Comparative Genome Analysis Reveals Important Genetic Factors Associated with Probiotic Property in *Enterococcus faecium* strains

Running title: Comparative Genome of *Enterococcus faecium* strains

Vikas C. Ghattargi [1,2], Meghana A. Gaikwad[1], Bharati S. Meti[2], Yogesh S. Nimonkar[1], Kunal Dixit[1], Om Prakash, Yogesh S. Shouche[1], Shrikant S. Pawar[1#], Dhiraj Dhotre[1#]

[1]National Centre for Microbial Resources (NCMR), National Centre for Cell Science (NCCS), Pune- 411021, Maharashtra, India

[2]Department of Biotechnology, Basaveshwar Engineering College, Bagalkot- 587102, Karnataka, India

Corresponding Author(s)

Dr. Dhiraj Dhotre, National Centre for Microbial Resource (NCMR), National Centre for Cell Science (NCCS), Pune- 411021, Maharashtra, India. Email: dhiraj@nccs.res.in

Dr Shrikant S. Pawar, National Centre for Microbial Resource (NCMR), National Centre for Cell Science (NCCS), Pune- 411021, Maharashtra, India. Email: shrikantpawar@nccs.res.in

ABSTRACT

Enterococcus faecium though commensals in human gut, few strains provide beneficial effect to humans as probiotics, few are responsible for nosocomial infection and few as non-pathogens. Comparative genomics of *E. faecium* will help to reveal the genomic differences responsible for the said properties. In this study, we compared *E. faecium* strain 17OM39 with a marketed probiotic, non-pathogenic non-probiotic (NPNP) and pathogenic strains. The core genome analysis revealed, 17OM39 was closely related with marketed probiotic strain T110. Strain 17OM39 was found to be devoid of known vancomycin, tetracycline resistance genes and functional virulence genes. Moreover, 17OM39 is 'less open' due to absence of frequently found transposable elements. Genes imparting beneficial functional properties were observed to be present in marketed probiotic T110 and 17OM39 strains. Additional, genes associated with colonization within gastrointestinal tract were detected across all the strains. Beyond shared genetic features; this study particularly identified genes that are responsible to impart probiotic, non-pathogenic and pathogenic features to the strains of *E. faecium.* The study also provides insights into the acquired and intrinsic drug resistance genes, which will be helpful for better understanding of the physiology of antibiotic resistance in *E. faecium* strains. In addition, we could identify genes contributing to the intrinsic ability of 17OM39 *E. faecium* isolate to be a potential probiotic.

The study has comprehensively characterized genome sequence of each strain to find the genetic variation and understand effects of these on functionality, phenotypic complexity. Further the evolutionary relationship of species along with adaptation strategies have been including in this study.

Keywords: non-pathogenic, pathogenic, Indian, comparative genome analysis, *In-silico* analysis

BACKGROUND

The genus *Enterococcus* is one of the diverse and ecologically significant group and members of this genus are ubiquitously distributed in nature viz. animals, human gastrointestinal tract (GIT) and plants (Lam et al. 2012; Qin et al. 2012; Geldart and Kaznessis 2017; McKenney et al. 2016; dos Santos et al. 2015; Byappanahalli et al. 2012; Rasouli Pirouzian et al. 2012). *Enterococcus* plays an important role in the ripening of cheese products by lipolysis and proteolytic properties leading to the development of aroma and flavour (Rasouli Pirouzian et al. 2012). In Mediterranean region, *Enterococcus* spp have been used in the preparation of various meat and fermented milk products for centuries (dos Santos et al. 2015). Further, they also exhibit the beneficial property of bacteriocin production (Rasouli Pirouzian et al. 2012; dos Santos et al. 2015) presenting activity against potential pathogens viz. group D streptococci and *Listeria* in various foods and in GIT (Rasouli Pirouzian et al. 2012).

E. faecium is widely and extensively studied for its leading cause of nosocomial infections in humans (Guggenbichler et al. 2011). It is a gut commensal and acts as opportunistic pathogen due to a variety of virulence factors, including lipopolysaccharides and biofilm formation (Natarajan and Parani 2015). Their pathogenic nature is evident in urinary tract infections, endocarditis, and surgical wound infection, displaying its capability of causing a wide range of infections (Kajihara et al. 2015). Another remarkable character of *E. faecium* is its tolerance to many antimicrobial drugs (Coque et al. 2005; O'Driscoll and Crank 2015). It has also acquired antibiotic-resistance gene against vancomycin and a multidrug resistance beta-lactamase gene (Miller et al. 2014). Besides, it has been shown that *E. faecium* is capable to acquire resistance to antibiotics by sporadic mutations and infections caused by these are normally difficult to treat (Hollenbeck and Rice 2012). The strains like Aus0004 and V583 are reported as pathogens (Lam et al. 2012).

Numerous studies in the last decade have validated the safety claim of *Enterococci* in foods and as probiotics (Huys et al. 2013; Araújo and Ferreira 2013; Giraffa 2002; Franz et al. 2003). The application of *Enterococci* as a starter culture e.g *Enterococcus faecium* SF68 (Switzerland) and as probiotic e.g *E. faecium* T110 (Japan) has been used widely (Benyacoub et al. 2005; Jong-Hoon Lee, Donghun Shin, Bitnara Lee, Hyundong Lee, Inhyung Lee 2017). Additionally, *E. faecium* T110 is a content of many commercial available probiotics and no cause of illness or death has been reported (Natarajan and Parani 2015). *E. faecium* is among one of the directly fed microorganism recognized by the Association of American Feed Control, 2016. It is permitted as probiotic supplement in the diet for poultry, dogs, piglets and mice (Kačániová et al. 2006; Kreuzer et al. 2012; Vahjen and Männer 2003; Benyacoub et al. 2005). Few strains of *E. faecium* (NRRL B-2354) act as surrogate microorganism used in place of pathogens for validation of thermal processing technologies (Kopit et al. 2014) and some are widely used as laboratory strains e.g. *E. faecium* 64/3 (Bender et al. 2015). These two strains are non-pathogenic and are used routinely without any known disease outbreak (VanRenterghem 2012).

Thus the diversity and plasticity of *E. faecium* are accountable for both probiotic and pathogenic nature (Abeijón et al. 2006; Hassanzadazar et al. 2014; Satish Kumar et al. 2011). The work described here elucidates the genetic divergence between strain 17OM39 with marketed probiotic, non-pathogenic, and pathogenic strains to identify the genes reported for pathogenicity, antibiotic resistance, and probiotic properties. An attempt has also been made to identify the genes present exclusively in probiotic strains.

RESULTS AND DISCUSSION

Strain selection

Whole genome sequences were downloaded from NCBI genome database and the strains were grouped into probiotic, non-pathogenic non-probiotic (NPNP) and pathogenic based on

the literature survey (Table1). The pathogenic group had six strains: DO, Aus0004, Aus0085, 6E6, E39 and ATCC 700221(Lam et al. 2012; Qin et al. 2012; Geldart and Kaznessis 2017; McKenney et al. 2016). The first four were isolated from human blood and latter two from human stool. The non-pathogenic group had two strains: NRRL B-2354 and 64/3(Kopit et al. 2014; Bender et al. 2015). The probiotic group had the marketed strain T110 (Natarajan and Parani 2015) and strain 17OM39 that was isolated from healthy human gut (Ghattargi et al. 2018).

General genomic features

Genome sizes ranged from approximately 2.57–2.99 Mb with strain DO exhibiting the smallest and 6E6 the largest genome. Average GC content varied between 37.90 ± 0.65% and the strains with high G+C% do not have higher CDS, this contradicts with the results stated earlier (Bonacina et al. 2017). The genomic features of the strains under study are provided in Table 1. No significant differences ($P \leq 0.05$, Kruskal–Wallis statistical test) could be noted between the groups with respect to their genome size, GC content, average number of genes and coding DNA sequence (CDS).

The RAST annotation has facilitated to determine the features, assigned to subsystems that are present in all organisms (Figure S1). The average number of annotated protein-encoding genes for the probiotic group was 2,570; for NPNP group was 2,639 and 3,093 for the pathogenic group. Annotation based on RAST for the strains under study suggests an abundance of carbohydrates and protein metabolism subsystems. The enriched carbohydrates metabolism is in agreement with the *E. faecium* ability to utilize a wide range of mono-, di-, oligo-saccharides (Devriese et al. 1987; Manero and Blanch 1999).

Comparisons of 17OM39 with other E. faecium strains

The availability of complete *E. faecium* genomes has helped to define core, accessory and unique genomic features for all the strains. The comparison of strain 17OM39 with other strains of probiotic, NPNP and pathogenic strains, revealed 1935 (85.53 %) core genes, 526 (20.64%) accessory and 87(3.41%) unique genes. The numbers of shared genes are plotted as the function of number of strains (Figure 1). The figure shows that pan-genome size grew continuously with the addition of strains indicating an open pan-genome and the results are in accordance with previous study for *E. faecium* genome (Mikalsen et al. 2015). In contrast to the pan-genome, the size of the core-genome gradually stabilized (Mikalsen et al. 2015).

Pan-Genome Analysis

The pan genome analysis revealed the presence of 1935 core genes and 5718 accessory genes (Figure 2). The number of strain-specific genes observed was 67, 87, 10, 64, 62, 16, 13, 36, 54 and 14 for strains 17OM39, T110, NRRL B-2354, 64/3, DO, AUS0004, AUS0085, 6E6, E39 and ATCC700221, respectively (Figure 2). Identification of core, accessory and unique gene families by orthoMCL analysis revealed the proportion of known, hypothetical and uncharacterized proteins in these groups (Figure S2A). A large percentage (61.19%) of unique genes was assigned to an uncharacterized group and further studies are required to examine these unexplored attributes.

orthoMCL analysis of **core genes** led to the identification of 850 genes present in single copy and 772 genes present in multiple copies across all ten strains. Functional analysis of the core genes showed distribution in a varied range of functional categories within Cluster of Orthologous Genes (COG) viz. growth, DNA replication, transcription, translation, carbohydrate and amino acid metabolism, stress response and transporters. In contrast to earlier reports, core genes were also found in the functional categories of secondary metabolism and motility (Palmer et al. 2012; Beukers et al. 2017). Categories representing

transport and metabolism of coenzyme, lipid, amino acid, nucleotide comprised of 16.24% of the core genes, while 11.30% of core genes were ascribed to carbohydrate metabolism which is in agreement with an earlier report (Beukers et al. 2017).

Functional analysis of the **accessory genes** showed diverse distribution in COG categories as similar to core gene annotations (Figure S3). An important observation was seen in subsystems a) carbohydrate metabolism and b) replication, recombination and repair systems. The former was abundant in the probiotic group ($p = 0.002$) while latter in the pathogenic group ($p = 0.039$) (Figure 3). This has been attributed to the properties of probiotic strains to utilize various carbohydrates(Ghattargi et al. 2018); while the pathogenic group had higher abundance of replication and recombination genes known to be associated with a large number of mobile elements(Lam et al. 2012; Qin et al. 2012). We also made an attempt to find the accessory genes being shared between the groups. The probiotic and pathogenic group shared 15 genes; four of them are general transporters, two are manganese-containing catalase gene, which provides resistance to hydrogen peroxide present in human GIT (Figure S2B) (Wang et al. 2014; King et al. 2000; Xu et al. 2016).

The important **unique genes** associated with the various strains are as follows: phosphotransferase (PTS) system for mannose/fructose/sorbose in probiotic strain 17OM39 is involved in sugar uptake (Postma et al. 1993; Monot et al. 2011). Marketed probiotic strain T110 has macrolide-efflux transmembrane protein which acts as drug efflux pump and plays a key role in drug resistance (Sangvik et al. 2005; Stadler and Teuber 2002). The important unique genes for others are hexosyltransferase in strain 64/3, type III restriction-modification system in strain NRRL B- 2354, Cro/CI family transcriptional regulator protein in strain 6E6, transposase for insertion sequence IS1661 in strain ATCC 700221, streptogramin A acetyltransferase in strain Aus0004, Patatin-like proteins in strain Aus0085, IS1668

169 transposase in strain DO, plasmid recombination enzyme in strain E39. An additional file
170 gives the detailed information on core, accessory and unique genes in more detail [see
171 Additional file 1, 2, 3].

172 Thus, the analysis conducted here has shown that pan-genome of *E. faecium* constructed on
173 the basis of 10 genomes is still open, while the core genome seems to have reached almost a
174 closed state. The small size of the core genome and a huge number of accessory genes
175 support the observation of the genomic fluidity of *E. faecium* (Bakshi et al. 2016).

176 ***Antibiotic resistance determinants***

177 *Enterococci* can exhibit resistance to a number of antibiotics, which has been attributed to
178 their innate resistance and also due to their ability to successfully acquire resistance through
179 horizontal gene transfer (HGT) (Tong et al. 2017; Hegstad et al. 2010a). Multiple-drug-
180 resistant strains of *E. faecium* have been increasingly associated with nosocomial infections
181 particularly the vancomycin resistance (McArthur et al. 2013). Thus screening of antibiotic
182 resistance determinants in genomes was necessary in order to understand if probiotic strains
183 harboured these genes.

184 Here we screened the genomes for antibiotic resistance genes using Comprehensive
185 Antibiotic Resistance Database (CARD) (McArthur et al. 2013). Genes conferring resistance
186 to kanamycin were found in all the genomes which have been attributed to the intrinsic
187 property within *E. faecium* (Galimand et al. 2011). The non-pathogenic group showed the
188 presence of general multidrug transporter. Within pathogenic strains, Aus004 and Aus0085
189 showed the presence of tetracycline, trimethoprim and vancomycin resistance gene. Strains
190 E39 and 6E6 showed the presence of genes responsible for trimethoprim and tetracycline
191 resistance. Pathogenic strain E39 presented daptomycin resistance gene and strain ATCC
192 700221 showed the presence of genes responsible for resistance to the antimicrobial activity

of cationic antimicrobial peptides and antibiotics such as polymyxin. Table 2 shows various antibiotic resistance genes found in each strain.

In our study, genes imparting resistance to one or more antibiotics were seen in different strains of *E. faecium*. Overall, the pathogenic group of *E. faecium* was found to have a higher prevalence of antibiotic resistance genes; a factor that contributes to the challenge of selecting therapeutic measures. The probiotic group was devoid of any major clinically relevant antibiotic resistance (Natarajan and Parani 2015; Ghattargi et al. 2018).

Virulence determinants

Virulence genes contribute to the pathogenicity of an organism (Comerlato et al. 2013). Despite the increasing knowledge of *E. faecium* as an opportunistic pathogen, the distribution of virulence factors is still poorly understood (Comerlato et al. 2013). Knowledge of the virulence characteristics helps to understand the complex pathogenic process of the pathogenic strains. This study also determines genes responsible for virulence factors such as adherence, biofilm formation and exo-enzyme production in probiotic, NPNP and pathogenic groups.

The ability to adhere to the GIT is reflected to be one of the main selection criteria for potential probiotics as it extends their persistence in the intestine (Ouwehand et al. 1999) and thus allows the bacterium to exert its probiotic effects for an extended time. However, adhesion is also considered a potential virulence factor for pathogenic bacteria (Kirjavainen et al. 2001). The intestinal mucus is an important site for bacterial adhesion and colonization (Finlay 1997) and thus adherence property is beneficial to humans in case of probiotics and it possesses adverse effects in pathogenic strains. The genes described as adherence factors (*acm, scm, EbpA, EbpC*) have been attributed to pathogenic effects and our study could find most of these genes in the pathogenic group. Excluding strain DO all other pathogenic strains

showed the presence of enterococcus surface protein (*esp)* gene which contributes as a major virulence factor (Heikens et al. 2007; Ramadhan and Hegedus 2005; Toledo-Arana et al. 2001; Baldassarri et al. 2001).

The *bopD* gene involved in biofilm was intact in all groups but the operon was absent in strain 17OM39, marketed probiotic strain T110 (Ghattargi et al. 2018) and non-pathogenic strains (Natarajan and Parani 2015). In an exo-enzyme group, hyaluronidase gene was found to be associated with marketed probiotic strain alone, while the gene in strain 17OM39 displayed an alteration in sequence at position 167 (G>T) suggesting this could affect its functionality due to the nonsense mutation. Gene *acm* in the probiotic and the non-pathogenic group was not functional due to the non-sense mutation at position 1060 (G>T). Also, the virulence gene *scm, efaA* and *srtC* are not well characterized as virulence determinants in *E. faecium* (Natarajan and Parani 2015) (Table 3).

Although one could expect a virulence trait depending on the source of isolation, our study did not find any such traits and differs from the earlier reports (Dahlén et al. 2012). Also, the strains showed significantly different patterns of virulence determinants, which underlines the findings of other author (Dahlén et al. 2012). The strain 17OM39 within the probiotic group was devoid of any clinically relevant functional virulence determinants.

Mobile genetic elements

Mobile genetic elements (MGEs) play an important role in HGT of genes within and between bacteria (Beukers et al. 2017; Jiang et al. 2017; Kaplan 2014; Von Wintersdorff et al. 2016). A number of MGEs have been described in *E. faecium* including transposons, plasmids, and bacteriophage (Hegstad et al. 2010b; Beukers et al. 2017).

Insertion sequences (ISs) are possibly the smallest and most independent transposable elements, thus playing an important role in shaping the bacterial genomes (Siguier et al.

2014). Based on the screening for IS elements (Table S1), the IS1542 was present only in probiotic strains and earlier studies on IS1542 have shown its presence in just 2 out of 65 human pathogenic strains, suggesting no direct relation with the strains pathogenicity (Huh et al. 2004). The IS element ISEfa12 was present only in non-pathogenic group and IS1216, 1S1216E, 1S1216V, IS16, IS6770, ISEf1, ISEfa10, ISEfa11, ISEfa5, ISEfa7, ISEfa8, ISEnfa3, ISS1W were present only in all the strains of the pathogenic group. The presence of insertion sequence families in all groups imply these elements are spread by HGT (Mikalsen et al. 2015). However, particular IS elements are distributed in only one group suggesting that these IS elements have evolved over the time (Werner et al. 2011; Mikalsen et al. 2015). Notably, presence of IS16 has been used as a marker within the hospital strains of *E. faecium* with 98% sensitivity and 100% specificity (Mikalsen et al. 2015). This observation was further supported by detecting IS16 in the only pathogenic group of *E. faecium* strains. Moreover, ISEfa11 and ISEfa5 are associated with vancomycin resistance genes viz. *VanS*, *VanX*, and *VanY* [34,65]. This correlation was also seen in this study (Figure 4).

E. faecium are known to harbour bacteriophages, so the presence of **prophage** was predicated in all the ten genomes [34, 66]. Bacteriophages contribute actively to bacterial evolution by integrating and excising from the genome (Matos et al. 2013). In certain conditions they provide new genetic properties to the bacterial host and leading to the development of new pathogens within species, as shown for *Escherichia coli*, *Vibrio cholera* and *Corynebacterium diphtheriae* (Davis and Waldor 2003; Backman et al. 1983; FREEMAN 1951). We could trace at-least one intact prophage in all the ten genomes (Table S2). In total, there were three incomplete (PHASTER score < 70), twenty-three intact phages (PHASTER score < 70-90) and nine phages whose completeness status was doubtful (PHASTER score < 90). The non-pathogenic strains had two intact prophages, while the probiotic strain 17OM39 had one intact and a doubtful prophage while the other strain T110 had two intact prophages.

Enterococcus faecium ATCC 700221 had the highest number of intact phages. We could not find any known functional virulence factors or genes associated with important functions within these bacteriophages regions. Further, Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) associated (Cas) system were found to be absent within the genomes as opposite to the closed neighbour, *E. faecalis* (Palmer and Gilmore 2010).

Genomic Islands are distinct DNA fragments differing between closely related strains, which usually are associated with mobility (Dobrindt et al. 2004; Juhas et al. 2009). Genomic island comparison between strain 17OM39 with the other strain of *E. faecium* was done to find out the genes transferred by HGT (Table S3). We identified a total of 11 genomic island in strain 17OM39 amounting to 3.5% of total genome. The genomic island (GI1) was common across all groups except for strain 6E6. The choloylglycine hydrolase gene was found to be present in the genomic island of probiotic stain T110, pathogenic strain 6E6 and non-pathogenic strain 64/3. The choloylglycine hydrolase gene imparts resistance to bile salts and thus helps to survive within the gut environment (Jones et al. 2008; Begley et al. 2006). In this study, the pathogenic group showed a large number of IS elements, transposons and antibiotic resistance genes within the genomic island. Pathogenic strains Aus0004 and Aus0085 showed the presence of virulence factor *esp* gene (enterococcal surface protein) and vancomycin resistance genes and only pathogenic strains alone showed the presence of tetracycline resistance gene within the genomic island. They also showed presence of a cell adhesion protein within the Genomic Island. A higher similarity was observed between genomic islands of probiotic and NPNP strains as compared to pathogenic group (Figure 5). The distribution of these MGE's within the genome is as shown in Figure 6 with strain AUS0004 having nearly 25% of its genome as mobile.

E. faecium behaves as probiotic, non-pathogenic and pathogenic these underlying mechanisms may be intrinsic to the strain or acquired by horizontal exchange of genetic

material. Thus genes associated with Genomic Islands can be considered as acquired properties (Juhas et al. 2009; Hollenbeck and Rice 2012; Dobrindt et al. 2004; Gold 2001; Marothi et al. 2005). Genes associated with probiotic properties in strain 17OM39 were found to be associated with the genome and not within any Genome islands. From this study, it is evident that the role of the MGEs in adding new capacities and also driving the evolution of the strains (Jones et al. 2008; Begley et al. 2006). Certainly, studies like one carried out here are helpful to understand the evolution of predominant strains. An additional file gives the detailed information on genes present in Genomic Islands in more detail [see Additional file 4].

Survival in Gastrointestinal Tract

Biologically active microorganisms are usually required at the target site to induce a health benefit or pathogenic effect, but must survive the host's natural barriers such as the acid, bile and must adhere to GIT (İspirli et al. 2015; Banwo et al. 2013; Ahmadova et al. 2013; Rao et al. 2013). Moreover, *E. faecium* is a part of the core microbiome and their number in human faeces ranges from 10^4 to 10^5 per gram (Santagati et al. 2012). Thus a list of genes encoding for survival and growth were observed in the strain 17OM39 and compared with other genomes. We found Permease IIC component gene responsible for catalysing the phosphorylation of incoming sugar substrates (Minelli and Benini 2008) only in the probiotic group. All the groups showed the presence of the genes that impart resistance to acid, bile and can hydrolyze bile salt. Moreover, these strains were also able to adhere and grow in the GI conditions. This finding correlates with the fact that *Enterococcus faecium* are normal inhabitants of the gut (Campos et al. 2004) (Table 4).

314 ***Probiotic properties***

315 Numerous characteristics should be taken into account when selecting a probiotic strain.
316 Some examples of desirable characteristics for a probiotic strain include the ability to survive
317 and retain viability in acid and bile concentrations, simulating the harsh environment of GIT
318 (Charteris et al. 1998; Rehaiem et al. 2014). A probiotic strain should have the ability for
319 producing antimicrobial substances but be devoid of acquired antibiotic resistance genes
320 (Charteris et al. 1998; Rehaiem et al. 2014; Fao et al. 2002; Ganguly et al. 2011). As stated
321 earlier the strain 17OM39 and marketed probiotic strains T110 were devoid of any clinically
322 relevant antibiotic resistance gene while all the strains were able to survive in GIT conditions.

323 For strains 17OM39 and T110 (marketed probiotic) we could trace complete pathways for
324 amino acid synthesis viz. valine, lysine, and methionine (Table 5). These are among the
325 essential amino acids and need to be supplied exogenously (Food and Nutrition Board 1989).
326 Vitamins such as folate and thiamine are the components of Vitamin-B and are considered as
327 essential nutrients for humans. Folate (folic acid) cannot be synthesized by human cells and
328 hence is necessary to be supplemented exogenously as it plays important role in DNA and
329 RNA synthesis and amino acid metabolism (Mahmood 2014; Ohrvik and Witthoft 2011;
330 Tuszyńska 2012). Thus, strains (T110 and 17OM39) producing such amino acids and
331 vitamins are considered beneficial for human use (Fijan 2014; Eck and Friel 2013). Genes
332 responsible for antibacterial activity (bacteriocin) specific against *Listeria* were found. Genes
333 for exopolisaccharide (EPS) and anti-oxidant (hydro-peroxidases) production were noted
334 which in-turn help the probiotic strains in establishing themselves in the gut. The non-
335 pathogenic group only had EPS gene cluster. Complete pathways for amino acid and vitamin
336 synthesis were absent in NPNP and pathogenic group. Thus certain strains contribute
337 beneficially to health as observed by the probiotic group strains.

338

339 ***Plasmids***

340 Earlier studies within *E. faecium* isolates have shown the abundance of plasmids by finding
341 1-7 number of plasmids in 88 out of 93 isolates (Rosvoll et al. 2010). Plasmids comprise a
342 substantial portion of the accessory genome and are accountable for much of antibiotic and
343 virulence traits to be acquired by the HGT (Rosvoll et al. 2010). Thus an attempt was made to
344 compare the plasmids of the strains considered in this study with respect to their virulence
345 factors, antibiotic resistance, phage regions and IS elements.

346 Of all the strains taken for this study, we could find a gene with 66% similar to the cytolysin
347 (*cyl*) gene in probiotic strain T110. Studies with the *cyl* gene have been an important
348 determinant in lethality of endocarditis (Rosvoll et al. 2010). The CARD analysis of the
349 plasmids shows the resistance to antibiotics viz vancomycin, streptothricin, erythromycin,
350 gentamicin and kanamycin by pathogenic group stains: 6E6, ATCC700221, Aus0085, and
351 E39 only (Table S4). No phage elements were associated with the probiotic strain T110,
352 while the non-pathogenic strain (NRRL B-2354) and pathogenic strains (ATCC 700221,
353 Aus0085, DO, and E39) harboured incomplete or complete prophage. The list of IS elements
354 found in the plasmids of strains is summarised in the TableS5.

355 ***Delineating Probiotic, Non-pathogenic and Pathogenic strains***

356 Multi-Locus Sequence Analysis (MLSA) based phylogeny could not distinguish between
357 pathogenic and non-pathogenic strains of *E. faecium* (Homan et al. 2002), but this could be
358 achieved on the basis of the core genome SNP based phylogeny (Sankarasubramanian et al.
359 2016; Heydari et al. 2013; Chen et al. 2013; Vliet and Kusters 2015). Thus, concatenated
360 1935 core genes were used to construct a phylogenetic tree of the 10 strains along with
361 *Enterococcus faecalis* symbioflor as an out-group. Phylogenetic reconstruction by using
362 Maximum likelihood method separated 10 strains in 3 distinct clusters as the three groups

considered (bootstrap >90) (Figure 7). We found no clustering based on the source of isolation, while strain 17OM39 was closely related to the probiotic strain T110. The same observation was made when repeated with the pan-genome (data not shown).

PCA plot was generated using Euclidean distances based on the prevalence of genes in antimicrobial resistance, survival in GIT and probiotic properties. The PCA plot representing the differences in property harboured between the groups is as shown in Figure 8. The BLAST Atlas was generated with the help of GVIEW server with strain Aus0004 as the reference genome (Figure 9). Of the strains, Aus0085 exhibited the highest relatedness to the reference strain. There were variable regions identified among non-pathogenic, pathogenic and probiotic group illustrating their dissimilarity in genome content. Pathogenic Island (2812458-2878042 and 1860143-1894650 bp) consist of majorly virulence-associated genes, IS elements transposes and integrase and antibiotic resistance-related genes. It also has vancomycin resistance gene cluster and presence of *esp* gene which correlates with the previous studies (Shankar et al. 2002; McBride et al. 2009). However, several phages and transposon-related loci from the reference strain appeared to be absent in marketed probiotic T110 and 17OM39 strains. This observation further supports their distinct segregation into independent clades.

CONCLUSIONS

This study has provided a valuable insight into genome-based investigations and has refined our knowledge on the genomic diversity of probiotic, non-pathogenic and pathogenic strains of *E. faecium*. The analysis has helped us to define core and accessory genome and also to understand the genomic relationships within them. The genomic features responsible for survival in GI tract, antibiotic resistance, virulence factors were known. We also have highlighted an abundance of mobile elements including prophage, insertion sequence

elements, genomic islands, and plasmids. Moreover, the analysis of intrinsic and acquired properties helped us to know the inherent probiotic properties of strain 17OM39.

MATERIAL AND METHODS

Bacterial sequences and strains

Whole Genome Sequence of *E. faecium* was retrieved from NCBI genomes and a total of ten strains were used in this study. All the genomes were RAST annotated (Overbeek et al. 2014).

Comparative analysis

Comparative analysis of ten whole genome sequences of *Enterococcus faecium* was done by an ultra-fast bacterial pan-genome analysis pipeline (BPGA) (Chaudhari et al. 2016) which performs GC content analysis, pan-genome profile analysis along with sequence extraction and phylogenetic analysis. Furthermore, the genome was investigated for the presence of putative virulence genes using Virulence Factor of Bacterial Pathogens Database (VFDB)(Chen et al. 2005). Screening of probiotic genes was done by performing BLAST of probiotic genes to the genome by online NCBI's BLASTX tool (Altschul et al. 1990). Comprehensive Antibiotic resistance Database (CARD) was used for analysis of antibiotic resistance (McArthur et al. 2013). Presence of CRISPR repeats was predicted using the CRISPRFinder tools (Grissa et al. 2008). PHASTER: a rapid identification and annotation of prophage sequences within bacterial genomes were used for identification of prophages within the genome (Arndt et al. 2016). Bacterial insertion elements (ISs) were identified by ISfinder (Siguier et al. 2006). Horizontal gene transfer was detected by genomic island tool: Islandviewer (Langille and Brinkman 2009; Dhillon et al. 2015). The clustering and annotation of protein sequences were done with the help of orthoMCL (Li et al. 2003). COG analysis was done with the help of webMGA server (Wu et al. 2011). STAMP software was

412 used to generate PCA plot (Parks et al. 2014). A blast atlas was generated with the help of
413 GVIEW Server (https://server.gview.ca/) (Petkau et al. 2010).

414 **List of abbreviations:**

415 Human gastrointestinal tract: GIT; coding DNA sequence: CDS; horizontal gene transfer:
416 HGT; Comprehensive Antibiotic Resistance Database: CARD; Mobile genetic elements:
417 MGEs; Insertion sequences: ISs; genomic island: GI; phosphotransferase system: PTS;
418 exopolysaccharide: EPS; ultra-fast bacterial pan-genome analysis pipeline: BPGA; Virulence
419 Factor of Bacterial Pathogens Database: VFDB

420 **DECLARATIONS**

421 **Consent for publication:** Not Applicable

422 **Ethics approval and consent to participate:** Not Applicable

423 **Availability of data and material:** Not Applicable

424 **Competing interests:** None

425 **Funding:** This work was supported by the Department of Biotechnology, Government of
426 India, under Grant National Centre for Microbial Resources (NCMR)
427 (BT/Coord.II/01/03/2016).

428 **Authors' contributions:** VG, DD, SP: Designed the concept; VG, MG, DD, YM: Execution
429 of the concept; DD, SP: Guide to the concept, BM, OP, YS, YM: checking and intellectual
430 inputs to the paper.

431 **Acknowledgements:** Additional financial support was provided by the Department of
432 Biotechnology, Basaveshwar Engineering College, Bagalkot, under TEQUIP Grant to Bharati
433 S. Meti.

434

435

436 **REFERENCES**

437 Abeijón MC, Medina RB, Katz MB, González SN. 2006. Technological properties of
438 Enterococcus faecium isolated from ewe ' s milk and cheese with importance for flavour
439 development. *Can J Microbiol* **52**: 237–245.

440 Ahmadova A, Todorov SD, Choiset Y, Rabesona H, Mirhadi Zadi T, Kuliyev A, Franco
441 BDG de M, Chobert JM, Haertlé T. 2013. Evaluation of antimicrobial activity, probiotic
442 properties and safety of wild strain Enterococcus faecium AQ71 isolated from
443 Azerbaijani Motal cheese. *Food Control* **30**: 631–641.
444 http://dx.doi.org/10.1016/j.foodcont.2012.08.009.

445 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search
446 tool. *J Mol Biol* **215**: 403–410.
447 https://www.sciencedirect.com/science/article/pii/S0022283605803602?via%3Dihub
448 (Accessed March 26, 2018).

449 Araújo TF, Ferreira CL de LF. 2013. The genus enterococcus as probiotic: Safety concerns.
450 *Brazilian Arch Biol Technol* **56**: 457–466.

451 Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a
452 better, faster version of the PHAST phage search tool. *Nucleic Acids Res* **44**: W16–W21.

453 Backman J, Shackleton J, Kent V, Prell L, Hays JD, Denton N, Prentice M. 1983. Shiga-Like
454 Toxin-Converting Phages from Eschenrchia coli. *Shiga-Like Toxin-Converting Phage*
455 *from Escherichia coli Strains That Cause Hemorrhagic Colitis or Infant Diarrhea* **226**:
456 694–696.

457 Bakshi U, Sarkar M, Paul S, Dutta C. 2016. Assessment of virulence potential of
458 uncharacterized Enterococcus faecalis strains using pan genomic approach –

459 Identification of pathogen–specific and habitat-specific genes. *Sci Rep* **6**: 38648.
460 http://www.nature.com/articles/srep38648.

461 Baldassarri L, Bertuccini L, Ammendolia MG, Gherardi G, Creti R. 2001. Variant esp gene
462 in Enterococcus faecium. *Lancet* **357**: 1802.

463 Banwo K, Sanni A, Tan H. 2013. Technological properties and probiotic potential of
464 Enterococcus faecium strains isolated from cow milk. *J Appl Microbiol* **114**: 229–241.

465 Begley M, Hill C, Gahan CGM. 2006. Bile Salt Hydrolase Activity in Probiotics Bile Salt
466 Hydrolase Activity in Probiotics. *Appl Environ Microbiol* **72**: 1729–1738.

467 Bender JK, Fiedler S, Klare I, Werner G. 2015. Complete Genome Sequence of the Gut
468 Commensal and Laboratory Strain Enterococcus faecium 64/3. *Genome Announc* **3**: 53–
469 60.
470 http://www.ncbi.nlm.nih.gov/pubmed/26586871%5Cnhttp://www.pubmedcentral.nih.go
471 v/articlerender.fcgi?artid=PMC4653773.

472 Benyacoub J, Perez PF, Rochat F, Saudan KY, Reuteler G, Antille N, Humen M, De Antoni
473 GL, Cavadini C, Blum S, et al. 2005. Enterococcus faecium SF68 enhances the immune
474 response to Giardia intestinalis in mice. *J Nutr* **135**: 1171–1176.
475 http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citat
476 ion&list_uids=15867299.

477 Beukers AG, Zaheer R, Goji N, Amoako KK, Chaves A V., Ward MP, McAllister TA. 2017.
478 Comparative genomics of Enterococcus spp. isolated from bovine feces. *BMC Microbiol*
479 **17**: 52. http://bmcmicrobiol.biomedcentral.com/articles/10.1186/s12866-017-0962-1.

480 Bonacina J, Suárez N, Hormigo R, Fadda S, Lechner M, Saavedra L. 2017. A genomic view
481 of food-related and probiotic Enterococcus strains. *DNA Res* **24**: 11–24.

482 Byappanahalli MN, Nevers MB, Korajkic A, Staley ZR, Harwood VJ. 2012. Enterococci in
483 the Environment. *Microbiol Mol Biol Rev* **76**: 685–706.
484 http://mmbr.asm.org/cgi/doi/10.1128/MMBR.00023-12.

485 Campos E, Campos E, Baldoma L, Baldoma L, Aguilar J, Aguilar J, Badia J, Badia J. 2004.
486 Regulation of Expression of the Divergent. *Society* **186**: 1720–1728.

487 Charteris WP, Kelly PM, Morelli L, Collins JK. 1998. Development and application of an in
488 vitro methodology to determine the transit tolerance of potentially probiotic
489 Lactobacillus and Bifidobacterium species in the upper human gastrointestinal tract. *J*
490 *Appl Microbiol* **84**: 759–768.

491 Chaudhari NM, Gupta VK, Dutta C. 2016. BPGA- an ultra-fast pan-genome analysis
492 pipeline. *Sci Rep* **6**: 24373. http://www.nature.com/articles/srep24373.

493 Chen C, Zhang W, Zheng H, Lan R, Wang H, Du P, Bai X, Ji S, Meng Q. 2013. Minimum
494 Core Genome Sequence Typing of Bacterial Pathogens : a. *J Clin Microbiol* **51**: 2582–
495 2591.

496 Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q. 2005. VFDB: A reference database for
497 bacterial virulence factors. *Nucleic Acids Res* **33**: 325–328.

498 Comerlato CB, de Resende MCC, Caierão J, d'Azevedo PA. 2013. Presence of virulence
499 factors in Enterococcus faecalis and Enterococcus faecium susceptible and resistant to
500 vancomycin. *Mem Inst Oswaldo Cruz* **108**: 590–595.

501 Coque TM, Willems RJL, Fortún J, Diz S, Loza E, Cantón R. 2005. Population Structure of
502 Enterococcus faecium Causing Bacteremia in a Spanish University Hospital : Setting the
503 Scene for a Future Increase in Vancomycin Resistance ? Population Structure of
504 Enterococcus faecium Causing Bacteremia in a Spanish University H. *Antimicrob*

505 *Agents Chemother* **49**: 2693–2700.

506 Dahlén G, Blomqvist S, Almståhl A, Carlén A. 2012. Virulence factors and antibiotic
507 susceptibility in enterococci isolated from oral mucosal and deep infections. *J Oral*
508 *Microbiol* **4**: 1–7.

509 Davis BM, Waldor MK. 2003. Filamentous phages linked to virulence of Vibrio cholerae.
510 *Curr Opin Microbiol* **6**: 35–42.

511 Devriese L a., Van De Kerckhove a., Kilpper-Balz R, Schleifer KH. 1987. Characterization
512 and Identification of Enterococcus Species Isolated from the Intestines of Animals. *Int J*
513 *Syst Bacteriol* **37**: 257–259.

514 Dhillon BK, Laird MR, Shay JA, Winsor GL, Lo R, Nizam F, Pereira SK, Waglechner N,
515 McArthur AG, Langille MGI, et al. 2015. IslandViewer 3: More flexible, interactive
516 genomic island discovery, visualization and analysis. *Nucleic Acids Res* **43**: W104–
517 W108.

518 Dobrindt U, Hochhut B, Hentschel U, Hacker J. 2004. Genomic islands in pathogenic and
519 environmental microorganisms. *Nat Rev Microbiol* **2**: 414–424.
520 http://www.nature.com/doifinder/10.1038/nrmicro884.

521 dos Santos KMO, Vieira ADS, Salles HO, Oliveira J da S, Rocha CRC, Borges M de F,
522 Bruno LM, Franco BDG de M, Todorov SD. 2015. Safety, beneficial and technological
523 properties of Enterococcus faecium isolated from Brazilian cheeses. *Brazilian J*
524 *Microbiol* **46**: 237–249.

525 Eck P, Friel J. 2013. Vitamins & Minerals Editorial. *Vitam Miner* **3**: 1–2.

526 Fao J, Working WHO, Report G, Guidelines D, London F. 2002. Guidelines for the

527 Evaluation of Probiotics in Food. 1–11.

528 Fijan S. 2014. Microorganisms with Claimed Probiotic Properties: An Overview of Recent
529 Literature. *Int J Environ Res Public Heal Int J Environ Res Public Heal Int J Environ*
530 *Res Public Heal* **11**: 4745–4767. www.mdpi.com/journal/ijerph (Accessed November 3,
531 2016).

532 Finlay BB. 1997. Common Themes in Microbial Pathogenicity Revisited. *Microbiol Mol Biol*
533 *Rev* **61**: 136–169.

534 Food and Nutrition Board NRC. 1989. *Recommended Dietary Allowances*.

535 Franz CMAP, Stiles ME, Schleifer KH, Holzapfel WH. 2003. Enterococci in foods - A
536 conundrum for food safety. *Int J Food Microbiol* **88**: 105–122.

537 FREEMAN VJ. 1951. Studies on the virulence of bacteriophage-infected strains of
538 Corynebacterium diphtheriae. *J Bacteriol* **61**: 675–688.

539 Galimand M, Schmitt E, Panvert M, Douthwaite S, Mechulam Y, Courvalin P. 2011. Intrinsic
540 resistance to aminoglycosides in Enterococcus faecium is conferred by the 16S rRNA m
541 5 C1404-specific methyltransferase EfmM. *RNA* **17**: 251–262.

542 Ganguly NK, Bhattacharya SK, Sesikeran B, Nair GB, Ramakrishna BS, Sachdev HPS,
543 Batish VK, Kanagasabapathy AS, Muthuswamy V, Kathuria SC, et al. 2011. ICMR-
544 DBT Guidelines for evaluation of probiotics in food. *Indian J Med Res* **134**: 22–25.

545 Geldart K, Kaznessis YN. 2017. Characterization of class IIa bacteriocin resistance in
546 Enterococcus faecium. *Antimicrob Agents Chemother* **61**: 1–17.

547 Ghattargi VC, Nimonkar YS, Burse SA, Davray D, Kumbhare S V. 2018. Genomic and
548 physiological analyses of an indigenous strain , Enterococcus faecium 17OM39. *Funct*

549 *Integr Genomics*.

550 Giraffa G. 2002. Enterococci from foods. *FEMS Microbiol Rev* **26**: 163–171.

551 Gold HS. 2001. Vancomycin-Resistant Enterococci : Mechanisms and Clinical Observations.
552 *Antimicrob Resist* **33**: 214–19.

553 Grissa I, Vergnaud G, Pourcel C. 2008. CRISPRcompar: a website to compare clustered
554 regularly interspaced short palindromic repeats. *Nucleic Acids Res* **36**: 52–57.

555 Guggenbichler JP, Assadian O, Boeswald M, Kramer A. 2011. Incidence and clinical
556 implication of nosocomial infections associated with implantable biomaterials -
557 catheters, ventilator-associated pneumonia, urinary tract infections. *GMS Krankenhhyg*
558 *Interdiszip* **6**: Doc18.
559 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3252661&tool=pmcentrez&
560 rendertype=abstract.

561 Hassanzadazar H, Ehsani A, Mardani K. 2014. Antibacterial activity of Enterococcus
562 faecium derived from Koopeh cheese against Listeria monocytogenes in probiotic ultra-
563 filtrated cheese. *Vet Res Forum* **5**: 169–175. http://www.urmia.ac.ir/vrf/Shared
564 Documents/pdf/vol-5 no-3/169-175-1171.pdf.

565 Hegstad K, Mikalsen T, Coque TM, Werner G, Sundsfjord A. 2010a. Mobile genetic
566 elements and their contribution to the emergence of antimicrobial resistant Enterococcus
567 faecalis and Enterococcus faecium. *Clin Microbiol Infect* **16**: 541–554.

568 Hegstad K, Mikalsen T, Coque TM, Werner G, Sundsfjord A. 2010b. Mobile genetic
569 elements and their contribution to the emergence of antimicrobial resistant Enterococcus
570 faecalis and Enterococcus faecium. *Clin Microbiol Infect* **16**: 541–554.
571 http://dx.doi.org/10.1111/j.1469-0691.2010.03226.x.

572 Heikens E, Bonten MJM, Willems RJL. 2007. Enterococcal surface protein esp is important
573 for biofilm formation of Enterococcus faecium E1162. *J Bacteriol* **189**: 8233–8240.

574 Heydari H, Wee WY, Lokanathan N, Hari R, Yusoff AM. 2013. MabsBase : A
575 Mycobacterium abscessus Genome and Annotation Database. *PLoS One* **8**: 2–6.

576 Hollenbeck BL, Rice LB. 2012. Intrinsic and acquired resistance mechanisms in
577 enterococcus. *Virulence* **3**: 421–569.
578 http://www.tandfonline.com/doi/abs/10.4161/viru.21282.

579 Homan WL, Tribe D, Poznanski S, Li M, Hogg G, Spalburg E, Embden JD a Van, Willems
580 JL, Willems RJL. 2002. Multilocus Sequence Typing Scheme for Enterococcus faecium
581 Multilocus Sequence Typing Scheme for Enterococcus faecium. *J Clin Microbiol* **40**:
582 1963–71.

583 Huh JY, Lee WG, Lee K, Shin WS. 2004. Distribution of Insertion Sequences Associated
584 with Tn 1546 -Like Elements among Enterococcus faecium Isolates from Patients in
585 Korea Distribution of Insertion Sequences Associated with Tn 1546 -Like Elements
586 among Enterococcus faecium Isolates from Patie. *J Clin Microbiol* **42**: 1897–1902.

587 Huys G, Botteldoorn N, Delvigne F, De Vuyst L, Heyndrickx M, Pot B, Dubois JJ, Daube G.
588 2013. Microbial characterization of probiotics-Advisory report of the Working Group
589 "8651 Probiotics" of the Belgian Superior Health Council (SHC). *Mol Nutr Food Res*
590 **57**: 1479–1504.

591 İspirli H, Demirbaş F, Dertli E. 2015. Characterization of functional properties of
592 Enterococcus faecium strains isolated from human gut. *Can J Microbiol* **61**: 1–10.
593 http://www.ncbi.nlm.nih.gov/pubmed/26485327.

594 Jiang X, Ellabaan MMH, Charusanti P, Munck C, Blin K, Tong Y, Weber T, Sommer MOA,

595 Lee SY. 2017. Dissemination of antibiotic resistance genes from antibiotic producers to
596 pathogens. *Nat Commun* **8**: 15784.
597 http://www.nature.com/doifinder/10.1038/ncomms15784.

598 Jones B V., Begley M, Hill C, Gahan CGM, Marchesi JR. 2008. Functional and comparative
599 metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. *Proc*
600 *Natl Acad Sci U S A* **105**: 13580–5.
601 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2533232&tool=pmcentrez&
602 rendertype=abstract.

603 Jong-Hoon Lee, Donghun Shin, Bitnara Lee, Hyundong Lee, Inhyung Lee D-WJ. 2017.
604 Genetic Diversity and Antibiotic Resistance of Enterococcus faecalis Isolates from
605 Traditional Korean Fermented Soybean Foods. *J Microbiol Biotechnol* **27**: 916–924.

606 Juhas M, Van Der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW. 2009.
607 Genomic islands: Tools of bacterial horizontal gene transfer and evolution. *FEMS*
608 *Microbiol Rev* **33**: 376–393.

609 Kačániová M, Kmeť V, Čcuboň J. 2006. Effect of Enterococcus faeciumon the digestive tract
610 of poultry as a probiotic. *Turkish J Vet Anim Sci* **30**: 291–298.

611 Kajihara T, Nakamura S, Iwanaga N, Oshima K, Takazono T, Miyazaki T, Izumikawa K,
612 Yanagihara K, Kohno N, Kohno S. 2015. Clinical characteristics and risk factors of
613 enterococcal infections in Nagasaki, Japan: a retrospective study. *BMC Infect Dis* **15**:
614 426. http://dx.doi.org/10.1186/s12879-015-1175-6.

615 Kaplan T. 2014. The Role of Horizontal Gene Transfer in Antibiotic Resistance. *Eukaryon*
616 **10**: 80–81.

617 King KY, Horenstein JA, Caparon MG. 2000. Aerotolerance and peroxide resistance in

618 peroxidase and PerR mutants of Streptococcus pyogenes. *J Bacteriol* **182**: 5290–5299.

619 Kirjavainen P V, Saxelin M, Rautelin H, Valtonen V, Salminen SJ, Ouwehand AC. 2001.
620 Good adhesion properties of probiotics : a potential risk for bacteremia ? *FEMS Immunol*
621 *Med Microbiol* **31**: 35–39.

622 Kopit LM, Kim EB, Siezen RJ, Harris LJ, Marco ML. 2014. Safety of the surrogate
623 microorganism Enterococcus faecium NRRL B-2354 for use in thermal process
624 validation. *Appl Environ Microbiol* **80**: 1899–1909.

625 Kreuzer S, MacHnowska P, Amus J, Sieber M, Pieper R, Schmidt MF, Brockmann GA,
626 Scharek-Tedin L, Johne R. 2012. Feeding of the probiotic bacterium Enterococcus
627 faecium NCIMB 10415 differentially affects shedding of enteric viruses in pigs. *Vet Res*
628 **43**: 1–12.

629 Lam MMC, Seemann T, Bulach DM, Gladman SL, Chen H, Haring V, Moore RJ, Ballard S,
630 Grayson ML, Johnson PDR, et al. 2012. Comparative analysis of the first complete
631 Enterococcus faecium genome. *J Bacteriol* **194**: 2334–2341.

632 Langille MGI, Brinkman FSL. 2009. IslandViewer: An integrated interface for computational
633 identification and visualization of genomic islands. *Bioinformatics* **25**: 664–665.

634 Li L, Stoeckert CJJ, Roos DS. 2003. OrthoMCL: Identification of Ortholog Groups for
635 Eukaryotic Genomes -- Li et al. 13 (9): 2178 -- Genome Research. *Genome Res* **13**:
636 2178–2189. http://genome.cshlp.org/cgi/content/full/13/9/2178.

637 López M, Sáenz Y, Álvarez-Martínez MJ, Marco F, Robredo B, Rojo-Bezares B, Ruiz-Larrea
638 F, Zarazaga M, Torres C. 2010. Tn1546 structures and multilocus sequence typing of
639 vanA-containing enterococci of animal, human and food origin. *J Antimicrob*
640 *Chemother* **65**: 1570–1575.

641 Mahmood L. 2014. The metabolic processes of folic acid and Vitamin B12 deficiency. *J Heal*
642 *Res Rev* **1**: 5. http://www.jhrr.org/text.asp?2014/1/1/5/143318.

643 Manero A, Blanch AR. 1999. Identification of Enterococcus spp . with a Biochemical Key.
644 *Appl Environ Microbiol* **65**: 4425–4430.

645 Marothi Y a, Agnihotri H, Dubey D. 2005. Enterococcal resistance: an overview. *Indian J*
646 *Med Microbiol* **23**: 214–219.

647 Matos RC, Lapaque N, Rigottier-Gois L, Debarbieux L, Meylheuc T, Gonzalez-Zorn B,
648 Repoila F, de Lopes MF, Serror P. 2013. Enterococcus faecalis Prophage Dynamics and
649 Contributions to Pathogenic Traits. *PLoS Genet* **9**.

650 McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova
651 MJ, De Pascale G, Ejim L, et al. 2013. The comprehensive antibiotic resistance
652 database. *Antimicrob Agents Chemother* **57**: 3348–3357.

653 McBride SM, Coburn PS, Baghdayan AS, Willems RJL, Grande MJ, Shankar N, Gilmore
654 MS. 2009. Genetic variation and evolution of the pathogenicity island of Enterococcus
655 faecalis. *J Bacteriol* **191**: 3392–3402.

656 McKenney PT, Ling L, Wang G, Mane S, Pamer EG. 2016. Complete Genome Sequence of
657 *Enterococcus faecium* ATCC 700221. *Genome Announc* **4**: e00386-16.
658 http://genomea.asm.org/lookup/doi/10.1128/genomeA.00386-16.

659 Mikalsen T, Pedersen T, Willems R, Coque TM, Werner G, Sadowy E, van Schaik W, Jensen
660 LB, Sundsfjord A, Hegstad K. 2015. Investigating the mobilome in clinically important
661 lineages of Enterococcus faecium and Enterococcus faecalis. *BMC Genomics* **16**: 282.
662 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4438569&tool=pmcentrez&
663 rendertype=abstract.

664 Miller WR, Munita JM, Arias CA. 2014. Mechanisms of antibiotic resistance in enterococci.
665 *Expert Rev Anti Infect Ther* **12**: 1221–36.
666 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4433168&tool=pmcentrez&
667 rendertype=abstract.

668 Minelli EB, Benini A. 2008. Relationship between number of bacteria and their probiotic
669 effects. *Microb Ecol Health Dis* **20**: 180–183.

670 Monot M, Boursaux-Eude C, Thibonnier M, Vallenet D, Moszer I, Medigue C, Martin-
671 Verstraete I, Dupuy B. 2011. Reannotation of the genome sequence of Clostridium
672 difficile strain 630. *J Med Microbiol* **60**: 1193–1199.

673 Natarajan P, Parani M. 2015. First complete genome sequence of a probiotic enterococcus
674 faecium strain t-110 and its comparative genome analysis with pathogenic and non-
675 pathogenic enterococcus faecium genomes. *J Genet Genomics* **42**: 43–46.
676 http://dx.doi.org/10.1016/j.jgg.2014.07.002.

677 O'Driscoll T, Crank CW. 2015. Vancomycin-resistant enterococcal infections:
678 Epidemiology, clinical manifestations, and optimal management. *Infect Drug Resist* **8**:
679 217–230.

680 Ohrvik VE, Witthoft CM. 2011. Human folate bioavailability. *Nutrients* **3**: 475–490.

681 Ouwehand AC, Kirjavainen P V, Shortt C, Salminen S. 1999. Probiotics : mechanisms and
682 established effects. *Int Dairy J* **9**: 43–52.

683 Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S,
684 Parrello B, Shukla M, et al. 2014. The SEED and the Rapid Annotation of microbial
685 genomes using Subsystems Technology (RAST). *Nucleic Acids Res* **42**: 206–214.

686 Palmer KL, Gilmore MS. 2010. Multidrug-resistant enterococci lack CRISPR-cas. *MBio* **1**:
687 1–10.

688 Palmer KL, Godfrey P, Griggs A, Kos VN, Zucker J, Desjardins C, Cerqueira G, Gevers D,
689 Walker S, Wortman J, et al. 2012. Comparative Genomics of Enterococci : Comparative
690 Genomics of Enterococci : Variation in. *MBio* **3**: 1–11.

691 Parks DH, Tyson GW, Hugenholtz P, Beiko RG. 2014. STAMP: Statistical analysis of
692 taxonomic and functional profiles. *Bioinformatics* **30**: 3123–3124.

693 Petkau A, Stuart-Edwards M, Stothard P, van Domselaar G. 2010. Interactive microbial
694 genome visualization with GView. *Bioinformatics* **26**: 3125–3126.

695 Postma PW, Lengeler JW, Jacobson GR. 1993. Phosphoenolpyruvate: Carbohydrate
696 phosphotransferase systems of bacteria. *Microbiol Rev* **57**: 543–594.
697 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=372926&tool=pmcentrez&r
698 endertype=abstract.

699 Qin X, Galloway-Pĕa JR, Sillanpaa J, Roh JH, Nallapareddy SR, Chowdhury S, Bourgogne
700 A, Choudhury T, Muzny DM, Buhay CJ, et al. 2012. Complete genome sequence of
701 Enterococcus faecium strain TX16 and comparative genomic analysis of Enterococcus
702 faecium genomes. *BMC Microbiol* **12**.

703 Ramadhan a a, Hegedus E. 2005. Biofilm formation and esp gene carriage in enterococci. *J*
704 *Clin Pathol* **58**: 685–686.

705 Rao S, Ihara Y, Sukegawa S, Arakawa F, Fujimura T, Murakami H, Morimatsu F. 2013.
706 Characterization of probiotic properties of Enterococcus faecium NHRD IHARA
707 isolated from porcine feces. *Biosci Biotechnol Biochem* **77**: 1967–9.
708 http://www.ncbi.nlm.nih.gov/pubmed/24018679.

709 Rasouli Pirouzian H, Hesari J, Farajnia S, Moghaddam M, Ghiassifar S. 2012. Effects of
710 Enterococcus faecalis and Enterococcus faecium, isolated from traditional Lighvan
711 cheese, on physicochemical and sensory characteristics of Iranian UF white cheese. *J*
712 *Agric Sci Technol* **14**: 1023–1034.

713 Rehaiem A, Belgacem Z Ben, Edalatian MR, Mart??nez B, Rodr??guez A, Manai M, Guerra
714 NP. 2014. Assessment of potential probiotic properties and multiple bacteriocin
715 encoding-genes of the technological performing strain Enterococcus faecium MMRA.
716 *Food Control* **37**: 343–350. http://dx.doi.org/10.1016/j.foodcont.2013.09.044.

717 Rosvoll TCS, Pedersen T, Sletvold H, Johnsen PJ, Sollid JE, Simonsen GS, Jensen LB,
718 Nielsen KM, Sundsfjord A. 2010. PCR-based plasmid typing in Enterococcus faecium
719 strains reveals widely distributed pRE25-, pRUM-, pIP501- and pHTβ-related replicons
720 associated with glycopeptide resistance and stabilizing toxin-antitoxin systems. *FEMS*
721 *Immunol Med Microbiol* **58**: 254–268.

722 Sangvik M, Littauer P, Simonsen GS, Sundsfjord A, Dahl KH. 2005. mef(A), mef(E) and a
723 new mef allele in macrolide-resistant Streptococcus spp. isolates from Norway. *J*
724 *Antimicrob Chemother* **56**: 841–846.

725 Sankarasubramanian J, Vishnu US, Gunasekaran P, Rajendhran J. 2016. A genome-wide
726 SNP-based phylogenetic analysis distinguishes different biovars of Brucella suis. *Infect* ,
727 *Genet Evol* **41**: 213–217. http://dx.doi.org/10.1016/j.meegid.2016.04.012.

728 Santagati M, Campanile F, Stefani S. 2012. Genomic diversification of enterococci in hosts:
729 The role of the mobilome. *Front Microbiol* **3**: 1–9.

730 Satish Kumar R, Kanmani P, Yuvaraj N, Paari KA, Pattukumar V, Arul V. 2011. Purification
731 and characterization of enterocin MC13 produced by a potential aquaculture probiont

Enterococcus faecium MC13 isolated from the gut of *Mugil cephalus*. *Can J Microbiol* **57**: 993–1001. http://www.nrcresearchpress.com/doi/abs/10.1139/w11-092.

Shankar N, Baghdayan a S, Gilmore MS. 2002. Modulation of virulence within a pathogenicity island in vancomycin- resistant Enterococcus faecalis. *Nature* **417**: 746–50. http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?db=m&form=6&dopt=r&uid=12066186.

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

Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* **34**: D32-6. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1347377&tool=pmcentrez&rendertype=abstract%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/16381877%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1347377.

Stadler C, Teuber M. 2002. The macrolide efflux genetic assembly of Streptococcus pneumoniae is present in erythromycin-resistant Streptococcus salivarius [5]. *Antimicrob Agents Chemother* **46**: 3690–3691.

Toledo-Arana A, Valle J, Solano C, Arrizubieta MJ, Cucarella C, Lamata M, Amorena B, Leiva J, Penades JR, Lasa I. 2001. The Enterococcal Surface Protein, Esp, Is Involved in *Enterococcus faecalis* Biofilm Formation. *Appl Environ Microbiol* **67**: 4538–4545.

Tong J, Lu XT, Zhang JY, Sui Q, Wang R, Chen M, Wei Y. 2017. Occurrence of antibiotic resistance genes and mobile genetic elements in enterococci and genomic DNA during anaerobic digestion of pharmaceutical waste sludge with different pretreatments. *Bioresour Technol* **235**: 316–324. http://dx.doi.org/10.1016/j.biortech.2017.03.104.

755 Tuszyńska M. 2012. Folic Acid - The Occurrence and the Role in Human Nutrition. *Veg*
756 *Crop Res Bull* **76**: 43–54. http://www.degruyter.com/view/j/vcrb.2012.76.issue--
757 1/v10032-012-0003-4/v10032-012-0003-4.xml.

758 Vahjen W, Männer K. 2003. The effect of a probiotic Enterococcus faecium product in diets
759 of healthy dogs on bacteriological counts of Salmonella spp., Campylobacter spp. and
760 Clostridium spp. in faeces. *Arch Anim Nutr* **57**: 229–233.

761 VanRenterghem B. 2012. 126th AOAC Annual Meeting & Exposition. *Food Saf Mag* 38–45.

762 Vliet AHM Van, Kusters JG. 2015. Use of Alignment-Free Phylogenetics for Rapid Genome
763 Sequence- Based Typing of Helicobacter pylori Virulence Markers and Antibiotic. *J*
764 *Clin Microbiol* **53**: 2877–2888.

765 Von Wintersdorff CJH, Penders J, Van Niekerk JM, Mills ND, Majumder S, Van Alphen LB,
766 Savelkoul PHM, Wolffs PFG. 2016. Dissemination of antimicrobial resistance in
767 microbial ecosystems through horizontal gene transfer. *Front Microbiol* **7**: 1–10.

768 Wang X, Tong H, Dong X. 2014. PerR-regulated manganese ion uptake contributes to
769 oxidative stress defense in an oral streptococcus. *Appl Environ Microbiol* **80**: 2351–
770 2359.

771 Werner G, Fleige C, Geringer U, van Schaik W, Klare I, Witte W. 2011. IS element IS16 as a
772 molecular screening tool to identify hospital-associated strains of Enterococcus faecium.
773 *BMC Infect Dis* **11**: 80. http://bmcinfectdis.biomedcentral.com/articles/10.1186/1471-
774 2334-11-80.

775 Wu S, Zhu Z, Fu L, Niu B, Li W, Yooseph S, Sutton G, Rusch D, Halpern A, Williamson S,
776 et al. 2011. WebMGA: a customizable web server for fast metagenomic sequence
777 analysis. *BMC Genomics* **12**: 444.

778 http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-12-444 (Accessed
779 October 14, 2016).

780 Xu L, He SS, Li DY, Mei C, Hou XL, Jiang LS, Liu FH. 2016. Hydrogen peroxide induces
781 oxidative stress and the mitochondrial pathway of apoptosis in RAT intestinal epithelial
782 cells (IEC-6). *Mol Biol* **50**: 270–277. http://dx.doi.org/10.1134/S0026893316020266.

783

TABLES

Features	Probiotic	Non-Pathogenic		Pathogenic					
	T110	17OM39	NRRLB-2354	64/3	DO	Aus0004	Aus0085	6E6	E39	ATCC 700221
Size (mb)	2.6	2.6	2.6	2.5	2.6	2.9	2.9	2.9	2.7	2.8
GC%	38.4	38.5	37.8	38.2	37.9	38.3	37.9	37.6	37.8	37.8
Genes	2,639	2,865	2,771	2,508	2,795	2,960	3,214	3,404	3,043	3,145
CDS	2,502	2,639	2,658	2,418	2,703	2,825	2,938	3,307	2,907	2,725
Pseudo Genes	54	148	47	32	32	70	181	73	44	326
rRNAs (5S, 16S, 23S)	6, 6, 6	6, 6,6	6, 6, 6	6, 6, 6	6, 6, 6	6, 6, 6	6, 6, 6	6, 6, 6	6, 6, 6	6, 6, 6
tRNAs	65	62	48	68	62	47	76	75	70	72
Plasmid	1	-	1	-	3	3	6	2	5	3
Accession code	NZ_ CP006030.1	LWHF 00000000.1	NC_ 020207.1	NZ_ CP012522.1	NC_ 017960.1	NC_ 017022.1	NC_ 021994.1	NZ_ CP013994.1	NZ_ CP011281.1	CP 014449.1
Source	probiotic	feces	cheese	feces	blood	blood	blood	feces	blood	feces
Country	Japan	India	-	Germany	USA	Australia	Australia	USA	USA	USA
Reference	Natarajan and Parani 2015	Ghattargi etal. 2018	Kopit et al. 2014	Bender et al. 2015	Lam et al. 2012	Qin et al. 2012	Qin et al. 2012	Geldart and Kaznessis 2017	Geldart and Kaznessis 2017	McKenney et al. 2016

Table 1 General Genome Features

Table 2 Antibiotic Resistance genes found in *Enterococcus* genomes as performed by CARD analysis, where + Present and - Absent

Antibiotics	T110	17OM39	NRRL B-2354	64_3	DO	Aus0004	Aus0085	6_E6	E39	ATCC 700221
Daptomycin	-	-	-	-	-	-	-	-	+	-
Trimethoprim	-	-	-	-	-	+	+	+	+	-
Multidrug	-	-	-	-	-	+	-	-	-	-
Macrolide	-	-	-	-	-	+	-	-	-	-
Polymyxin	-	-	-	-	-	-	-	-	-	+
Tetracyline	-	-	-	-	-	+	+	+	+	-
Vancomycin	-	-	-	-	-	+	+	-	-	-

Table 3 Virulence factors found in *Enterococcus* genomes, where + Present; - Absent; * Non-functional due to presence of stop codon.

CATEGORY	GENES	T110	17OM39	NRRL B-2354	64/3	DO	Aus004	Aus0085	6E6	E39	ATCC 700221
Adherence	***acm***	*	*	*	*	+	+	+	+	+	+
	EbpA	-	-	+	+	+	+	+	+	+	+
	EbpC	-	-	+	+	+	+	+	+	+	+
	srtC	+	+	+	+	+	+	+	+	+	+
	EcbA	-	-	-	-	+	+	+	+	+	+
	efaA	+	+	+	+	+	+	+	+	+	+
	Esp	-	-	-	-	-	+	+	+	+	+
	Scm	*	*	*	*	+	+	+	+	+	+
	SgrA	-	-	-	-	+	+	+	+	+	+
Bioflim	***bopD***	*	*	*	*	+	+	+	+	+	+
Exoenzymes	***EF0818***	+	*	-	-	-	-	-	-	-	-

Table 4 Number of genes responsible for survival in GI track within *Enterococcus* genomes.

CATEGORY	GENES	T110	17OM39	NRRL B-2354	64/3	DO	Aus0004	Aus0085	6_E6	E39	ATCC 700221
Acid resistance	*LBA0995* *LBA1524* *LBA1272* *gadC* *rrp-1*	**3/5**	**3/5**	**3/5**	**3/5**	**4/5**	**4/5**	**5/5**	**4/5**	**4/5**	**3/5**
Bile resistance	*LBA1430* *clpE* *dps* *LBA1429*	**4/4**	**4/4**	**4/4**	**4/4**	**4/4**	**4/4**	**4/4**	**4/4**	**4/4**	**4/4**
Competitive	*copA* *met* *pts14C*	**3/3**	**3/3**	**2/3**	**2/3**	**2/3**	**2/3**	**2/3**	**2/3**	**2/3**	**2/3**
Adherence	*lsp* *FbpA* *ispA*	**2/3**	**2/3**	**3/3**	**3/3**	**3/3**	**3/3**	**3/3**	**3/3**	**3/3**	**3/3**
Persistence	*LJ1656* *msrB* *LJ1654* *clpC*	**4/4**	**4/4**	**4/4**	**4/4**	**4/4**	**4/4**	**4/4**	**4/4**	**4/4**	**4/4**
Bile salt hydrolase	*bsh*	**1/1**	**1/1**	**1/1**	**1/1**	**1/1**	**1/1**	**1/1**	**1/1**	**1/1**	**1/1**
Growth	*treC*	**1/1**	**1/1**	**1/1**	**1/1**	**1/1**	**1/1**	**1/1**	**1/1**	**1/1**	**1/1**
Adaptation	*Lr1265* *Lr1584*	**1/2**	**1/2**	**1/2**	**1/2**	**1/2**	**2/2**	**2/2**	**2/2**	**2/2**	**2/2**

19

20 **Table 5** Probiotic properties found in *Enterococcus* genomes, where + Present and - Absent

Properties	17OM39	T110	64/3	NRRL B-2354	DO	Aus0004	Aus0085	6_E6	E39	ATCC 700221
Anti-oxidant	+	+	-	-	-	-	-	-	-	-
Anti-bacterial	+	+	-	-	-	-	-	-	-	-
EPS	+	+	+	+	-	-	-	-	-	-
Amino-acid	valine, lysine, methionine	valine, lysine	-	-	-	-	-	-	-	-
Vitamins	Folate, Thiamine	Folate, Thiamine	-	-	-	-	-	-	-	-

21

22

FIGURES

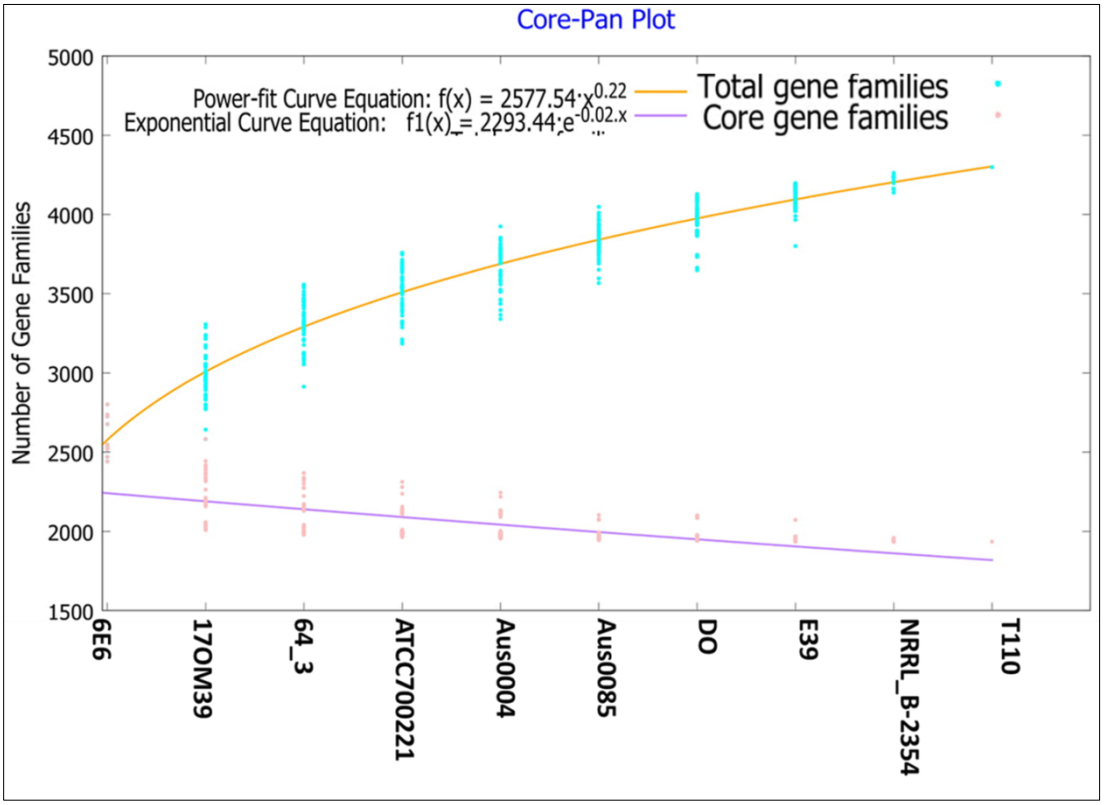

Figure 1 Core and pan genome for *E. faecium* strains. The number of shared genes is plotted as the function of number of strains (n) added sequentially. 1935 genes are shared by all 10 genomes. The orange line represents the least-squares fit to the power law function f(x)=a.x^b where a= 2577.54, b= 0.222602. The red line represents the least-squares fit to the exponential decay function f1(x)=c.e^(d.x) where c= 2293.44, d= -0.0232013.

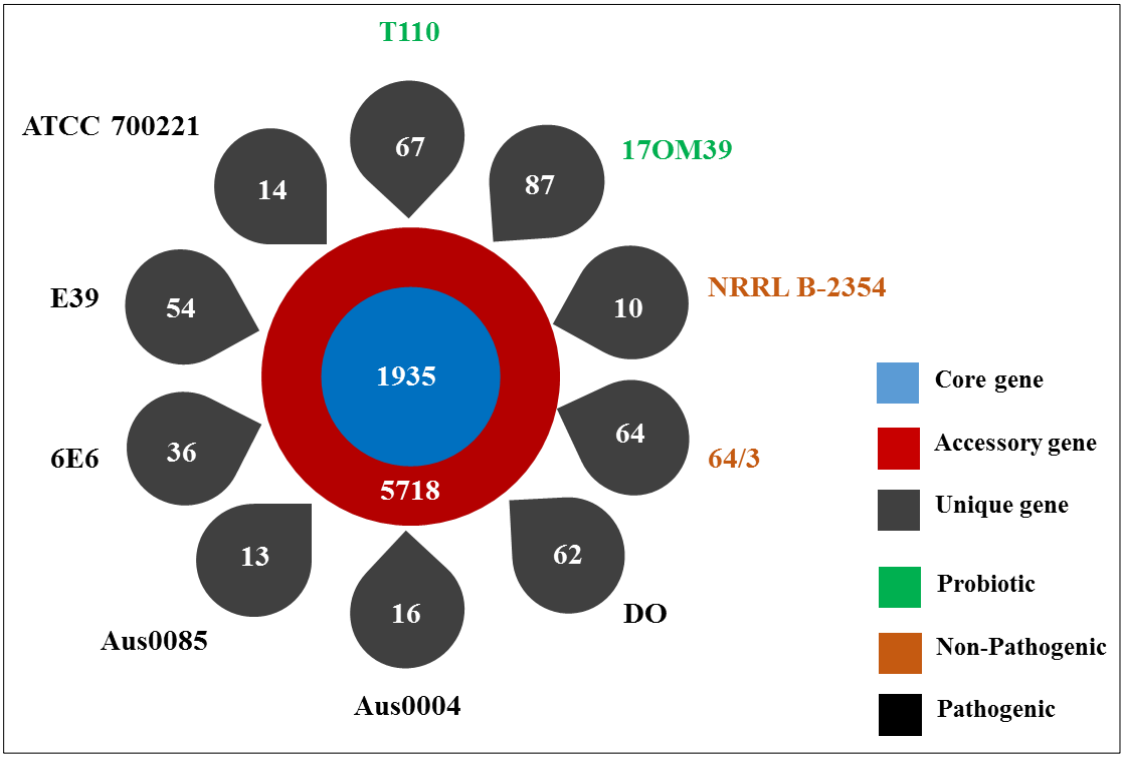

Figure 2 Number of Core, Accessory and Unique gene families of *Enterococcus* genomes. The inner circle represents the core genome consisting of 1935 genes in single copy. The outer red circle represents the accessory genome for all then ten strains adding to a sum of 5718 genes, while the outer petals represents the unique genes associated with all the strains. The strains green coloured are probiotics, brown are non-pathogenic and the black are pathogenic.

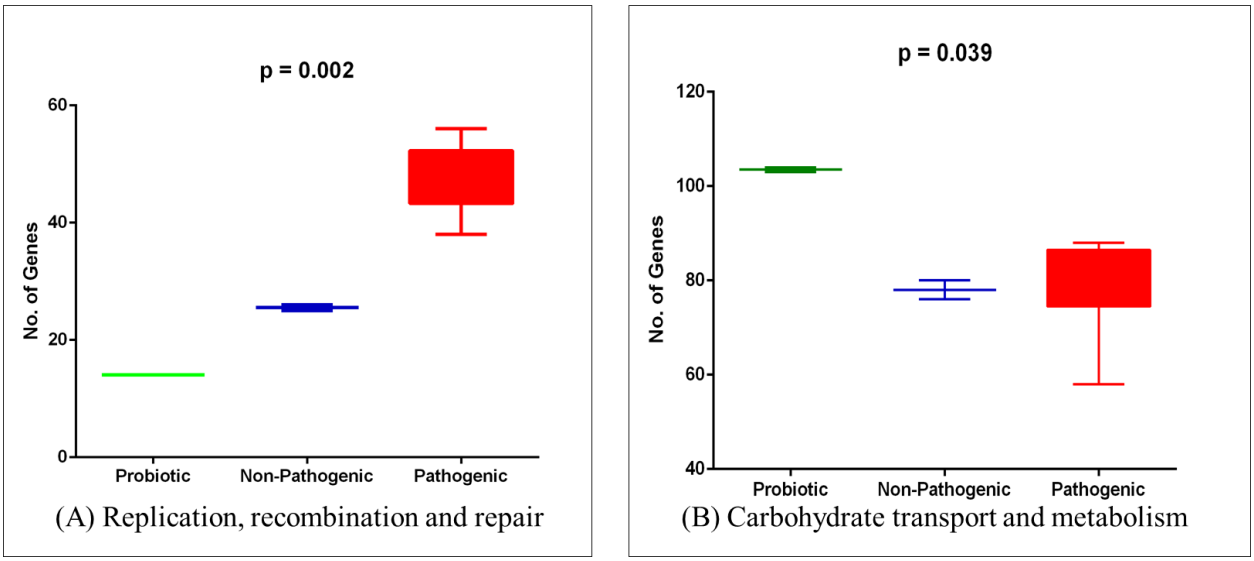

Figure 3 Showing the Significant COG's in the accession genome. (A) Replication, recombination and repair (B) Carbohydrate transport and metabolism

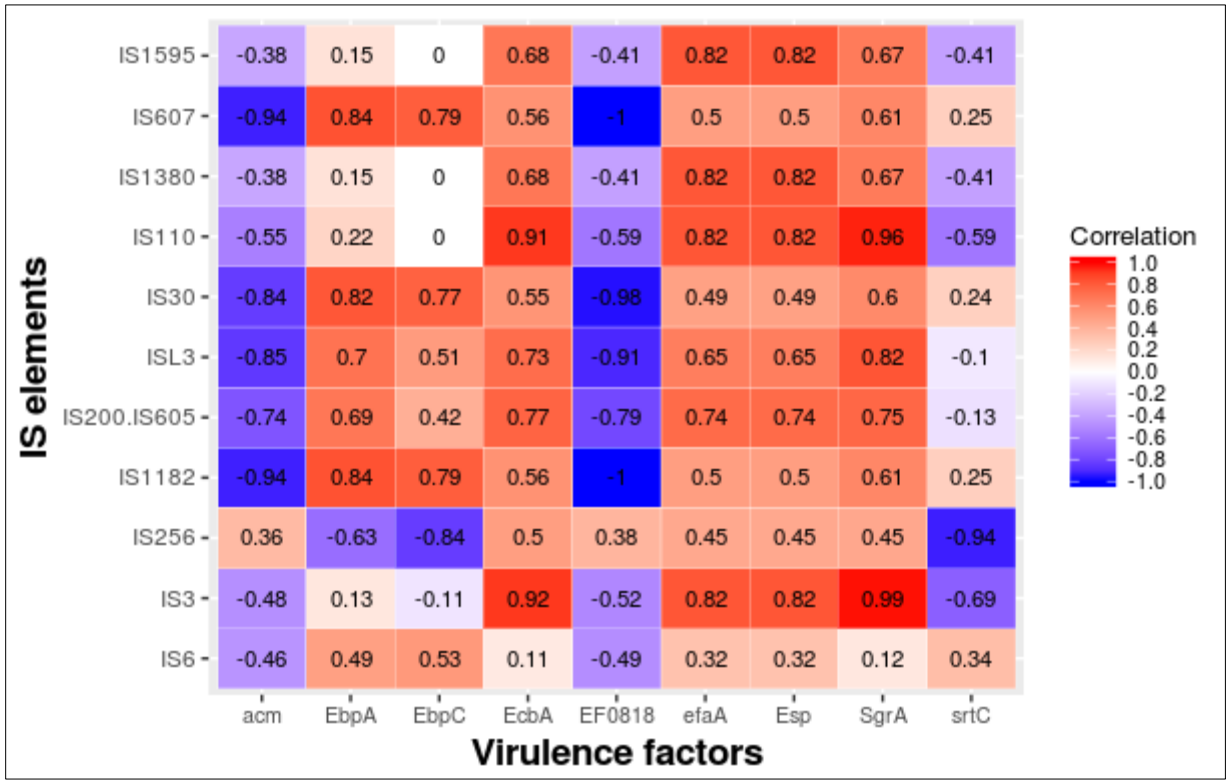

Figure 4 The heat map showing the correlation between the IS elements and virulence factors found across the genome. Red color indicated the strong positive correlation while the blue indicates negative correlation.

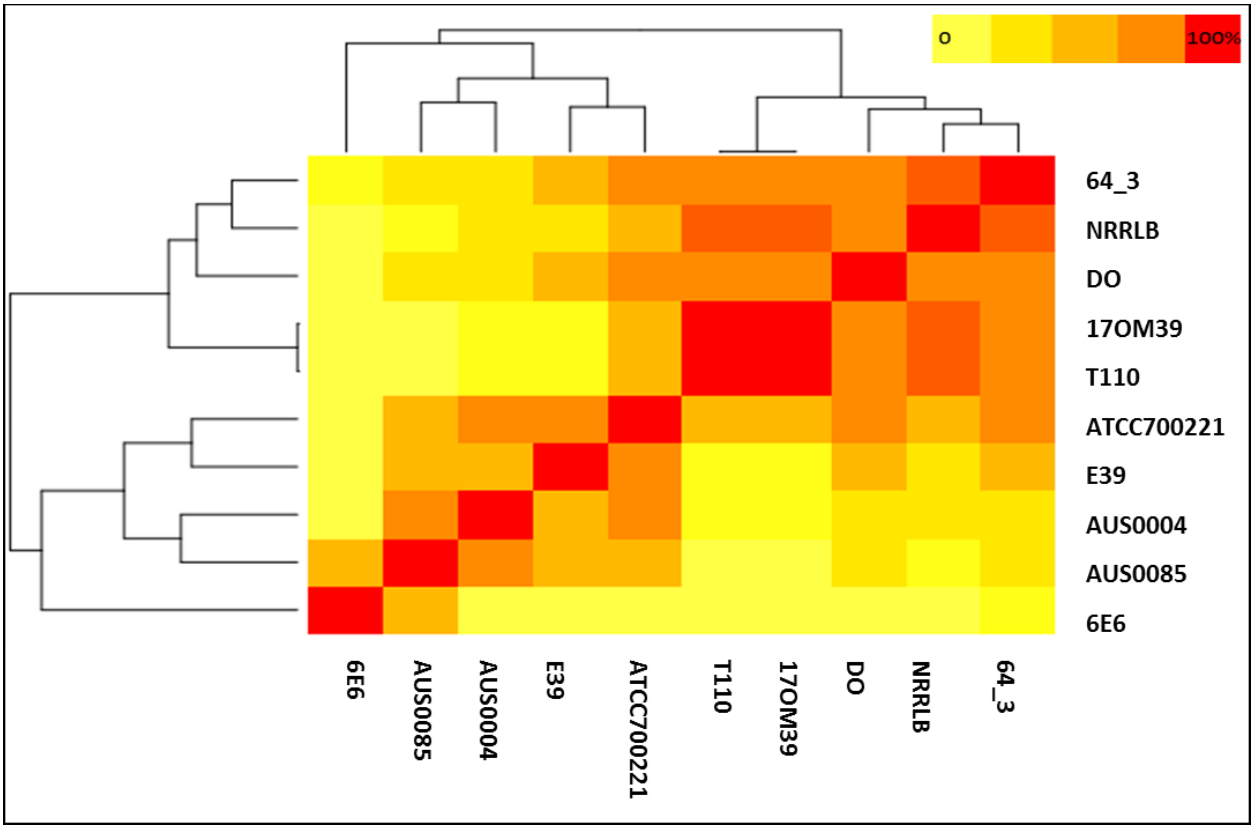

Figure 5 Heat map showing the similarities between the Genomic Islands of the strains considered in this study. The lightyellow shows least percent similarity while the red indicate 100% similarity with the genomes of the strains considered in the study. The pathogenic strains (E39, Aus0004, Aus0085 and 6E6) shows clustering while the probiotic strains (T110 & 17OM39), nonpathogenic strains (64_3 & NRRLB) and pathogenic strain DO present another clustering. The color scheme is as shown in the top right corner.

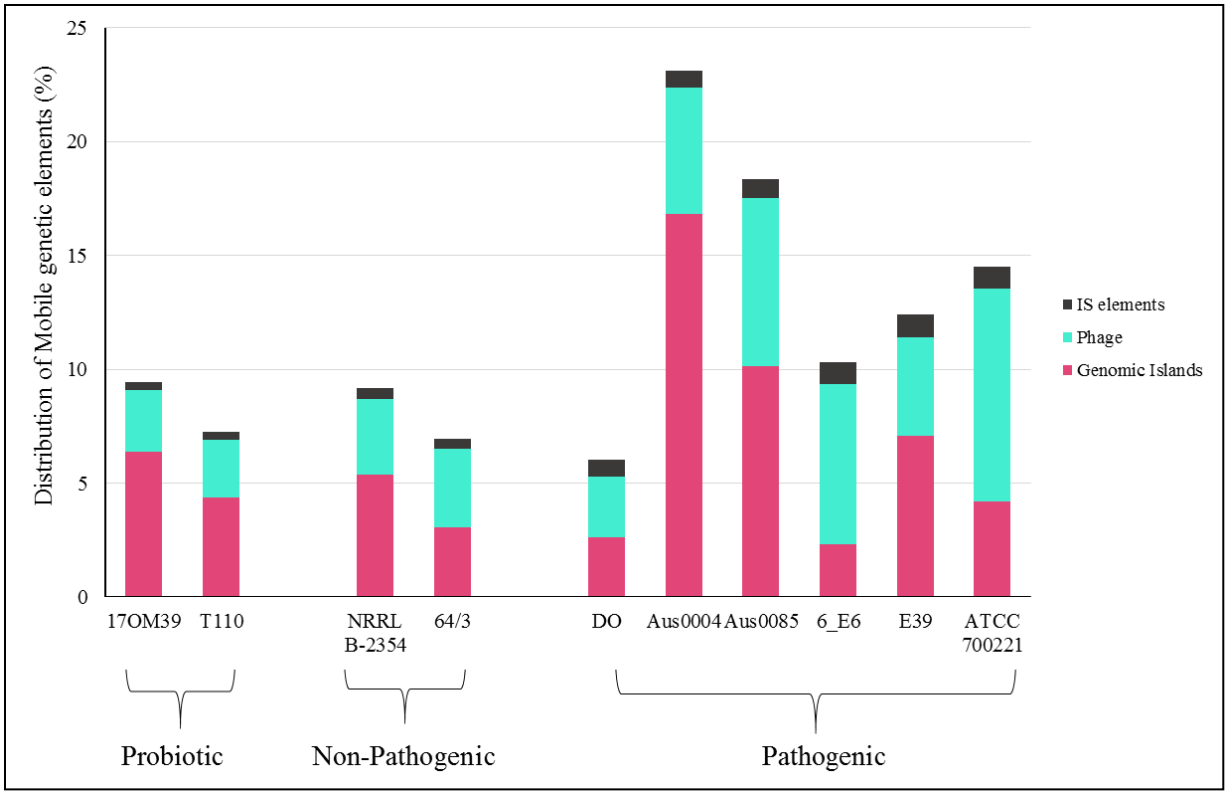

Figure 6 Proportion of Mobile Genetic Elements across *Enterococcus* genomes. The pick colour shows the proportion of genomic islands present in the each strain, light green for bacteriophages and black of IS elements across all the strains. Strain AUS004 has nearly quarter of its genome packed with these mobile genetic elements.

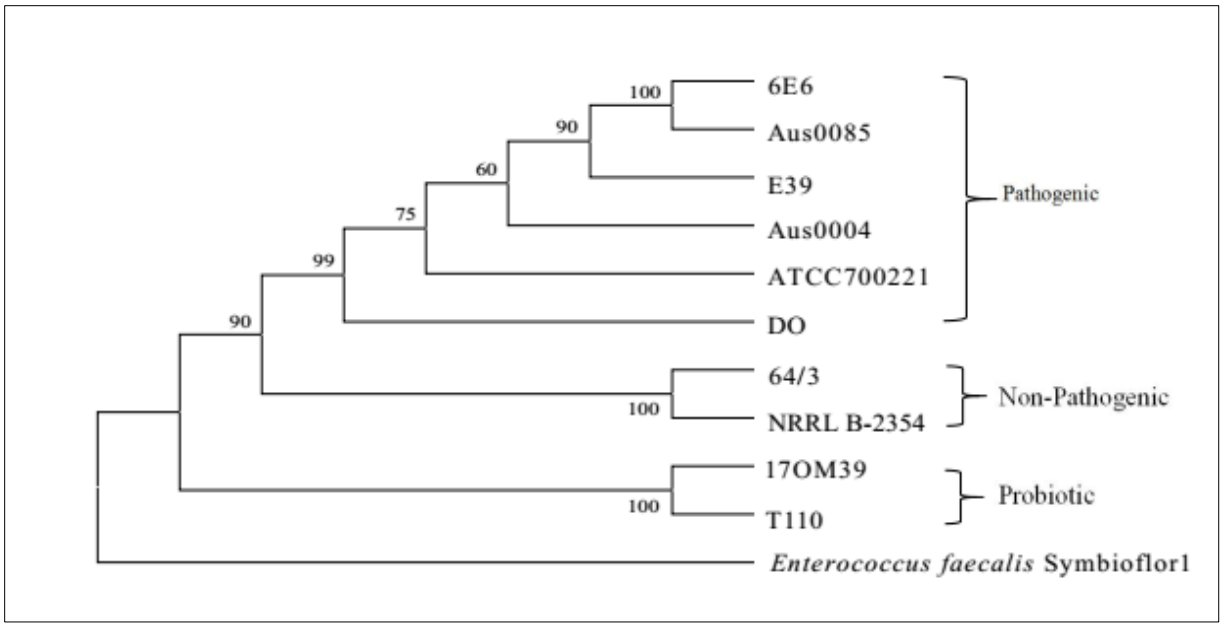

Figure 7 Core Genome Phylogeny. Phylogenetic tree of 10 *Enterococcus faecium* strains using the Maximum Likelihood method based on the GTR + G substitution model. The tree with the highest log likelihood (-17644.1414) is shown. Evolutionary analyses were conducted in MEGA6. A concatenated tree of 1945 genes was considered in the final dataset.

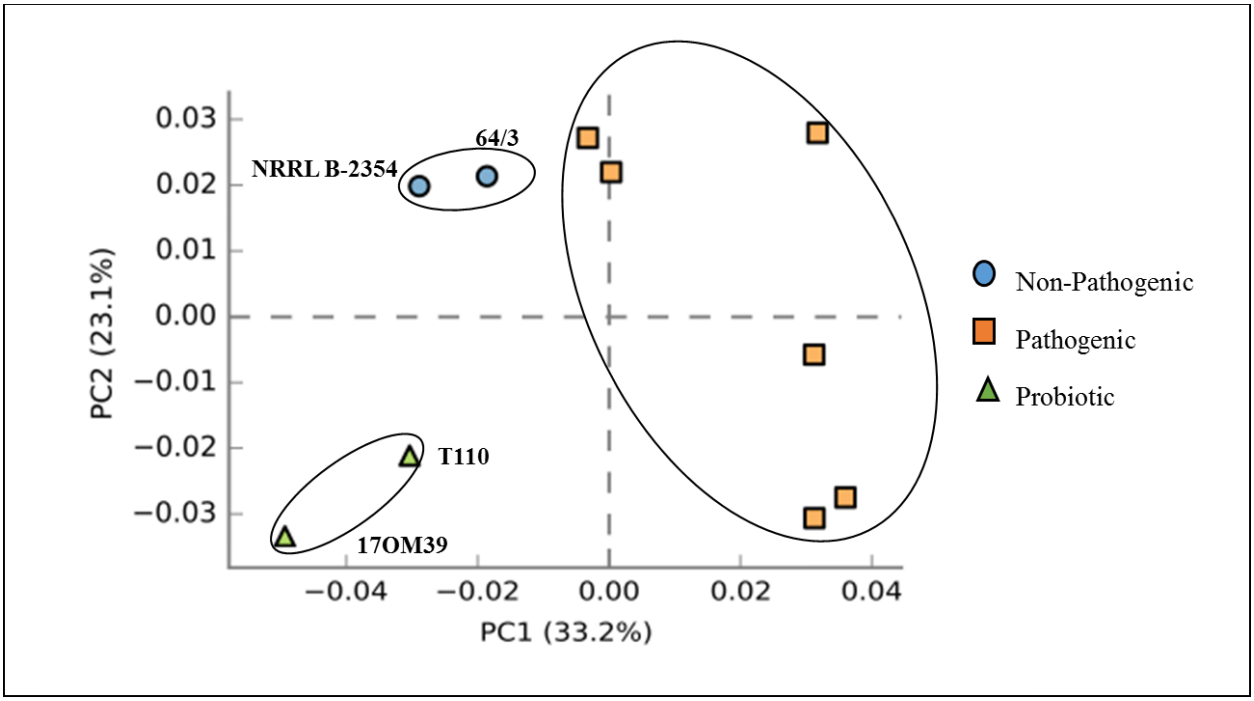

Figure 8 PCA plot comparing probiotic, pathogenic and non-pathogenic *Enterococcus* genomes based on presence and absence of genes responsible for survival in GI track, virulence factors and antibiotic resistance. The probiotic strains are shown in green, non-pathogenic in blue and pathogenic in red colour and the clustering is indicated by the oval shaped rings on the strains. From the plot, it can be noted that strain 17OM39 is different from the marketed probiotic strain T110.

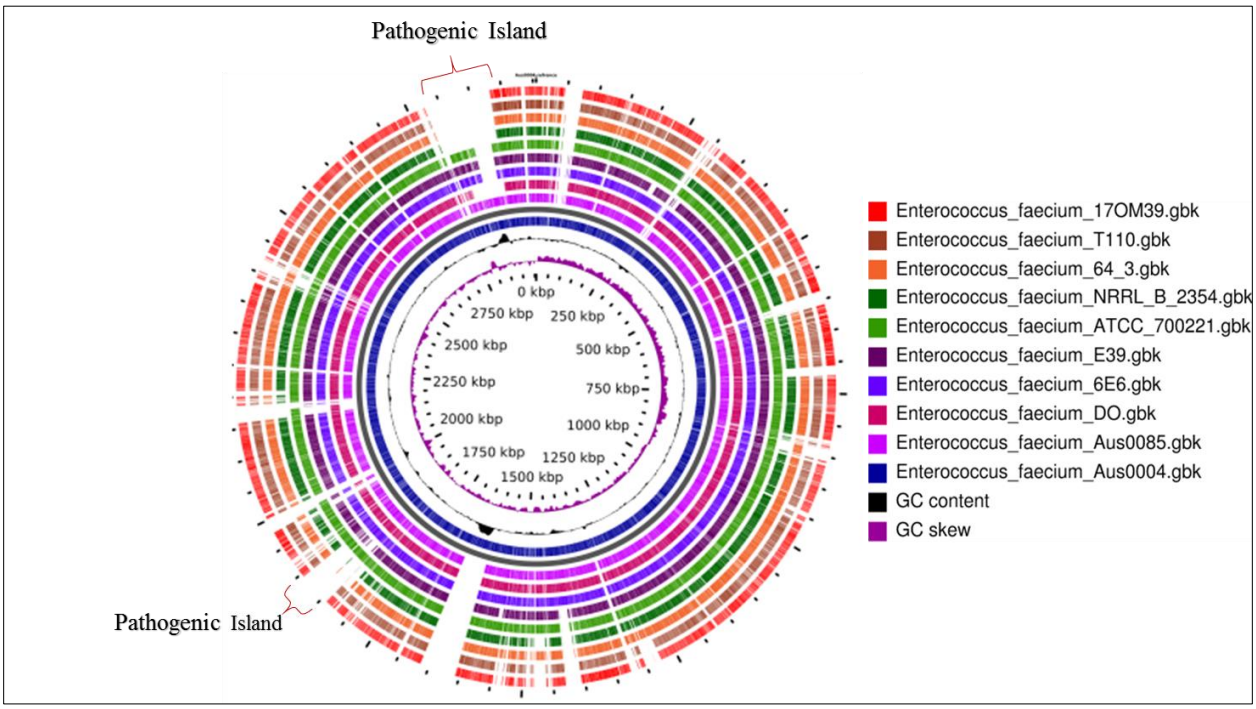

Figure 9 Blast Atlas of *Enterococcus* genomes, with strain Aus004 as a reference genome followed by Aus0085, DO, 6E_6, E_39, ATCC_7200221, NRRLB_2354, 64_3, T110 and the outermost as 17OM39. The two pathogenic islands (has most of virulence factors and antibiotic resistance genes) are shown in figure.

SUPPLEMENTARY MATERIAL

FIGURES

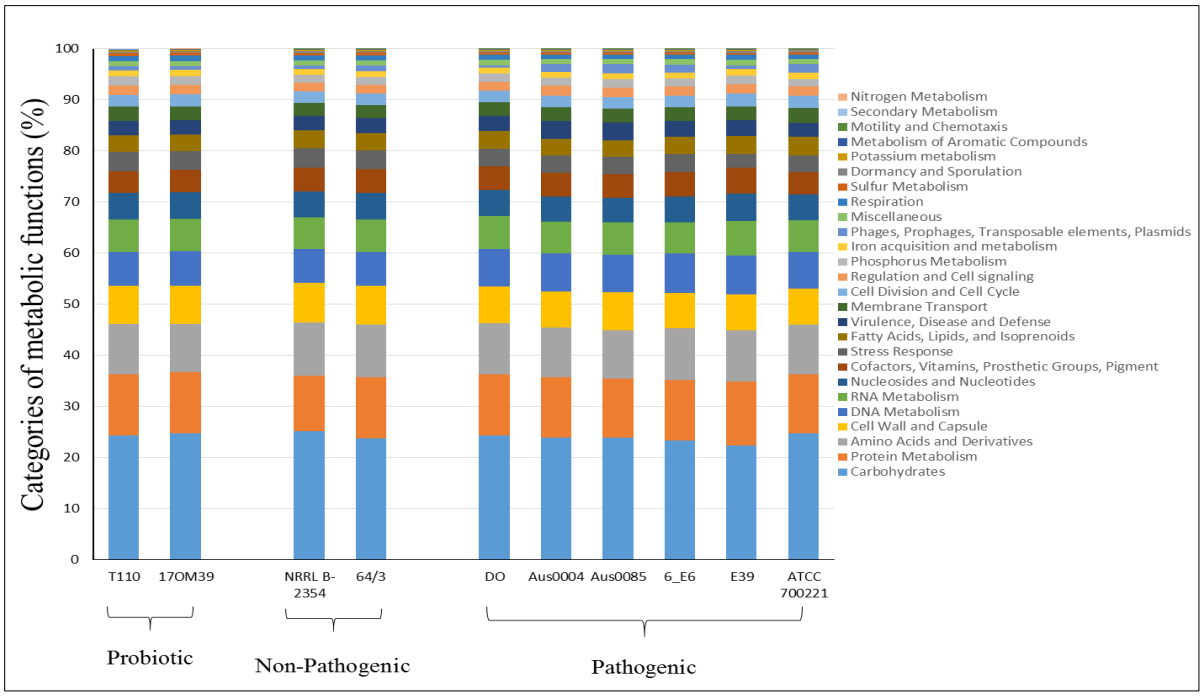

Fig S1. Features assigned to subsystems from RAST present in all ten *Enterococcus* strains.

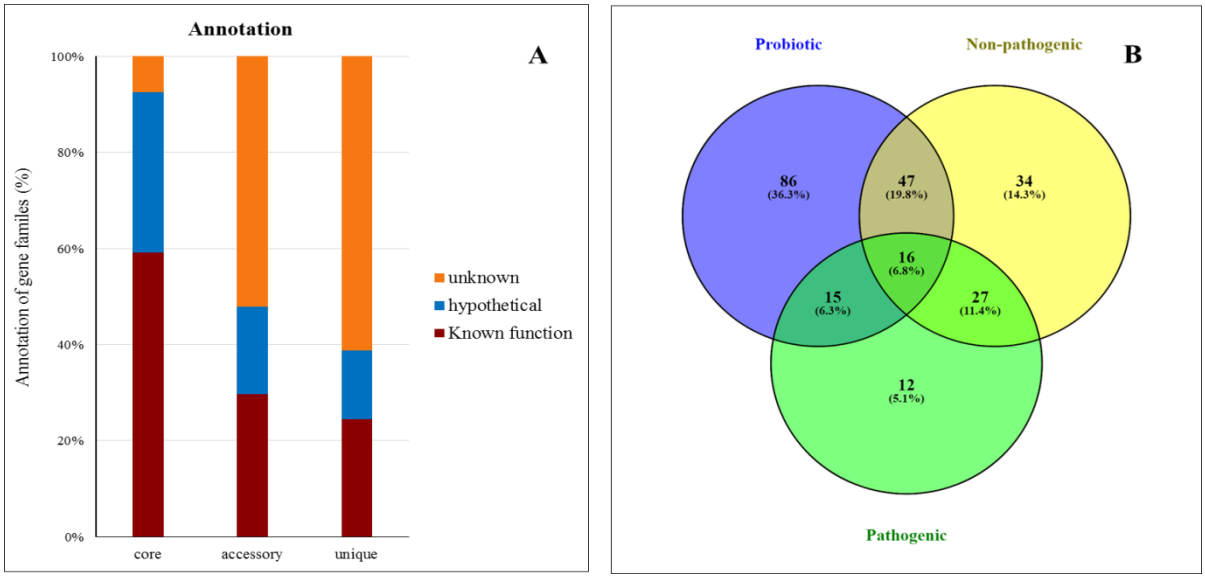

Figure S2 (A) Proportion of known, hypothetical and unknown proteins in the group of core, accessory and unique genes **(B)** Venn Diagram for accessory genome between probiotic, non-pathogenic and pathogenic group.

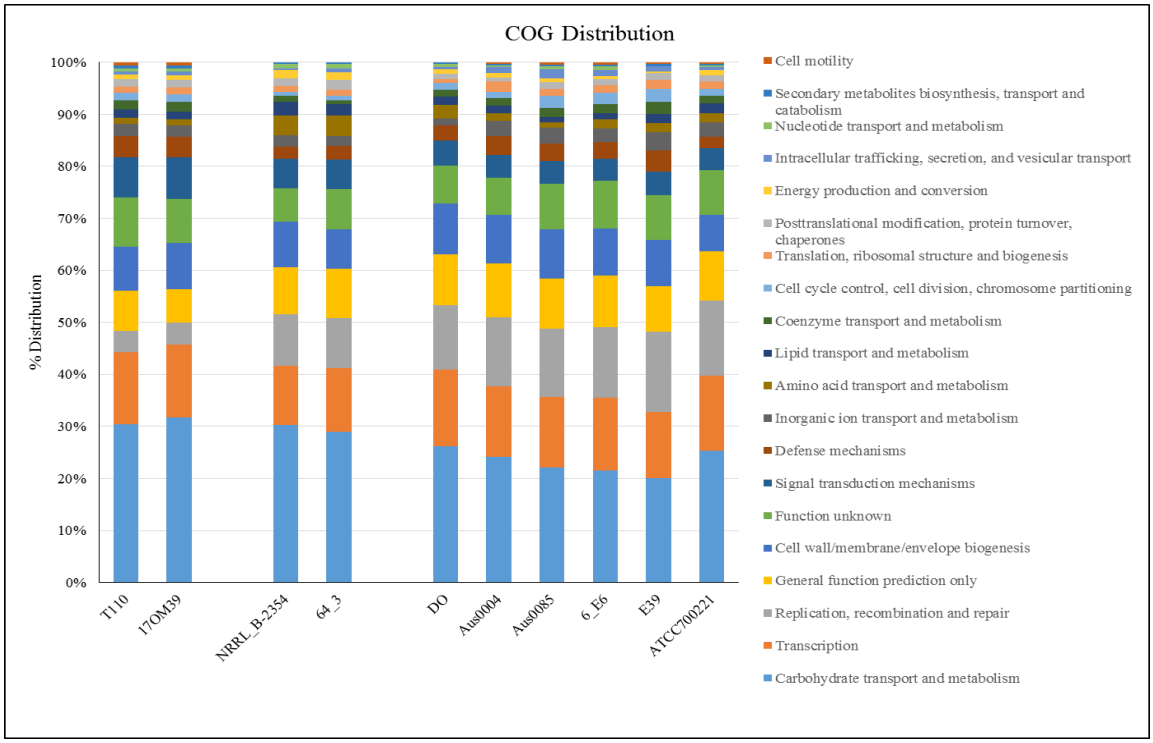

Figure S3 Functional analysis of the accessory genes in COG categories

TABLES

IS elements	Probiotic	Non-Pathogenic	Pathogenic
IS1216	-	+	+
IS1216E	-	+	+
IS1216V	-	+	+
IS1251	-	+	+
IS1476	-	-	+
IS1485	+	+	+
IS1542	+	-	-
IS16	-	-	+
IS1678	-	-	+
IS19	+	+	+
IS256	-	-	+
IS6770	-	+	+
ISEf1	-	-	+
ISEfa10	-	+	+
ISEfa11	-	-	+
ISEfa12	-	+	-
ISEfa13	-	-	+
ISEfa4	-	-	+
ISEfa5	-	-	+
ISEfa7	-	-	+
ISEfa8	-	-	+
ISEfm1	+	+	+
ISEfm2	+	-	+
ISEnfa110	-	-	+
ISEnfa3	-	-	+
ISEnfa4	+	-	+
ISLgar5	+	-	+
ISS1W	-	+	+
ISSsu5	-	-	+

Table S1 IS elements found in *Enterococcus* genomes by ISfinder tool. + Present, - Absent

Phage	17OM39	T110	NRRL B-2354	64/3	DO	Aus0004	Aus0085	6_E6	E39	ATCC 700221
Intact	1	2	2	2	1	3	3	3	2	4
Questionable	1	-	-	-	1	1	1	1	2	2
Incomplete	-	-	-	-	1	1	1	-	-	-
Total Size (bp)	73100	68700	87000	89100	74500	164900	220600	208400	117000	268300

Table S2 Number of Phage elements present in *Enterococcus* genomes as intact, questionable and incomplete.

Genome	Genomic Islands	Size (bp)
17OM39	11	172279
T110	9	117774
NRRL B-2354	12	142123
64/3	7	78409
DO	7	73539
Aus0004	35	496552
Aus0085	24	304164
6_E6	8	69233
E39	17	191228
ATCC 700221	10	119456

Table S3 Number of Genomic Islands in *Enterococcus* genomes

Genes	6_E6	ATCC_700221	Aus0085	DO	E39
Aminoglycoside	+	+	+	+	+
Chloramphenicol	-	-	-	+	-
Dihydrofolate	-	+	-	-	-
Erythromycin	+	+	+	+	+
Gentamicin	+	+	-	-	-
Lincosamide	-	-	+	-	-
Streptothricin	-	+	-	+	+
Tetracycline	-	-	-	+	-
Vancomycin	+	+	-	-	+

Table S4 Antibiotic Resistance genes found in *Enterococcus* plasmids as performed by CARD analysis, where + Present and - Absent

IS elements	6_E6	ATCC 700221	Aus 0004	Aus 0085	DO	E39	NRRLB 2354	T110
IS1062	-	-	-	-	+	-	-	-
IS1182	-	+	-	-	+	+	-	-
IS1216	+	+	-	+	+	+	+	-
IS1216E	+	+	-	+	+	+	+	-
IS1216V	+	+	-	+	+	+	+	-
IS1251	-	+	-	-	-	+	+	-
IS1297	+	-	-	+	+	+	+	-
IS1476	-	+	-	-	+	-	-	-
IS1485	+	+	-	+	+	+	+	-
IS16	+	-	-	-	+	-	-	-
IS19	+	+	-	+	+	+	+	-
IS256	+	+	-	+	-	+	+	-
IS6770	-	+	-	-	+	-	-	-
ISCco2	+	+	-	+	+	+	-	-
ISEf1	+	+	-	+	+	+	+	-
ISEfa10	+	+	-	-	-	+	-	-
ISEfa11	+	+	-	-	+	+	+	-
ISEfa12	-	-	-	-	+	-	-	-
ISEfa13	+	-	-	-	-	-	-	-
ISEfa4	+	-	+	+	+	+	-	-
ISEfa5	+	+	-	-	+	+	+	-
ISEfa7	+	+	-	+	+	+	-	-
ISEfa8	+	+	-	+	+	+	-	-
ISEfm1	+	+	-	+	+	+	+	-
ISEfm2	+	+	-	-	+	+	-	-
ISEnfa3	-	-	-	+	-	-	-	-
ISEnfa4	+	+	-	+	-	+	+	-
ISLgar5	+	-	-	+	-	-	-	-
ISLpl1	-	-	-	-	-	-	+	-
ISPp1	-	-	-	-	-	-	+	-
ISS1CH	+	-	-	+	+	+	+	-
ISS1D	+	-	-	+	+	+	+	-
ISS1E	+	-	-	+	+	+	+	-
ISS1M	+	-	-	+	+	+	+	-
ISS1N	+	-	-	+	+	+	+	-
ISS1W	+	+	-	+	+	+	+	-

Table S5 IS elements found in *Enterococcus* plasmids by ISfinder tool. + Present, - Absent