains studied including the two clinical isolates, remains in close agreement. For instance, the _wzb_ ORF as part of the _eps_ cluster from _L. rhamnosus_ strain GG and both clinical isolates are predicted to share 100% amino acid sequence identity. Remaining genes of the _eps_ cluster from both clinical isolates showed closest identity to two _eps_ cluster genes from _L. rhamnosus_ strain LMS2-1, _Lactobacillus zeae_ strain KCTC 3804 and _Lactobacillus pentosus_ strain IG1 and a single _eps_ cluster gene each from _L. rhamnosus_ strains GG, _Lc705_ and _L. rhamnosus_ LRHMDP3.

### Table 3. Number of genes from reference strain _L. rhamnosus_ ATCC 53103 (GG) and _L. rhamnosus_ Lc705 not detected in both assemblies.

| Reference probiotic _L. rhamnosus_ strains | Number of genes |
|-------------------------------------------|-----------------|
| _L. rhamnosus_ ATCC 53103 (GG)            | 275*            |
| _L. rhamnosus_ Lc705                      | 331*            |

* _L. rhamnosus_ LRHMDP2 and _L. rhamnosus_ LRHMDP3.

**Exopolysaccharide cluster**

Most of the lactic acid bacteria (LAB) produce exopolysaccharide (EPS) and the composition of exopolysaccharide has been shown to influence adhesion properties of probiotic strains [17,18]. Metabolic shunting away from glycolysis is considered to be an important driver of EPS formation [19]. The organization of an exopolysaccharide (eps) gene cluster from LAB includes genes encoding glycosyltransferases (key enzymes for biosynthesis of the EPS repeating unit located in the central region of the eps cluster), polysaccharide synthesis export protein and transcriptional regulator, in addition to chain length determinant protein (wzb ORF) at one end of the _eps_ gene cluster and protein-tyrosine phosphatase (wzd ORF) on the other end of the _eps_ gene cluster [17,19,20]. To date, the _eps_ cluster genes of only two probiotic _L. rhamnosus_ strains GG and ATCC 9595 have been characterized in detail [18,21].
and HN001. Biological function is unknown for four ORFs from the *L. rhamnosus* LRHMDP2 *eps* gene cluster and 3 ORFs from the *L. rhamnosus* LRHMDP3 *eps* gene cluster whereas in *L. rhamnosus* strains GG, Lc705 and ATCC 9595, biological function could be attributed to most of the predicted translated products.

The genomic organisation of the *eps* cluster genes from *L. rhamnosus* GG showed that, except for one ORF (wzr), all other ORFs of the *eps* cluster are in reverse orientation. Whereas 11 ORFs in *L. rhamnosus* LRHMDP2 and 10 ORFs in *L. rhamnosus* LRHMDP3 are in reverse orientation to the remaining ORFs of the *eps* cluster [Figure 2](#fig2). In both clinical isolates, in addition to wzr (cell envelope-associated transcriptional attenuator LytR-CpsA-Psr, subfamily F2), priming glycosyltransferase, *welE* (undecaprenyl-phosphate galactosephosphotransferase) also maintained a forward direction. These findings imply that in the clinical isolates, in addition to organisation, transcriptional direction of wzr and welE could regulate expression of *eps* cluster genes.

Minor variation in structure of the polysaccharide is found to influence adherence and biofilm formation and the nature of immune responses. A deletion mutant for the priming glycosyltransferase (*welE*) of *L. rhamnosus* GG showed altered heteropoly saccharide composition with increased adherence and biofilm capacity compared with the galactose-rich EPS from wild type *L. rhamnosus* GG [18]. A homolog for the priming glycosyltransferase *welE* from *L. rhamnosus* HN001 and a homolog for *welA* from *L. pentosus* IG1 were detected in the *eps* cluster from both clinical isolates. In *S. pneumoniae*, *welA* encodes the initial transferase transferring glucose-1-phosphate to undecaprenol phosphate, a similar function to *WbaP*, a galactosyl-1-phosphate transferase of the *Salmonella enterica* O-antigen gene cluster that catalyzes the first step of O-antigen synthesis [22,23].

Phosphotransferase system (PTS) transporters, ATP-binding cassette (ABC) transporters and sugar transporters

Genome-wide studies, using microarray-based expression analysis and also signature-tagged mutagenesis (STM) have shown that, in many bacteria, catabolism of complex carbohydrates could play a crucial role in the pathogenesis of disease [24]. Lactobacilli, including probiotic lactobacilli with a broad spectrum pattern of carbohydrate utilisation, were found to encode PTS transporters as a dominant mechanism for carbohydrate transport [13,25,26]. In *L. rhamnosus* LRHMDP2 and *L. rhamnosus* LRHMDP3, 83 and 86 PTS transporters respectively, were identified (equivalent to...
encoded multiple clusters for cellobiose uptake and hydrolysis as were complete (encoding for all subunits). Both clinical isolates of *Lactobacillus rhamnosus* isolates, were not homologous to the genes in the genome of *Lactococcus lactis* strain. Beta-glucoside-specific IIA, IIB, IIC identified in both clinical isolates included multidrug transporter, antimicrobial peptide transporter, bacteriocin/lantibiotic transporter, phosphate, nitrate/sulfonate/bicarbonate, alkaline sulfonates and ferric iron, zinc and manganese transporters, oligopeptide and amino acid, glutamine, proline, glycine, betain transporters, osmotically activated L-carnitine/cholin (OpuCA, CB, CC, CD) transporters, excinuclease, spermidine, putrescine transport, glycerol-3-phosphate (UgpCAE), ribose (RbsDACB), N-acetyl-D-glucosamine, maltose/maltodextrin transporters and multisugar transporters. Most of the ABC transporters were also detected in *L. rhamnosus* GG except for ferric iron, lantibiotic, alkane sulfonates, L-carnitine/cholin, ribose, N-acetyl-D-glucosamine and maltose/maltodextrin transporters. In both clinical isolates, ferric iron ABC transporters (iron-binding protein and permease protein) were associated with two-component sensor kinase (LRHMDP2_1746, LRHMDP3_482) and response regulator (LRHMDP3_478) were associated with two-component sensor system. Nearest homologs for ferric iron ABC transporters in *L. casei* and *L. paracasei* strains.

Sixty nine ATP-binding cassette (ABC) transporters identified in both clinical isolates as well as in probiotic *L. casei* [27]. Unlike *L. rhamnosus* GG, maltose/maltodextrin metabolism in the clinical isolates could be associated with the ABC transporter, a mechanism also adopted by *L. lactis* [26,28]. The PTS for mannose-specific IIA, IIB, IID, galactitol-specific IIA, IIC, mannitol-specific IIA, lactose-specific IIB, cellulose-specific IID, IIB and PTS system beta-glucoside-specific IIA, IIB, IIC identified in both clinical isolates, were not homologous to the genes in the genome of *L. rhamnosus* GG (Table 5a, 5b).

Table 4. Exopolysaccharide cluster genes from *L. rhamnosus* LRHMDP2.

| ORF | Size (bp) | Predicted encoded function | Predicted domain(s) present in encoded ORF | Best BLASTx hit (accession no.) | % Amino acid identity |
|-----|-----------|----------------------------|-------------------------------------------|-------------------------------|-----------------------|
| wzb | 768       | Manganese-dependent protein-tyrosine phosphatase | No hit in Pfam database | Wzb (ABV54212.1) | L. rhamnosus GG 100 |
| wzr | 894       | Cell envelope-associated transcriptional attenuator LysT-CpsA-Psr, subfamily F2 | 1 transmembrane domain and LysT_cpsA_psr domain (PF03816) | Wzr (YP_003174725.1) | L. rhamnosus LC705 95 |
| welE | 669       | Undecaprenyl-phosphate galactosephosphotransferase | Bacterial sugar transferase (PF02397) | WelE (AB571033.1) | L. rhamnosus HN001 85 |
| 198 | Degenerate transposase | TnpB super family (PF05717) | Transposase (ZP_04441884.1) | Lactobacillus rhamnosus LMS2-1 97 |
| 516 | polysaccharide biosynthesis export protein | Polysacc_synt (PF01943) and 6 transmembrane domains | Oligosaccharide translocase (YP_003352680.1) | Lactococcus lactis subsp. lactis KF147 43 |
| 1425 | Glycosyltransferase family 2 | Glycosyltransferase family 2 (PF05035) | WchA (CCC15460.1) | L. pentosus GI1 51 |
| 1086 | Glycosyltransferase family 2 group 2 | Glycosyltransferase family 2 (PF05035) | Polysaccharide pyruvyltransferase (PF04230) | Clostridium clariflavum DSM 19732 44 |
| 978 | hypothetical protein | low complexity | conserved hypothetical protein (ZP_04841887.1) | L. rhamnosus LMS2 95 |
| 210 | hypothetical protein | 2 transmembrane domains | Endoplasmic reticulum metalloproteinase 1 (CCCH00546.1) | Fibrellaestuarina BUZ 2 34 |
| 1101 | hypothetical protein | 9 transmembrane domains | hypothetical protein (ZP_06987467.1) | Bacteroides sp. 3_1_19 25 |
| 792 | Glycosyltransferase | No hit in Pfam database | WciB (ABQ58967.1) | Streptococcus oralis 53 |
| 1122 | UDP-galactopyranosemutase | UDP-galactopyranosemutase(PF03275) | UDP-galactopyranosume mutase (ZP_095047298.1) | L. zeae KCTC 3804 88 |
| 756 | Tyrosine-protein kinase EpsD | ATPase MipZ (PF09140)andCbiA (PF01656) | Wze (ZP_09453943.1) | L. zeae KCTC 3804 89 |
| 915 | Tyrosine-protein kinase transmembrane modulator EpsC | Chain length determinant protein (PF02706) and 1 transmembrane domain | Wzd (ACN94846.1) | L. rhamnosus GG 79 |

~3% of coding sequences (Table 6). Not all PTS transporters were complete (encoding for all subunits). Both clinical isolates encoded multiple clusters for cellobiose uptake and hydrolysis as observed in *L. casei* [27]. Unlike *L. rhamnosus* GG, maltose/maltodextrin metabolism in the clinical isolates could be associated with the ABC transporter, a mechanism also adopted by *L. casei* and *Lactococcus lactis* [26,28]. The PTS for mannose-specific IIA, IIB, IIC, IID, galactitol-specific IIA, IIC, mannitol-specific IIA, lactose-specific IIB, cellulose-specific IID, IIB and PTS system beta-glucoside-specific IIA, IIB, IIC identified in both clinical isolates, were not homologous to the genes in the genome of *L. rhamnosus* GG (Table 5a, 5b).

Sixty nine ATP-binding cassette (ABC) transporters identified in clinical isolates included multidrug transporter, antimicrobial peptide transporter, bacteriocin/lantibiotic transporter, phosphate, nitrate/sulfonate/bicarbonate, alkaline sulfonates and ferric iron, zinc and manganese transporters, oligopeptide and amino acid, glutamine, proline, glycine, betain transporters, osmotically activated L-carnitine/cholin (OpuCA, CB, CC, CD) transporters, excinuclease, spermidine, putrescine transport, glycerol-3-phosphate (UgpCAE), ribose (RbsDACB), N-acetyl-D-glucosamine, maltose/maltodextrin transporters and multisugar transporters. Most of the ABC transporters were also detected in *L. rhamnosus* GG except for ferric iron, lantibiotic, alkaline sulfonates, L-carnitine/cholin, ribose, N-acetyl-D-glucosamine and maltose/maltodextrin transporters. In both clinical isolates, ferric iron ABC transporters (iron-binding protein; LRHMDP2_1748, LRHMDP3_480 and permease protein; LRHMDP2_1750, LRHMDP3_478) were associated with two-component sensor kinase (LRHMDP2_1746, LRHMDP3_482) and response regulator (LRHMDP2_1747, LRHMDP3_481). Nearest homologs for ferric iron ABC transporters (iron-binding protein and permease protein), two-component sensor kinase and response regulator could be identified in *L. casei* and *L. paracasei* strains.

Genome-wide analysis has shown that virulence attributed to ABC transporters is associated with uptake of nutrients, metal ions such as iron, zinc, and manganese and cell attachment in a given physiological niche [29,30]. However, ABC transport systems, particularly important in bacterial virulence in disease models, were detected in clinical isolates as well as probiotic *L. rhamnosus* GG [29,30]. ABC transporters for antimicrobial peptides were detected in both clinical isolates as well as in probiotic *L. rhamnosus*.
### Table 5. Exopolysaccharide cluster genes from *L. rhamnosus* LRHMDP3.

| ORF | Size (bp) | Predicted encoded function | Predicted domain(s) present in encoded ORF | Best BLASTx hit (accession no.) | % Amino acid identity |
|-----|-----------|----------------------------|-------------------------------------------|---------------------------------|-----------------------|
| wzb | 768       | Manganese-dependent protein-tyrosine phosphatase | No hit in Pfam database | Wzb (ABV54212.1) | 100 |
| wzr | 894       | Cell envelope-associated transcriptional attenuator LytR-CpsA-Par, subfamily F2 | 1 transmembrane domain and LytR_cpsA_psr domain (PF03816) | Wzr (YP_003174725.1) | 95 |
| weE | 669       | Undecaprenyl-phosphate galactosephosphotransferase | Bacterial sugar transferase (PF02397) | WeE (AB571033.1) | 85 |
| WelE | 669       | hypothetical protein | low complexity | conserved hypothetical protein (ZP_04441887.1) | 95 |
| 198 |          | Degenerate transposase | TnpB_s556 super family (PF05717) | Transposase (ZP_04441884.1) | L. rhamnosus HMS2-1 |
| 516 |          | polysaccharide biosynthesis export protein | Polysacc_synt (PF01943) and 6 transmembrane domains | Oligosaccharide translocase (YP_003352680.1) | Lactococcus lactis subsp. lactis KF147 |
| 1425 |          | Glycosyltransferase family 2 | Glycosyltransferase family 2 (PF00535) | WelA (CCC15460.1) | 51 |
| 1011 |          | hypothetical protein | Polysaccharide pyruvyltransferase (PF04230) | Polysaccharide pyruvyltransferase (YP_005047298.1) | Clostridium clariflavum DSM 19732 |
| 1086 |          | Glycosyltransferase group 2 family protein | Glycosyltransferase family 2 (PF00535) | WelL (CCC15461.1) | Lactobacillus pentosus IGI |
| 981 |          | hypothetical protein | 9 transmembrane domains | hypothetical protein (ZP_06987467.1) | Bacteroides sp. 3_1_19 |
| 1323 |          | Glycosyltransferase | No hit in Pfam database | WclB (ABQ58967.1) | Streptococcus oralis |
| 792 |          | UDP-galactopyranosemutase | UDP-galactopyranosemutase(PF03275) | UDP-galactopyranosemutase (ZP_09453942.1) | L. zeae KCTC 3804 |
| 1122 | Tyrosine-protein kinase EpsD | ATPase MipZ (PF09140) and CibA (PF01656) | Wze (ZP_09453943.1) | L. zeae KCTC 3804 | 89 |
| wze | 756       | Tyrosine-protein kinase transmembrane modulator EpsC | Chain length determinant protein (PF02706) and 1 transmembrane domain | Wzd (ACN94846.1) | L. rhamnosus GG | 79 |

**ORF**: open reading frame. **BLASTx**: Basic Local Alignment Search Tool. **Pfam**: Protein families. **Domain(s)**: protein domain(s). **Protein**: protein name. **Organism**: bacterial species. **% Amino acid identity**: percentage of amino acid identity between the query sequence and the best BLASTx hit.
GG. Resistance towards host defence molecules such as antimicrobial peptides is identified as an important virulence phenotype and considered as a target for antimicrobial therapy for infectious diseases of humans [31].

Surface-exposed proteins

Surface-exposed proteins were identified using genome-wide analysis for predicted signal sequences; sortase motif, lipid attachment motif, choline binding motif, YSIRK signal peptide for Gram-positive cell-wall attached proteins, membrane proteins and exported proteins (Table 7). In *L. rhamnosus* LRHMDP2 and *L. rhamnosus* LRHMDP3, as well as in *L. rhamnosus* GG, surface proteins with YSIRK signal peptides found in other Gram-positive cell-wall attached proteins and putative choline binding motifs, were absent. However, 12 LPXTG-type cell wall surface anchor proteins and putative choline binding motifs, proteins with YSIRK signal peptides found in other Gram-positive LRHMDP3, as well as in *L. rhamnosus* GG, were absent. However, 12 LPXTG-type cell wall surface anchor proteins could be identified in *L. rhamnosus* LRHMDP2 and 10 in *L. rhamnosus* LRHMDP3 (Table 8). As expected, surface-exposed pilus-specific proteins (SpaCBA) of *L. rhamnosus* GG were not detected in either clinical isolate (see comparative genomic analysis). LPXTG-type Gram-positive anchor regions with designated biological functions were identified in both clinical isolates. These included a putative cell wall surface anchor family protein with a repeating unit of a collagen binding domain, a subtilisin-like serine protease, a putative cell wall anchored protein SasC (LPXTG motif), similar to MabA, a modulator of *L. rhamnosus* GG adhesion and biofilm formation and a cell wall surface anchor family protein (internalinJ), renamed as mucus binding factor (MBF) in *L. rhamnosus* GG. Other LPXTG-type Gram-positive anchor regions are currently identified as hypothetical proteins with no assigned biological function.

The putative cell wall surface anchor family protein with a repeat unit of collagen binding protein domain B is unique to clinical isolates and no homolog could be identified in *L. rhamnosus* GG. Collagen binding domains (CBD) from the bacterial adhesion domain superfamily are commonly detected in Firmicutes [32]. Collagen is one of the most common components of the extracellular matrix (ECM) and binding to ECM frequently precedes the invasion of host tissues. Organic material in dentinal matrix mainly consists of collagen type I fibrils and it has been suggested that adhesion to collagen components within dentinal tubules assists bacterial invasion [33]. In earlier studies we elucidated the microbiome invading dentinal tubules in situ [6]. Based on the phases of infection, we could identify lactocabacilli consistently in the limited and established stages of infection of dental pulp [6]. Persistence of *Enterococcus faecalis* in root canals of failed endodontic treatment was associated with collagen binding protein and serine protease activity [34]. Similarly, in *S. mutans*, a collagen-binding protein, Cnm, is implicated as a virulence factor that could contribute to infection [35]. Growing evidence for association between Microbial Surface Component Recognizing Adhesive Molecule (MSCRAMM) and disease processes implies that a cell wall surface anchor family protein with repeat units of a collagen binding domain and subtilisin-like serine protease activity could play a crucial role in the apparent invasive behaviour of *L. rhamnosus*, which could be confirmed by in-situ mRNA expression studies in the context of tissue pathophysiology. A subtilisin-like serine protease identified in both clinical isolates was not detected in *L. rhamnosus* GG, but is present in *L. rhamnosus* strain Lc705 [13].

In both clinical isolates the putative cell wall anchored protein SasC (LPXTG motif), a 200 amino acid Gram-positive anchor (LRHMDP2_2569 and LRHMDP3_2533), showed 97% identity to the LPXTG-motif cell wall anchor domain of the extracellular matrix protein, MabA (LGG_01865) from *L. rhamnosus* GG. However, in both clinical isolates, except for the LPXTG motif, no DUF1542 (domain of unknown function 1542) could be identified. MabA containing LPXTG-motif cell wall anchor domain with multiple DUF1542 repeat domains, was characterised as a modulator of adhesion and biofilm formation in probiotic *L. rhamnosus* GG [36]. In *S. aureus*, SasC comprises an N-terminal signal peptide with a YSIRK motif, a C-terminal LPXTG motif and a FIVAR (found in various architectures) motif that includes 17 DUF1542 repeat domains. SasC is identified as a novel factor involved in cell aggregation and biofilm formation, and is implicated in *S. aureus* colonization during infection [37]. The putative cell wall anchored protein in *L. rhamnosus* GG is devoid of the saccharide-binding FIVAR domain that mediates hyaluronate and fibrinectin binding [36,38]. In both clinical isolates, a putative homolog of the cell-wall-anchored protein SasC (LPXTG Motif)

---

### Table 6. PTS, ABC and sugar transporters in *L. rhamnosus* GG, LRHMDP2, LRHMDP3.

| Organism | PTS transporters | ABC transporter | sugar transporters |
|----------|------------------|-----------------|-------------------|
| LRHMDP2  | 85               | 69              | 70                |
| LRHMDP3  | 86               | 70              | 71                |
| GG       | 75               | 85              | 56                |

*Count from annotations.  
Count from KEGG.  
doi:10.1371/journal.pone.0090643.t006

### Table 7. Surface exposed proteins in *L. rhamnosus* GG, LRHMDP2, LRHMDP3.

| Surface exposed proteins | *L. rhamnosus* GG | *L. rhamnosus* LRHMDP2 | *L. rhamnosus* LRHMDP3 |
|--------------------------|-------------------|------------------------|------------------------|
| Putative Signal Peptides* | 253               | 235                    | 179                    |
| Candidate lipoprotein signal peptides* | 145               | 141                    | 143                    |
| Putative choline-binding motifs | 0                 | 0                      | 0                      |
| Signal peptide YSIRK for Gram-positive cell wall-attached proteins* | 0                 | 0                      | 0                      |
| Membrane proteins | 756               | 715                    | 706                    |
| Exported proteins | 98                | 135                    | 136                    |

*Identified by SignalP4.1 using a cutoff of Y score lower limit of 0.3.  
Matches to pattern [DEKR]{(6)}-[LIVMVWFSTAG]{(2)}-[LIVMVFWSTAGCQ]{(2)}-[AGS]{(2)}-C by prosite scan;  
Identified from proteins predicted to have signal sequences by [YS]{(2)}-[SA]{(2)}-[IL]{(2)}-[RK]{(2)}-[x(2)]-[G-x(2)]-[S].

doi:10.1371/journal.pone.0090643.t007
Table 8. LPXTG-type Gram-positive anchor regions.

| LRHMDP2: |  |
|---|---|
| gi|411186888|gb|EKS54010.1| hypothetical protein LRHMDP2_251 |
| gi|411186173|gb|EKS53298.1| hypothetical protein LRHMDP2_549 |
| gi|411185004|gb|EKS52134.1| Subtilisin-like serine protease |
| gi|411183640|gb|EKS50777.1| hypothetical protein LRHMDP2_1916 |
| gi|411183327|gb|EKS50466.1| internalin J |
| gi|411182566|gb|EKS49713.1| Adhesion exoprotein |
| gi|411182429|gb|EKS49578.1| hypothetical protein LRHMDP2_2346 |
| gi|411181795|gb|EKS48956.1| putative cell-wall-anchored protein SasC (LPXTG motif) |
| gi|411181505|gb|EKS48676.1| Phage tail fiber protein |
| gi|411181513|gb|EKS48684.1| hypothetical protein LRHMDP2_2752 |
| gi|411181406|gb|EKS48582.1| hypothetical protein LRHMDP2_2786 |
| gi|411181363|gb|EKS48543.1| Cell wall surface anchor family protein putative |
| LRHMDP3: |  |
|---|---|
| gi|411185129|gb|EKS52258.1| hypothetical protein LRHMDP3_629 |
| gi|411184803|gb|EKS51934.1| hypothetical protein LRHMDP3_1001 |
| gi|411184385|gb|EKS51518.1| Subtilisin-like serine protease |
| gi|411182972|gb|EKS50114.1| cell surface protein |
| gi|411182504|gb|EKS49652.1| hypothetical protein LRHMDP3_2273 |
| gi|411182229|gb|EKS49382.1| hypothetical protein LRHMDP3_2400 |
| gi|411181925|gb|EKS49083.1| putative cell-wall-anchored protein SasC (LPXTG motif) |
| gi|411181547|gb|EKS48716.1| cell wall surface anchor family protein |
| gi|411181369|gb|EKS48548.1| hypothetical protein LRHMDP3_2824 |
| gi|411181377|gb|EKS48556.1| Phage tail fiber protein |
| gi|411181289|gb|EKS48474.1| Cell wall surface anchor family protein putative |

doi:10.1371/journal.pone.0090643.t008

locus, LRHMDP2_2895 (comprising 1716 amino acids) and LRHMDP3_2912 (comprising 1744 amino acids) with no putative conserved domains, was also detected, showing 96% identity to LGG_01865.

Altered domain structure of the cell wall anchored proteins in both clinical isolates and in probiotic GG due to the absence of YSIRK signal peptide surface proteins, as well as absence of DUF1542 repeat domains in clinical isolates, requires further investigation.

The cell wall surface anchor family protein (internalin J) identified in both clinical isolates showed 98% sequence identity to internalin J (LGG_02337) from L. rhamnosus GG. This protein was characterised and renamed as mucus binding factor (MBF), together with MabA protein, is postulated to provide ancillary support in the adhesion process once Spa C has expressed proteins with adhesive properties that contribute to virulence. Paradoxically, in addition to the presence of five CBA pilus cluster that is considered to establish primary contact in adhesive processes. This mechanism is supported by ancillary involvement of MBF and MabA proteins for prolonged persistence of L. rhamnosus GG in the gastrointestinal tract.

Comparative genomics based on COG analysis of bifidobacteria, lactobacilli and other probiotic bacteria has also indicated that virulence genes are not a separate category, but instead pathogenic strains are characterised by variation in genes assigned to COG categories M (cell wall/membrane biosynthesis) and O.
(post-translational modifications and chaperones) [11]. Cai et al. proposed a model for niche-associated evolution in L. casei showing evolution of L. casei niche generalists from the Lactobacillus ancestor with gain and/or loss of plasmids and genomic islands. They have also proposed that enhanced fitness of dairy specialist L. casei occurred by further gene loss from niche generalists capable of adaptation to wider range of environmental conditions [42]. Genetic divergence between L. rhamnosus clinical isolates implicated in infection of dental pulp and probiotic L. rhamnosus GG provides an indication of possible niche adaptation. This is a phenomenon prevalent in the evolution of lactobacilli associated with mucosal surfaces and also from food-related habitats [43,44].

Conclusions

Dental caries is a progressive chronic infection associated with polymicrobial aetiology. Progression of lactobacilli through dental tubules and the presence of apparently encapsulated L. rhamnosus in the initial stages of pulp infection are indicative of an important role in advanced stages of the infection. In both clinical isolates, the absence of the L. rhamnosus GG SpaCBA pilus cluster (Table S9), existence of a modified MabA-like protein, a unique repeat unit of a collagen binding protein domain and the presence of an altered exopolysaccharide gene cluster (Table 4, Table 5, Table S9), suggests the surface components of the clinical isolates of L. rhamnosus may employ a different mechanism for invasion of dental pulp. Surface components of L. rhamnosus clinical isolates are postulated to facilitate the observed invasion of dental pulp by adherence both to cellular elements and extracellular matrices. Invasion could also be facilitated by the evasion of host defense mechanisms, through the masking effect of an extracellular polysaccharide layer. The presence of capsular polysaccharide (CPS) or extracellular polysaccharide (EPS) is widespread in Gram-positive and Gram-negative pathogens. These serve dual roles in survival and pathogenic action by modulating adaptive immune responses and by regulating phagocytosis [45]. Therefore, transcriptional orientation of the eps cluster genes, the presence of two genes homologous to priming glycosyltransferases, the absence of mlACBD genes involved in the dTDP-rhamnose biosynthetic pathway and the presence of a family 2 glycosyltransferase in the eps cluster of both clinical isolates of L. rhamnosus, is predicted to alter EPS composition and could influence pathogenicity. Interestingly, the presence of putative CDSs encoding CtsR (class III stress gene repressor), RpoN (654 RNA polymerase), mannose PTS and internalin J as identified in the clinical isolates, correlates with the presence of homologues in L. monocytogenes, a firmicute involved in invasive gastrointestinal infection.

The impact of acquisition of new genes by the two clinical isolates as evident in the segregation of the clinical isolates from other probiotic L. rhamnosus strains and specifically, the SNP divergence from L. rhamnosus GG, will be better understood with future transcriptomic and functional studies.

Materials and Methods

Collection and processing of carious teeth

Carious teeth were collected from adult patients attending the Westmead Centre for Oral Health. The protocol was approved by the Human Ethics Committee, Sydney West Area Health Service. The decision by the clinician, independent from the study, to extract the carious tooth, was based on clinical examination and radiographs; alternatively, extraction of the tooth was the patient’s informed decision. The patient signed the consent form to participate in the research. The carious tooth was not included in the study if the patient had antibiotic administered in the two weeks preceding the extraction or if the patient had reported a history of Hepatitis C or HIV. Early stage of pulp infection was inferred based on radiographic image(s) showing integrity of periodontal ligament with no shadow effect for peri-apical infection and with no clinical signs or symptoms of irreversible pulpitis, based on patient descriptors such as thermal sensitivity, pain and postural variation. Stages of infection were subsequently re-assessed during downstream processing of the tissues. Carious teeth meeting the above criteria were transported in sterile and reduced PBS glycerol to the Institute of Dental Research within 2 h of extraction and were processed. Carious dentine was excavated to near totality and the scored tooth was rinsed in sterile 18-MΩ water twice before splitting to avoid contamination from the carious dentine. A longitudinal groove was carefully carved around the carious tooth to bisect the carious lesion using a sterilized bur and pressure syringe and sterile 18-MΩ water; the tooth was split sagittally with sterile hand instruments and the extent and location of the carious lesion and patho-physiology of pulp tissue was noted. Based on these observations, classification of the carious lesion for stages of infection was re-assessed. An intact pulp with mild erythema was provisionally classified as at an early (limited) stage of infection. Exposed pulp localized on one side of the bisected tooth was carefully lifted with sterile tweezers, rinsed in sterile, reduced PBS with 15% glycerol and placed in a sterile screw capped cryo-tube containing 0.4 ml sterile PBS with 15% glycerol. The resuspended pulp was homogenized using a 2 ml sterile glass homogenizer and stored at −80°C after subsequent processing.

PCR analysis of bacterial phyta in infected pulp tissue

Earlier findings based on FISH studies [6] indicating that fewer taxa are detected in early stages of infection was confirmed by PCR analysis of bacterial phyta in infected pulp tissue (Table S1).

Isolation and identification of lactobacilli from infected vital pulp

An aliquot (20 μl) of homogenized pulp sample was streaked on freshly prepared Man, Rogosa and Sharpe (MRS) agar and incubated anaerobically at 37°C for 48 h. Each bacterial colonial form with distinct morphology was subcultured in MRS medium (2 ml). Cultures were incubated anaerobically at 37°C for 24 h. An aliquot (0.85 ml) of bacterial culture was collected in screw capped cryo-tubes containing 0.15 ml sterile glycerol, homogenized and stored at −80°C for future use. Remaining culture was processed for routine Gram stain to study bacterial morphology and for purification of genomic DNA as described [6]. Briefly, an aliquot (40 μl) of homogenised pulp was digested at 56°C for 40 min in cell lysis buffer consisting of 2 mg ml⁻¹ proteinase K (Qiagen), 2 mg ml⁻¹ lysozyme (Roche) and 1000 U ml⁻¹ mutanolysin (Sigma) in 10 mM sodium phosphate buffer pH 6.7. After the addition of 0.1% SDS (Sigma), DNA was extracted using QIAamp DNA Mini kit (Qiagen) to give a final elution volume of 200 μl as described [6]. Genomic DNA isolated from bacterial cultures showing morphology of Gram-positive rods was subjected to PCR using Lactobacillus genus-specific and species-specific primers. Qualitative PCR and sequence identity for the 16S rRNA gene was performed as described previously [10].

Design and evaluation of PCR primers

Primers were designed using Primer Express (Applied Biosystems) for the detection of bacteria from the families, Streptococcaceae, Acidaminococcaceae, Lachnospiraceae and Coriobacteriaceae as well as from
Primers used for the detection of Lactobacillus, Prevotella and F. nucleatum have previously been reported (Table S2a). Primers for the L. rhamnosus exopolysaccharide and pilus cluster gene set (Table S8) were designed using OligoExplorer 1.2 (Gene Link) and the specificity of the primers determined by DNA sequencing of the amplicons produced (Australian Genome Research Facility).

DNA amplification for identification of bacteria

Amplificaons of 16S rRNA genes for given taxa were obtained from carious dentine or infected pulp using HotStarTaq Master Mix kit (Qiagen), using 200 nM of each of the appropriate primer set (Invitrogen or Sigma; Table S2) and µL extracted DNA solution from the infected pulp. The PCR reaction was carried out using a GeneAmp PCR system 9700 (Applied Biosystems) in a 25 µL reaction volume at 95°C for 10 min followed by 40 cycles at 95°C for 15 s and 56°C for 1 min or 2 min, respectively.

DNA amplification for identification of L. rhamnosus

exopolysacchiae and pilus cluster genes

In order to amplify specific genes, PCR was performed with HotStarTaq Master Mix (Qiagen), 200 nM of the appropriate primers (Integrated DNA Technologies; Table S8) and µL extracted DNA from eight L. rhamnosus clinical isolates. Amplification was performed at 95°C for 15 min followed by 40 cycles of 95°C for 15 s and 58°C for 1 min or 55°C for 1 min at 60°C for 2 min (Table S8) for the detection of exopolysachriease and pilus cluster genes. DNA isolated from L. rhamnosus ATCC 53103 and L. rhamnosus ATCC 9595 was included as a positive control. The PCR products were subjected to electrophoresis as described above.

DNA sequencing

PCR amplicons were purified using a MO BIO kit (MO BIO Laboratories Inc.) and sequenced at the Australian Genome Research Facility using the appropriate forward primer. Sequence identity of 16S rDNA PCR amplicons was established using BLAST to search nucleotide collection (nr/nt) databases accessed through NCBI (http://www.ncbi.nlm.nih.gov/).

Whole genome sequencing using Roche GS FLX+

Two genomic DNA samples (1–2 µg DNA, A260/A280 ratio in the range 1.8–2.0) were delivered to the Ramaciotti Centre for Gene Function Analysis, University of New South Wales, Australia for whole genome sequencing using Roche GS FLX+. The sequencing was undertaken according to standard Roche/FLX procedures.

Genome Assembly and BLAST analysis

The sequence reads were assembled de novo with Newbler assembler 2.6 using the default parameters. The contigs assembled by Newbler were then joined by CAP3 [46] using default parameters. The final sets of contigs were cross-checked against the Lactobacillus rhamnosus GG genome sequence using blastn as well as blastx with an E-value cut-off of 10e-6. The absence of a gene was defined as no BLAST hits against the reference or a normalized blastx score of less than 2, where normalized blastx score was calculated as BLAST score divided by alignment length.

Genomic Distance Analysis and SNP Detection

Genomic distance between the eleven probiotic L. rhamnosus strains; GG, Lc705, HN001, R0011, LMS2-1, CASL, LOCK900, LOCK908, ATCC 8530, ATCC 21052, MTCC 5462 and the 2 clinical isolates L. rhamnosus LRHMDP2 and L. rhamnosus LRHMDP3 was analyzed by whole-genome BLAST comparison using the NCBI web-based BLAST, with an E-value threshold of 10e-6. The whole genome sequence of L. rhamnosus GG was used as a query sequence in the BLAST search against the other genomes. The Tree View option was then selected to generate a dendrogram using the neighbour-joining method that clusters sequences according to their distances from the query sequence.

SNP divergence between the probiotic L. rhamnosus strain GG and the two clinical isolates L. rhamnosus LRHMDP2 and L. rhamnosus LRHMDP3 was investigated using the web-based program SNPs Finder (http://snpsfinder.lanl.gov) [47]. The 3 genomes were uploaded at the same time for SNP detection. SNPs were evaluated in homologous regions of 600 bp that shared sequence similarity of at least 95%. Protein coding information for L. rhamnosus GG was used as the gene coordinate file in SNP detection for anchoring gene positions.

Supporting Information

Figure S1 Dendrogram of genomic difference between the 2 clinical isolates, L. rhamnosus LRHMDP2 and L. rhamnosus LRHMDP3 and 11 other L.rhamnosus strains based on genomic BLAST. Genomic distance between eleven L. rhamnosus strains; GG, Lc705, HN001, R0011, LMS2-1, CASL, LOCK900, LOCK908, ATCC 8530, ATCC 21052, MTCC 5462 and the 2 clinical isolates L. rhamnosus LRHMDP2 and L. rhamnosus LRHMDP3 was analyzed by whole-genome BLAST comparison using the NCBI web-based BLAST, with an E-value threshold of 10e-6. The whole genome sequence of L. rhamnosus GG was used as a query sequence in the BLAST search against the other genomes. The dendrogram was generated using the neighbour-joining method that clusters sequences according to their distances from the query sequence. Percent relatedness is indicated on the scale.

Table S1 PCR analysis of bacterial taxa in infected dental pulp tissue.

Table S2 Primers used to amplify taxa (a) and Specificity of primer pairs for detecting a given taxon (b).

Table S3 Genes from the reference L. rhamnosus strain GG not present in both assemblies.

Table S4 Genes from the reference L. rhamnosus strain Lc705 not present in both assemblies.

Table S5 New genes present in L. rhamnosus LRHMDP2 (a) and L. rhamnosus LRHMDP3 (b).

Table S6 Strain specific genes for L. rhamnosus LRHMDP2 (a) and L. rhamnosus LRHMDP3 (b).

Table S7 SNP divergence between L. rhamnosus GG and L. rhamnosus LRHMDP2 and LRHMDP3.
Table S8 Primers used to amplify selected exopolysaccharide and pilus cluster genes.

Table S9 Targeted genetic profile of exopolysaccharide and pilus cluster genes for clinical isolates from pulp sample.

Acknowledgments

We thank Dr. Nattida Charadram for collection of teeth with carious lesions and extraction of pulp.

References

1. Paster BJ, Dewhirst FE. (2009) Molecular microbial diagnosis. Periodontology 2000 51:36–44.
2. Belda-Ferre P, Alcaraz LD, Cabrera-Rubio R, Romero H, Simon-Soro A, et al. (2012) The oral metagenome in health and disease. The ISME journal 6:46–56.
3. Griffin AL, Beall CJ, Campbell JH, Firestone ND, Kumar PS, et al. (2012) Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing. The ISME journal 6:1176–1185.
4. Ling Z, Kong J, Jia P, Wei G, Wang Y, et al. (2010) Analysis of oral microbiota in children with dental caries by PCR-DGGE and barcoded pyrosequencing. Microbial ecology 60:677–690.
5. Zaura E, Keijger BJ, Huse SM, Crielard W (2009) Defining the healthy “core microbiome” of oral microbial communities. BMC microbiology 9:259.
6. Nadkarni MA, Simonian MR, Harty DW, Zozulin H, Jacques NA, et al. (2010) Lactobacilli are prominent in the initial stages of polymicrobial infection of dental pulp. J Clin Microbiol 48:1732–1740.
7. Schaudinn C, Carr G, Gorur A, Jaramillo D, Costerton JW, et al. (2009) Imaging of endodontic biofilms by combined microscopy (FISH/cLSM - SEM). Journal of microscopy 235:124–127.
8. Huttenhower C, HMP Consortium (2012) Structure, function and diversity of the healthy human microbiome. Nature 486:207–214.
9. Chen Z, Wilkins MR, Hunter N, Nadkarni MA (2013) Dhrall Genome Sequences of Two Clinical Isolates of Lactobacillus rhamnosus from Initial Stages of Dental Pulp Infection. Genome Announce 1:1.
10. Byun R, Nadkarni MA, Chlouer KL, Martin FE, Jacques NA, et al. (2004) Quantitative analysis of diverse Lactobacillus species present in advanced dental caries. J Clin Microbiol 42:3128–3136.
11. Lukjanenko O, Usery DW, Wassenaar TM (2011) Comparative genomics of Bifidobacterium, Lactobacillus and related probiotic genera. Microbial ecology 63:651–673.
12. Aza R, Bartel D, Best AA, De Jongh M, Ditz T, et al. (2008) The RAST Server: rapid annotations uses subsystems technology. BMC genomics 9:75.
13. Kankainen M, Paulin L, Tyrykynen S, von Ossowski I, Reunanen J, et al. (2009) Comparative genomic analysis of Lactobacillus rhamnosus GG reveals pilus containing a human- mucus binding protein. Proceedings of the National Academy of Sciences of the United States of America 106:17193–17198.
14. Macklaim JM, Gloor GB, Anakam KC, Cribby S, Reid G (2011) At the crossroads of vaginal health and disease, the genome sequence of Lactobacillus gasseri ATCC 33323. BMC microbiology 10:77.
15. Francl AL, Thongaram T, Miller MJ (2010) The PTS transporters of Lactobacillus casei and its regulation by inducer exclusion. Research in microbiology 159:94–102.
16. Davidson AL, Dussa E, Orelle C, Chen J (2008) Structure, function, and evolution of bacterial ATP-binding cassette systems. Microbiol Mol Biol Rev 72:317–364.
17. Abranches J, Miller JH, Martinez AR, Simpson-Haidaris PJ, Burne RA, et al. (2009) Comparative Genome Analysis of Tissue Invasive Enterococcus faecalis. PLoS Genetics 5:e1000439.
18. Hubble TS, Hatton JF, Nallapareddy SR, Murray BE, Gillespie MJ (2003) Identification of a Gene Cluster for the Biosynthesis of a Long, Galactose-Rich Exopolysaccharide in Lactobacillus casei. Journal of Bacteriology 185:2390–2398.
19. Welman AD, Maddex BS, Miller JH, Fox PR (2003) Comparative analysis of the exopolysaccharide biosynthesis gene clusters from four strains of Lactobacillus casei. Microbiology (Reading, England) 151:1309–1315.
20. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabinowitch E, et al. (2006) Generic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS genetics 2:e31.

Data Availability: The Whole Genome Shotgun project described in this paper has been deposited at DDBJ/EMBL/GenBank under the accession numbers AMQW00000000 for L. rhamnosus LRHMIDP2 and AMQX00000000 for L. thalmansis LRHMIDP3.

Author Contributions

Conceived and designed the experiments: MAN. Performed the experiments: MAN. Analyzed the data: MAN NC MRW. Contributed reagents/materials/analysis tools: MAN NH ZC MRW. Wrote the paper: MAN NH ZC MRW.
44. Makarova KS, Koonin EV (2007) Evolutionary genomics of lactic acid bacteria. Journal of bacteriology 189:1199–1208.
45. Avci FY, Kasper DL (2010) How bacterial carbohydrates influence the adaptive immune system. Annual review of immunology 28:107–130.
46. Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. Genome research 9:868–877.
47. Song J, Xu Y, White S, Miller KW, Wolinsky M (2005) SNPsFinder—a web-based application for genome-wide discovery of single nucleotide polymorphisms in microbial genomes. Bioinformatics (Oxford, England) 21:2083–2084.
48. Stothard P, Wishart DS (2005) Circular genome visualization and exploration using CGView. Bioinformatics (Oxford, England) 21:537–539.