The Virulence Associated *fsr* Quorum Sensing System Play Key Roles in *Enterococcus Faecalis*

Jinglin Yue
Peking University Ditan Teaching Hospital
Pengcheng Du
Beijing Key Laboratory of Emerging Infectious Diseases
Mingxi Hua
Beijing Key Laboratory of Emerging Infectious Diseases
Xinzhe Liu
Beijing Key Laboratory of Emerging Infectious Diseases
Ang Duan
Beijing Key Laboratory of Emerging Infectious Diseases
Nan Chen
Beijing Key Laboratory of Emerging Infectious Diseases
Chen Chen (chenchen1@ccmu.edu.cn)
Peking University Ditan Teaching Hospital

Research Article

Keywords: Enterococcus faecalis, Genomes, fsr quorum sensing system, Antibiotic resistance genes

Posted Date: June 9th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-562032/v1

Abstract

Enterococcus faecalis, a major nosocomial pathogen, has become a top leading cause of hospital-acquired infections including urinary tract infections, endocarditis and bacteremia. It is important to study the epidemiology and virulence characteristics of *E. faecalis* isolates in order to tailor infection prevention and antibiotic prescribing. In this study, comparative genomic analysis was conducted on 537 isolates from different human origins. The isolates from bloodstream and intra-abdominal lining had the largest and smallest average genome size respectively, while the isolates from open natural orifices (gastrointestinal tract, urinary tract, respiratory tract, wound and eye) had medium average genome size. The phylogenetic relationships were expounded that the strain isolation niche is uncorrelated with strain phylogeny. Six clonal complexes generally appeared in different isolation sources. Furthermore, genomic analysis revealed differences at the accessory genome, the functions of different genes mainly pointed to the virulence, drug resistance and metabolism of *E. faecalis*. Interestingly, *fsr* quorum sensing system genes affecting biofilm formation had a highest proportion in the blood-derived strains. This study showed the genomic characteristics of different human origins and suggested that *fsr* quorum-sensing system maybe a contributing factor of bacteremia due to *E. faecalis* infection.

Introduction

Enterococcus faecalis is a gram-positive facultative anaerobe inhabiting the gastrointestinal tract in humans as normal commensals [1]. *E. faecalis* has emerged as an important healthcare-associated pathogen during the past few decades causing 15% of urinary tract infections in intensive careunit (ICU) patients and 10%-15% of cases of endocarditis and has became the second leading cause of healthcare-associated bacteremia currently [2]. Nosocomial *E. faecalis* infections are frequently caused by hospital-acquired strains which are thought to have acquired factors important for virulence, colonization, and/or for fitness in the hospital environment [3–5].

Major virulence factors in *E. faecalis* are supposed to be six categories including adherence, antiphagocytosis, biofilm formation, exoenzyme, quorum sensing system and toxin. Among, biofilm formation is regulated by quorum-sensing system in several important pathogens, including *Pseudomonas aeruginosa* and *Staphylococcus aureus*. In *E. faecalis*, the *fsr* quorum sensing system, comprising 3 genes (*fsrA*, *fsrB* and *fsrC*), is an important regulator with both positive and negative effects, regulating *gelE*, *sprE* and *bopD* expression that are important for biofilm formation, along with genes implicated in several metabolic pathways. The earliest reports of *E. faecalis* in associated with infection-related biofilm were probably from the studies that identified strains in infected vascular ports from patients and in a urinary stone [6]. Moreover, isolates from endocarditis and intravascular catheter-associated bloodstream infections display particularly robust formation of biofilms [7]. Recently, it has been reported some studies of the function and diversity of *fsr* quorum sensing system of *E. faecalis* without considering isolation niche. Therefore, we are now interested in the roles of the *fsr* quorum sensing system in different isolation niche of *E. faecalis* of infectious patients.

In this study, a total of 537 genomes of *E. faecalis* (including 52 genomes sequenced in this work and 485 genomes retrieved from GenBank database) were subjected to comparative genomic analysis. We collected strains isolated from blood, intra-abdominal, GI tract, urinary tract, wound, eye and respiratory tract to characterize the population structure and virulence characteristics. Our results revealed that the genomic characteristics of different human origins and implied that *fsr* quorum sensing system maybe a contributing factor in *E. faecalis* bloodstream infections.

Methods

Bacterial strains

We enrolled all of the *E. faecalis* collected by the clinical laboratory of Beijing Ditan Hospital from 2011 to 2014. The 52 isolates were identified by VITEK2 firstly, then the classification was validated by 16S rRNA amplification and sequencing. The published 485 *E. faecalis* genomes available on September, 2020 were downloaded from GenBank database.

Total DNA extraction, Sequencing, assembly and annotation

The strains were cultured and then total DNA was extracted using the Genomic DNA Kit (Qiagen, USA) according to the manufacturer's instructions. Sequencing libraries were generated using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NEB, USA), and the whole genome of the 52 strains was sequenced using the Illumina NovaSeq platform. The clean read data were obtained using fastQC [8]. The paired-end reads were assembled de novo using SPAdes v3.13.0 [9]. The assembled genomes were annotated using Prokka [10].

Phylogenetic analysis

The Illumina reads and assembled genomes from GenBank database were first mapped to the complete genome sequence of *E. faecalis* strain V583 (accession number: AE016830.1) using bowtie 2 software, and the results were filtered using Samtools [11]. The variant sites conserved in all strains (core genome SNPs, cgSNPs) were retained, and the sequence of these sites in each strain were concatenated and employed for phylogenetic analysis using FastTree v2.1.11 [12]with the maximum likelihood method.

Multilocus sequence typing

Multilocus sequence typing (MLST) was conducted by the reference method. The genome sequences were compared with the nucleotide sequences of housekeeping genes (*gdh*, *gyd*, *pstS*, *gki*, *aroE*, *xpt*, and *yqiL*) in the MLST database (https://pubmlst.org/organisms/enterococcus-faecalis), to determine the number of alleles and assigned STs. A clonal complex was defined as STs sharing alleles at five or six of seven loci (http://eburst.mlst.net/).

Construction of accessory genome

The core- and pan- genome of the genomes were analyzed using Roary [13]. The pan-genome was constructed by counting the total number of non-redundant gene families within the complete dataset. The core-genome was constructed by counting the total number of gene families commonly shared by all genomes. The accessory-genome was the result of that the pan-genome minus the core-genome. We calculated the P value of the gene presence rates of accessory genome between two groups randomly for all eight source groups. If there is a P value <=0.01, this gene is the meaningfully differential gene concerned.

Identification of virulence genes

A BLAST search was performed with all predicted genes from the 537 strains against the Virulence Factor Database (VFDB) to identity potential virulence genes (sequence identity >=80% and alignment length / subject sequence length >=80%).

Identification of antibiotic resistance genes

Antimicrobial resistance genes were identified by performing searches using ABRicate (https://github.com/tseemann/abricate) (sequence identity >=80% and alignment length / subject sequence length >=80%). [If you publish the results of Abricate please cite both the software and the appropriate database you used with –db : CARD - doi:10.1093/nar/gkw1004; ARG-ANNOT - doi:10.1128/AAC.01310-13]

Statistical analysis

Statistical analyses were performed using R v3.6.2, applying the Kruskal-Wallis rank sum test, followed by the multiple comparison test after Kruskal-Wallis (kruskalmc) to compare the gene presence rates of accessory genome in different groups. The p-values <0.05 were considered significant.

Results

Isolation and characteristics of *E. faecalis*.

A total of 537 genomes of *E. faecalis* were subjected to genomic analysis, including 52 genomes that obtained from the isolations in our hospital. These genomes covered 537 strains isolated from blood (42 strains), intra-abdominal (34 strains), GI tract (369 strains), urinary tract (55 strains), respiratory tract (3 strains), wound (19 strains), eye (5 strains) and unknown source (10 strains) (Additional file 1).

The analysis of *E. faecalis* genomes displayed low G+C contents ranging from 37.0 to 38.0% with an average genome size 2.98 Mb. The largest genome size is 3.10±0.19 Mb, whereas the lowest genome size is 2.90±0.13 Mb (Figure 1). Considering the average gene size of *E. faecalis* is ~1kb, the largest and the smallest genome have 200 gene differences, which displayed large genomic diversity in *E. faecalis*.

Then, we investigated the differences of genome size and GC contents among the strains from different sources, and found that the isolates from bloodstream had a largest genome size but lowest GC contents. Also, we observed significant differences in genome size between strains from blood and the other two sources (blood vs. GI tract: p-value<0.01; blood vs. intra-abdominal: p-value<0.01), and significant differences in GC content between blood and intra-abdominal (p-value<0.01). These results indicated different genomic characteristics of *E. faecalis* that isolated in different body sites.

The phylogeny and MLST structures of *E. faecalis*

To investigate the phylogenetic relationship of 537 strains, we performed complete genomic comparison analysis on these strains. First, the core genome contained 1425 gene with 55167 cgSNPs. According to 55167 cgSNPs, the phylogenic analysis showed that 537 strains were divided into three main clades on the phylogenetic tree. Clade A composed of 189 strains of which were isolated from other six source groups expect eye and respiratory system. While Clade B and C both including all eight source groups contained 167 and 181 strains respectively (Figure 2).

Then, further MLST analysis on 537 isolates from different sources and geographic locations revealed 106 different sequence types that grouped into nine major clonal complexes (CC4, CC21, CC25, CC30, CC40, CC64, CC179, CC476 and CC483) (Figure 3). The nine CCs widely distributed among strains of different origins and almost nine CCs contained strains from GI tract. For strains of non-open natural orifices, the MLST of strains from bloodstream were ST9, ST742, ST959, ST743, ST674, ST79 and ST32, while the MLST of strains from intra-abdominal lining were mainly ST428, ST858, ST857, ST872, ST483, ST721 and ST479. For strains of open natural orifices, the MLST of strains from urinary tract were ST619, ST856, ST689, ST747, ST470, ST590, ST745, ST744, ST525 and ST943. Additionally, the strains from GI tract had the largest number and revealed 68 different sequence types.

Therefore, the scatter of different source strains in the three phylogenic clades and population structures of *E. faecalis* defined by MLST suggested that the correlation between strain isolation niche and strain phylogeny might be weaken.

The accessory genome of *E. faecalis*

Pan-genome analysis of 537 *E. faecalis* revealed that the accessory genome composed 17243 genes. To investigate the different gene contents from 8 sources, we pair-wised compared the present/absent of genes in different sources. We identified 2546 genes of the accessory genome were significantly distributed in one sources of samples at least. Molecular functionally analysis on these genes showed that 53.1% genes were unknown function. 9.4% genes were associated to replication, recombination and repair, which were most frequent genes in the right part (Figure 4). Then, we constructed a matrix on the significant p-value between 8 sources, and found three clusters, namely Cluster A (Low similarity, LS), Cluster B (High similarity, HS), Cluster C (Middle similarity, MS).

Cluster A suggested that stains from blood and GI tract had a wide and huge gene difference compared to other six isolation groups in information storage and processing, metabolism and cellular processes and signaling categories. More specifically, above genes about cell wall/membrane/envelope biogenesis, cell motility, defense mechanisms and coenzyme transport and metabolism were almost involved in virulence and drug resistance of *E. faecalis*. Strains from eye and respiratory tract showed considerably less different from other six isolation groups in Cluster B reversely. Cluster C is just a control group due to the presence of samples of unknown sources.

Antibiotic resistance genes

Next, We analyzed that antibiotic factors in the isolates according to the against with *E. faecalis* virulence genes recorded in multiple database. A total of 59 antibiotic resistance genes were found within the investigated *E. faecalis* genomes (Figure 5). The percentage of antibiotic resistance genes varied greatly between isolation groups. Based on the average percentage of antibiotic resistance genes, the 537 investigated strains could be divided into two classes, namely Group A consisted of strains from GI tract, eye and respiratory tract and Group B consisted of strains from blood, wound, intra-abdominal lining and urinary tract. In generally, the Group B represented by strains from blood have a obviously higher average presence percentage and numbers of antibiotic type, comparing to Group A represented by strains from GI tract.

Moreover, we found two types of vancomycin resistance genes: vanA-cluster (vanYA, vanZA, vanSA, vanHAX2 and vanRA) and vanB-cluster (vanHBX1, vanHBX2, vanSB, vanVB, vanWB, vanYB and vanRB). There was a lower gene percentage of vanB-cluster in all source groups obviously. However, vanA-cluster showed a higher percentage in Group B and the strains isolated from blood had the largest presence percentage. Although the distribution of antibiotic resistance genes varied greatly between strains, six antibiotic resistance genes, consisting of *lsaA* (ABC-F subfamily protein), *efrA/B* (heterodimeric ABC transporter efflux pump), *emeA* (multidrug efflux pump), *dfrE* (chromosome-encoded dihydrofolate reductase) and *mphD* (PTS system fructose-specific EIID component), were commonly present in all investigated isolates.

Virulence genes

We analyzed that virulence factors in the isolates according to the comparison with *E. faecalis* virulence genes recorded in VFDB. Thirty-six genes encoding six groups of common virulence factors in *E. faecalis* were identified: the adherence-related genes, the antiphagocytosis-related genes, the biofilm formation-associated genes, the genes encoding exoenzyme, the quorum sensing system-associated genes and the genes encoding toxins (Figure 6). The adherence-related genes (*srtC, ebpC, efaA, ebpA and ebpB*) and the biofilm formation-associated genes (*bopD*) were almost identified in all eight source groups. The biosynthesis of capsular polysaccharides by E. faecalis is encoded by the csp operon, which includes 11 open reading frames (cpsA to cpsK). The presence percentages of nine genes *cpsC, cpsD, cpsE, cpsF, cpsG, cpsH, cpsI, cpsJ* and *cpsK* were higher in isolates from blood, wound, intra-abdominal and urinary tract, whereas there were lower presence percentage of above nine genes in isolates from GI tract, eye and

respiratory tract. Moreover, some members of the cytolysin (cyl) operon were detected in the *E. faecalis* genomes. Normally, the cyl operon comprises eight genes, i.e. *cylA/B/I/L/M/R1/R2/S*. The presence of six genes (*cylI/L/M/R1/R2/S*) were higher in isolates from eye, urinary tract and wound but were lower in strains from blood and intra-abdominal lining. Next, we have interests in focusing on quorum sensing system-associated genes and observed that the presence of these genes (*fsrA, fsrB, fsrC, gelE* and *sprE*) were higher in isolates from blood compared to isolates from GI tract, especially *fsrA* and *fsrB*.

The diversity of *fsr-sprE* region of *E. faecalis*.

To illustrate the relationship between *fsr* quorum sensing system and *E. faecalis* bloodstream infections, we analyzed the genomic characteristics of *fsr-sprE* region and gene presence rates in all eight isolation groups. We divided the genomic patterns in this region into three types: Cluster I/II/III based on the presence or absence of these five genes (*fsrA, fsrB, fsrC, gelE* and *sprE*) (Figure 7, Table 1). Cluster I genomic pattern refers to that the genome of the *fsr-gelE* region of the strain is compared with the genome of the *fsr-gelE* region of the blood-derived *E. faecalis* V583 strain. All five genes were well compared, and the upstream and downstream of the region were also well compared, with no deletion or insertion. The Cluster II genomic pattern refers to the comparison of the genome in the *fsr-gelE* region of the strain with that of the blood-derived *E. faecalis* V583 isolate, there was deletion of one gene in 5 genes, and there was insertion of gene cluster in the upstream of this region, and deletion of gene cluster in the downstream. Cluster III genomic pattern refers to the comparison of the genome in the *fsr-gelE* region of the strain with that of the blood-derived *E. faecalis* V583 isolate, there were deletions of 3 or more genes in 5 genes, and deletions of gene clusters were found downstream of this region.

In general, the *fsr-gelE* regional genomic patterns of all the strains from eight different sources were mainly clustered in Cluster I and Cluster III, while Cluster II only existed in the strains from urinary tract and GI tract. *E. faecalis* was less likely to cause infections in the eyes and respiratory system. In addition to these two uncommon infection origins, the detection rate of *fsr*-gelE regional genome Cluster I pattern in strains of blood, wound, abdominal, urethral, and GI tract decreased in turn (Figure 8).

Discussion

Analysis of 110 E. faecalis strains isolated worldwide from about 20th century revealed four major CCs (CC2, CC9, CC10 and CC21) by use of eBURST [14]. And analysis of more recently isolated strains from Europe (primarily isolated from 2006-2009) found that epidemiology lineages is enriched in six large E. faecalis CCs (CC2, CC16, CC21, CC30, CC40 and CC87) [15].We then performed a phylogenetic reconstruction using cSNPs of all 537 *E. faecalis* strains. The results suggested no correlation between the strain isolation niche and phylogeny, which corroborates the inference drawn by previous studies.

Multidrug resistant E. faecalis has emerged as leading causes of hospital acquired infection and become a public threat worldwide over the past 30 years [16]. With the acquisition of resistance to vancomycin in the mid 1980's, few antibiotics are extremely effective for treating infections caused by E. faecalis. To

address this problem, knowledge of the epidemic characteristics, pathogenesis and antibiotic resistance mechanisms of this pathogen is necessary and important. Since then, researchers with an interest in various aspects of E. faecalis biology had performed large amount of studies to uncover the molecular characteristics, resistance profiles and mechanisms of this pathogen [17,18]. In this study, we conducted a comparative genomic analysis of 537 E. faecalis strains. These strains were originated from a wide range of human sources, including blood, eye, intra-abdominal, GI tract, urinary tract, respiratory tract, wound and unknown. The *fsr* quorum sensing system genes affecting biofilm formation had a highest proportion in the blood-derived strains by contrast. Therefore, we had a guess that the *fsr* quorum-sensing system may be a contributing factor of *E. faecalis* bloodstream infections.

In order to further explore the relationship between *fsr* quorum sensing system and Enterococcus faecalis bloodstream infection, we focused on the genomic characteristics and related genes of the *fsr-gelE* region in 537 Enterococcus faecalis genomes. Cluster I and Cluster III were the main *fsr-gelE* regional genomic patterns in all the 8 strains from different sources, while Cluster II was only found in the strains from urinary tract and gastrointestinal tract. Since the colonization site of Enterococcus faecalis is in the gastrointestinal tract, the detection of Enterococcus faecalis in other tissue sites is usually due to infection [19]. In our study, we observed that the Cluster III genomic pattern was the dominant pattern in gastrointestinal strains and the Cluster I genomic pattern was the dominant pattern in blood-derived strains. In addition, a large number of studies have found that mobile elements such as bacteriophages, plasmids and transposons of bacteria are usually aggregated on large genomic islands, which are usually flanked by short repeated DNA sequences, characterized by abnormal G + C content [20, 21]. Combined with the above evidences, we tend to believe that the genomic pattern of Cluster III is the original genomic structure of Enterococcus faecalis, and the *fsr-gelE* region may be acquired through horizontal transfer and other mechanisms. *fsrA/B/C* is the population density sensing system of Enterococcus faecalis, which can sense bacterial density and regulate toxicity. Gelatinase and serine protease are encoded by *gelE* and *sprE*, respectively, and are positively regulated by *fsr* quorum sensing system. In general, these proteases and *fsr* quorum sensing systems contribute to virulence, host tissue degradation, and biofilm formation [22]. Compared to the gastrointestinal tract, where Enterococcus faecalis originally colonized, blood flow is more fluid, and the acquisition of *fsr-gelE* region contributes to the formation of biofilm, which helps Enterococcus faecalis colonize.

Conclusions

This study showed the genomic characteristics of different human origins and suggested that *fsr* quorum-sensing system maybe a contributing factor of bacteremia due to *E. faecalis* infection.

Declarations

Authors' contributions:

Chen Chen designed the study, Jinglin Yue engaged in the acquisition, analysis and interpretation of the data, Pengcheng Du, Mingxi Hua, Xinzhe Liu, Ang Duan and Nan Chen Participate in data acquisition and analysis. Jinglin Yue drafted the article and all the authers had final approval of the version to be submitted.

References

1. Kao PHN, Kline KA. Dr. Jekyll, et al. How Enterococcus faecalis Subverts the Host Immune Response to Cause Infection[J]. J Mol Biol, 2019, 431(16):2932-45.
2. Jett BD, Huycke MM, Gilmore MS. Virulence of enterococci[J]. Clin Microbiol Rev, 1994, 7(4):462–78.
3. Koch S, Hufnagel M, Theilacker C, Huebner J. Enterococcal infections: host response, therapeutic, and prophylactic possibilities[J]. Vaccine, 2004, 22(7):822–30.
4. Franz CMAP, Stiles ME, Schleifer KH, Holzapfel WH. Enterococci in foods–a conundrum for food safety[J]. Int J Food Microbiol, 2003, 88(2–3):105–22.
5. Anderson DJ, Olaison L, McDonald JR, et al. Enterococcal prosthetic valve infective endocarditis: report of 45 episodes from the International Collaboration on Endocarditis-merged database[J]. Eur J Clin Microbiol Infect Dis, 2005, 24(10):665–70.
6. Guiton PS, Hung CS, Hancock LE, et al. Enterococcal biofilm formation and virulence in an optimized murine model of foreign body-associated urinary tract infections[J]. Infect Immun, 2010, 78(10):4166–75.
7. Roilides E, Simitsopoulou M, Katragkou A, Walsh TJ. How Biofilms Evade Host Defenses[J]. Microbiol Spectr, 2015, 3(3).
8. Leggett RM, Ramirez-Gonzalez RH, Clavijo BJ, et al. Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics[J]. Front Genet, 2013, 4:288.
9. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing[J]. J Comput Biol, 2012, 19(5):455–77.
10. Seemann T. Prokka: rapid prokaryotic genome annotation[J]. Bioinformatics, 2014, 30(14):2068–9.
11. Yu X, Sun S. Comparing a few SNP calling algorithms using low-coverage sequencing data[J]. BMC Bioinformatics, 2013, 14:274.
12. Tully BJ, Sachdeva R, Graham ED, Heidelberg JF. 290 metagenome-assembled genomes from the Mediterranean Sea: a resource for marine microbiology[J]. PeerJ, 2017, 5:e3558.
13. Page AJ, Cummins CA, Hunt M, et al. Roary: rapid large-scale prokaryote pan genome analysis[J]. Bioinformatics, 2015, 31(22):3691–3.
14. Ruiz-Garbajosa P, Bonten MJM, Robinson DA, et al. Multilocus sequence typing scheme for Enterococcus faecalis reveals hospital-adapted genetic complexes in a background of high rates of recombination[J]. J Clin Microbiol, 2006, 44(6):2220–8.
15. Kuch A, Willems RJL, Werner G, et al. Insight into antimicrobial susceptibility and population structure of contemporary human Enterococcus faecalis isolates from Europe[J]. J Antimicrob

Chemother, 2012, 67(3):551–8.
16. Raven KE, Reuter S, Gouliouris T, et al. Genome-based characterization of hospital-adapted Enterococcus faecalis lineages[J]. Nat Microbiol, 2016, 1:15033.
17. Rybak MJ. Pharmacodynamics: relation to antimicrobial resistance[J]. Am J Infect Control, 2006, 34(5 Suppl 1).
18. Courvalin P. Vancomycin resistance in gram-positive cocci[J]. Clin Infect Dis, 2006, 42 Suppl 1:S25-S34.
19. In: Gilmore MS, Clewell DB, Ike Y, et al. Enterococci: From Commensals to Leading Causes of Drug Resistant Infection[M/OL]. Boston: Massachusetts Eye and Ear Infirmary, 2014.PMID: 24649510.
20. Daubin V, Ochman H. Bacterial genomes as new gene homes: the genealogy of ORFans in E. coli[J]. Genome Res, 2004;14(6):1036–42.
21. Lawrence JG, Ochman H. Amelioration of bacterial genomes: rates of change and exchange[J]. J Mol Evol, 1997, 44(4):383–97.
22. Ali L, Goraya MU, Arafat Y, et al. Molecular Mechanism of Quorum-Sensing in Enterococcus faecalis: Its Role in Virulence and Therapeutic Approaches[J]. Int J Mol Sci, 2017, 18(5).

Tables

Table 1. Different genome patterns of *Enterococcus faecalis fsr-gelE* region

sprE	*gelE*	*fsrC*	*fsrB*	*fsrA*	Cluster
0	0	0	0	0	III
1	1	1	0	0	III
1	1	1	1	1	I
1	1	1	1	0	II
1	1	0	1	1	II
1	0	0	0	0	III
1	1	0	0	0	III
1	0	1	0	0	III

Figures

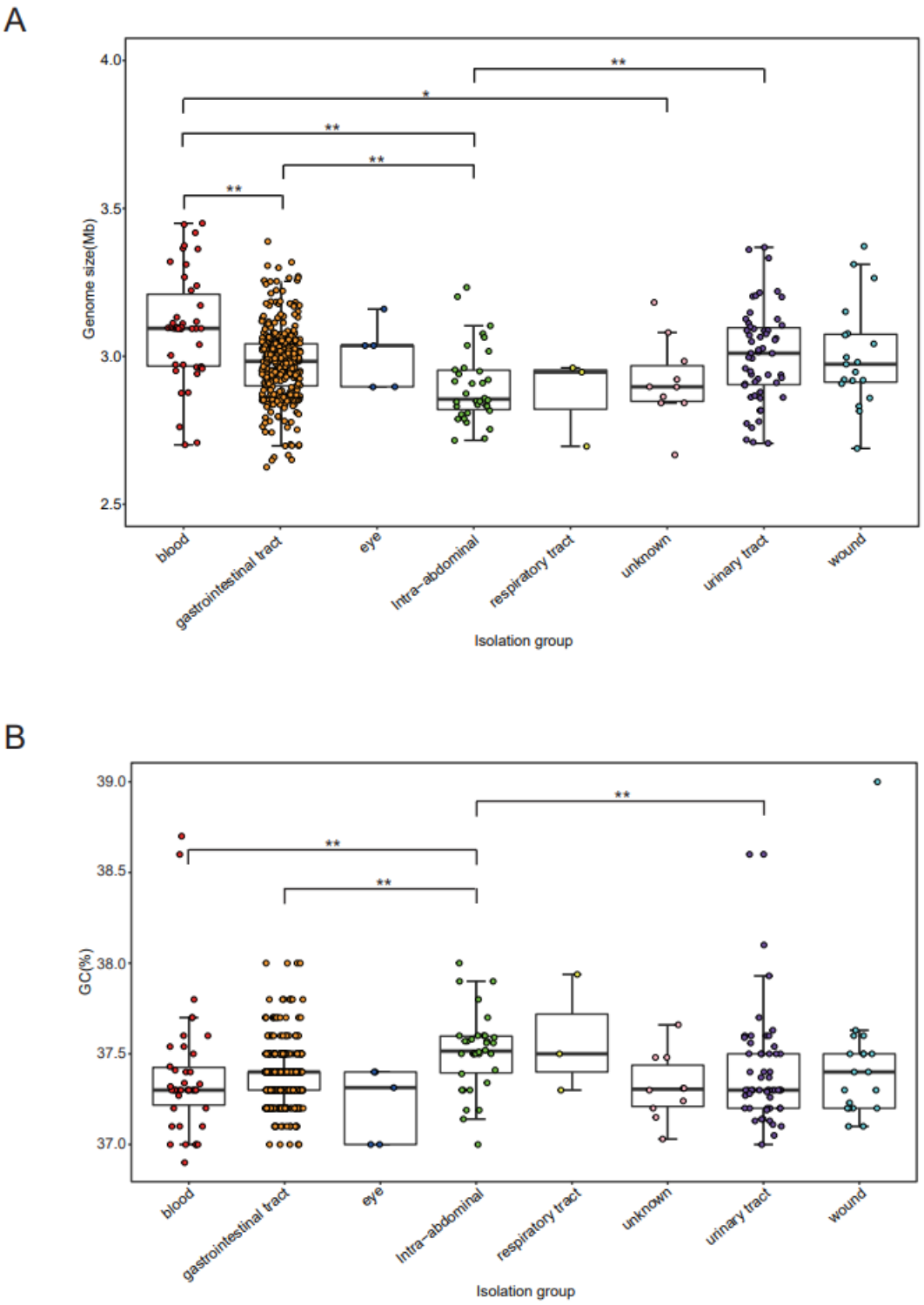

Figure 1

Genome size (A) and GC content (B) of Enterococcus faecalis strains isolated from different niches. An asterisk (*) indicates a p-value <0.05; double asterisks (**) indicate a p-value <0.01 (one-way ANOVE test)

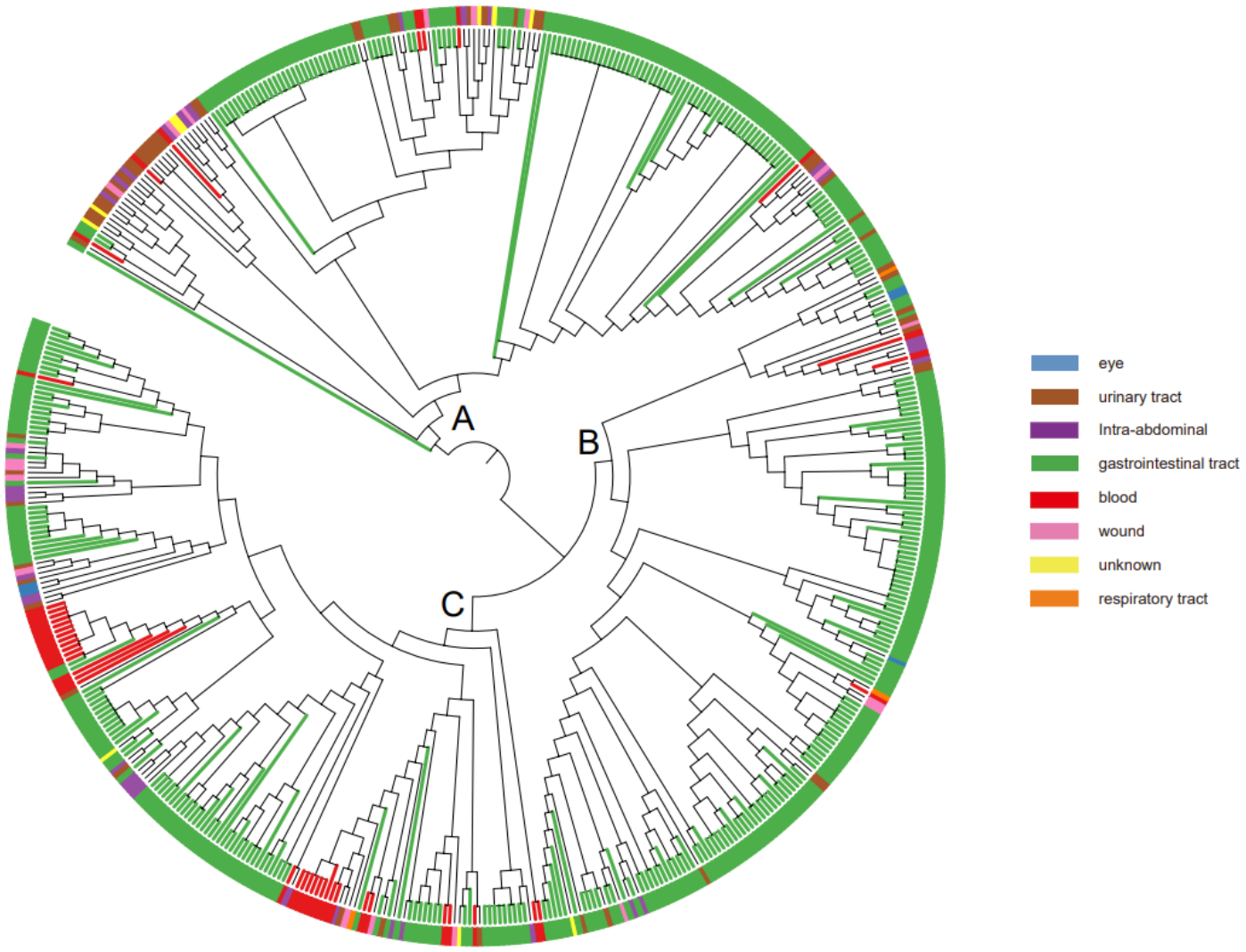

Figure 2

Phylogenetic tree constructed based on the core genes of Enterococcus faecalis isolates. The phylogenetic tree was constructed using the DNA sequences of xxx core genes of 537 isolates of E. faecalis. The color of out circle represents the origin of the isolate. Isolates derived from blood and GI tract are highlighted with red and green branches, respectively.

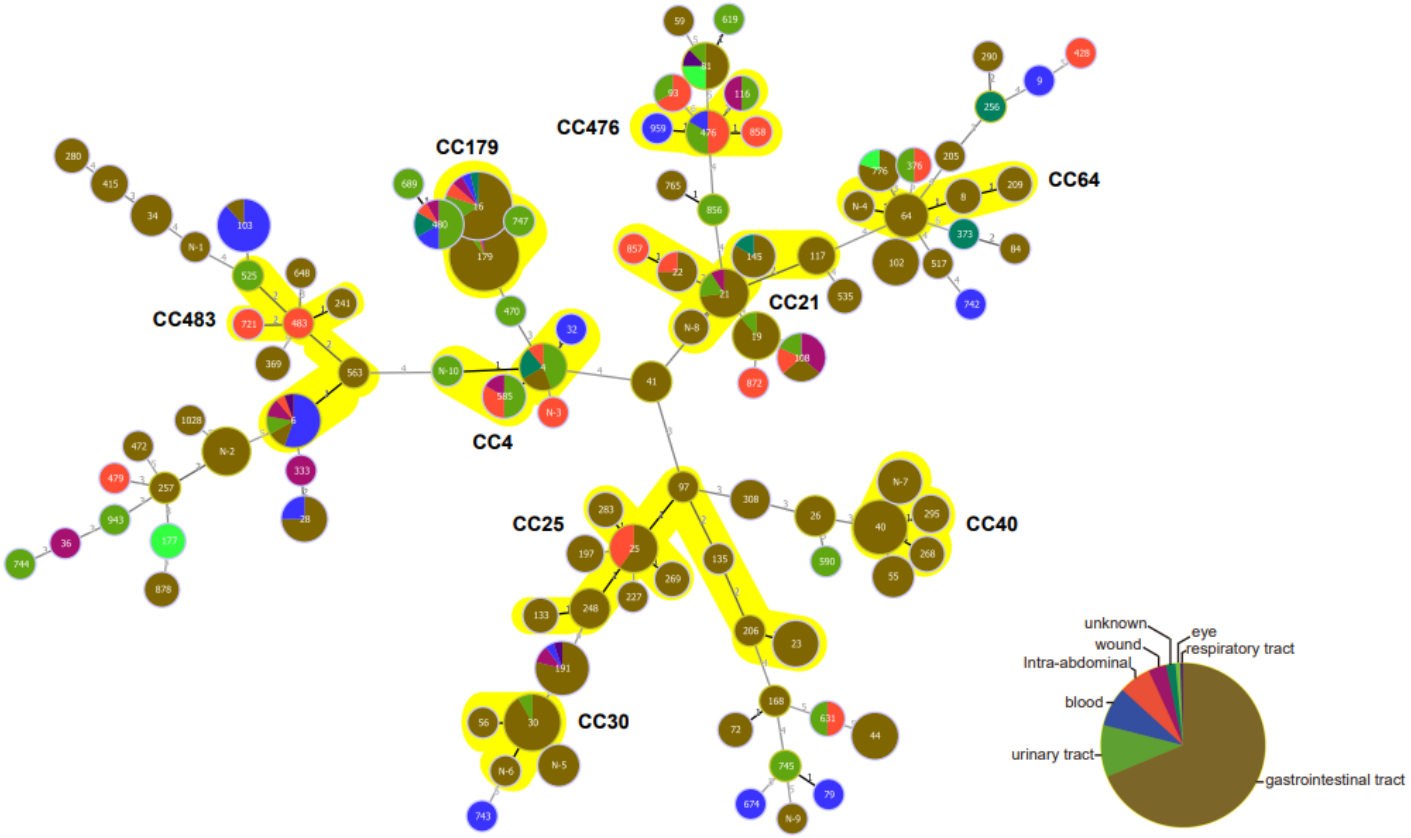

Figure 3

Clustering of 537 Enterococcus faecalis STs by use of eBURST. Colors indicate isolation sources. Each circle represents an ST, and the type number is indicated in the circle. The area of each circle corresponds to the number of isolates. Black solid lines connect single-locus variants, dark grey solid lines connect double-locus variants, and light gray solid lines connect STs that differ in three and more than three loci. Pie charts indicate ST distribution. Clonal complexes are indicated.

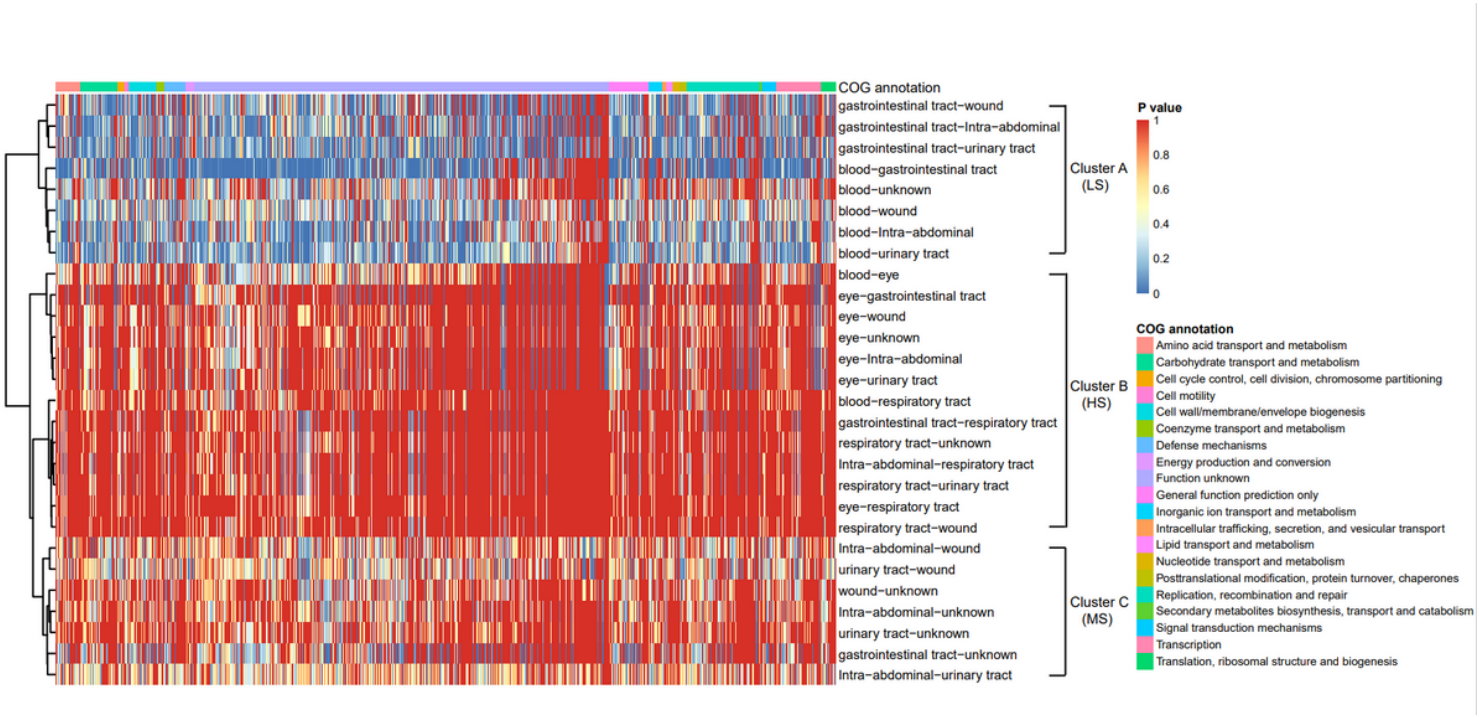

Figure 4

A matrix based on p-value of selected 2546 genes in 537 Enterococcus faecalis accessory genome. Three clusters could be identified: Cluster A (low similarity), Cluster B (high similarity) and Cluster C (middle similarity).

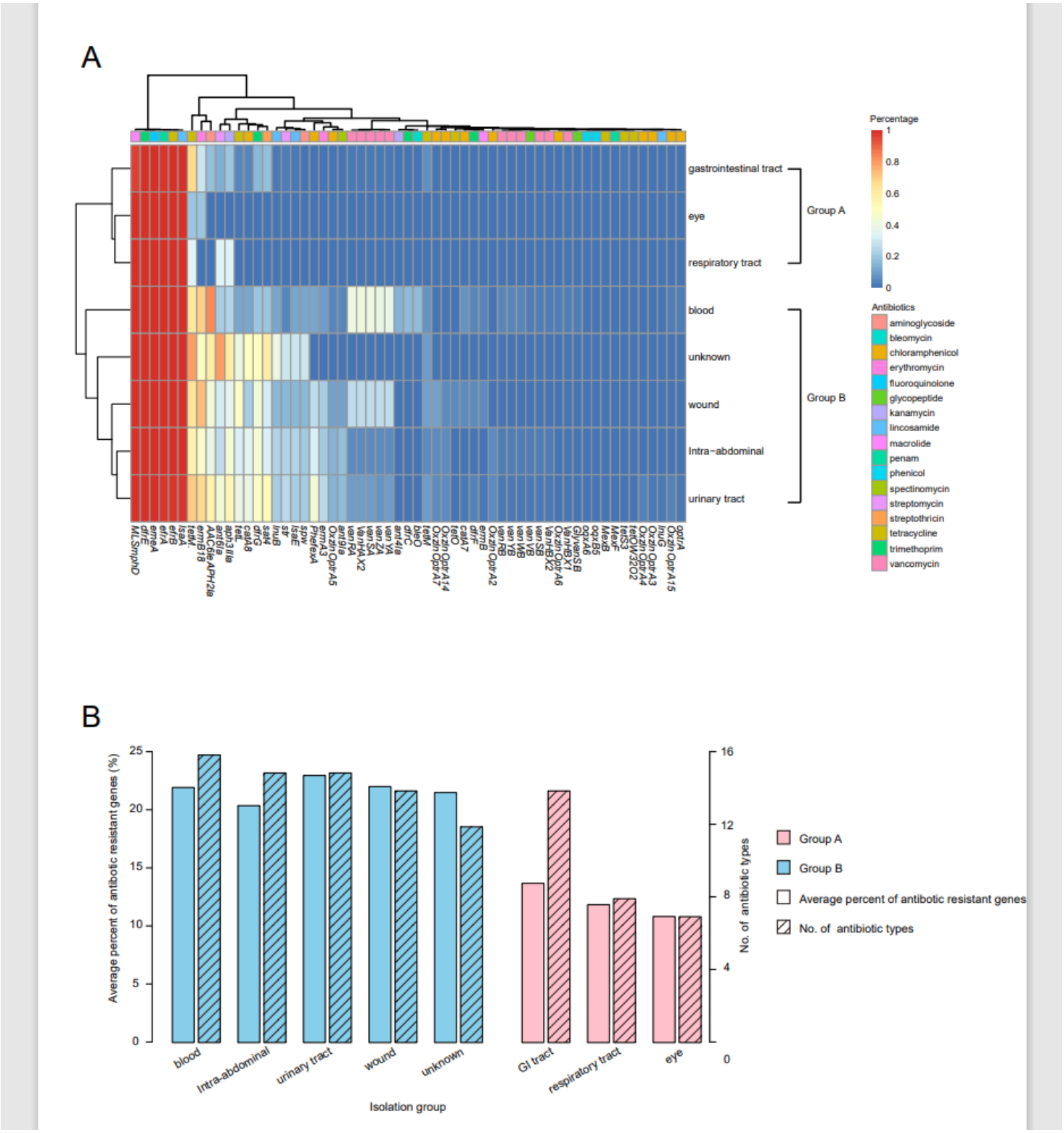

Figure 5

Distribution of antibiotic resistance genes in different sources of 537 Enterococcus faecalis genomes. (A) A box indicates that the presence percentage of a specific antibiotic resistance genes. Two clusters could be identified based on the distribution of antibiotic resistance genes: Group A and B. (B) the average percent of antibiotic resistance genes and number of involved antibiotic types in eight isolation source groups.

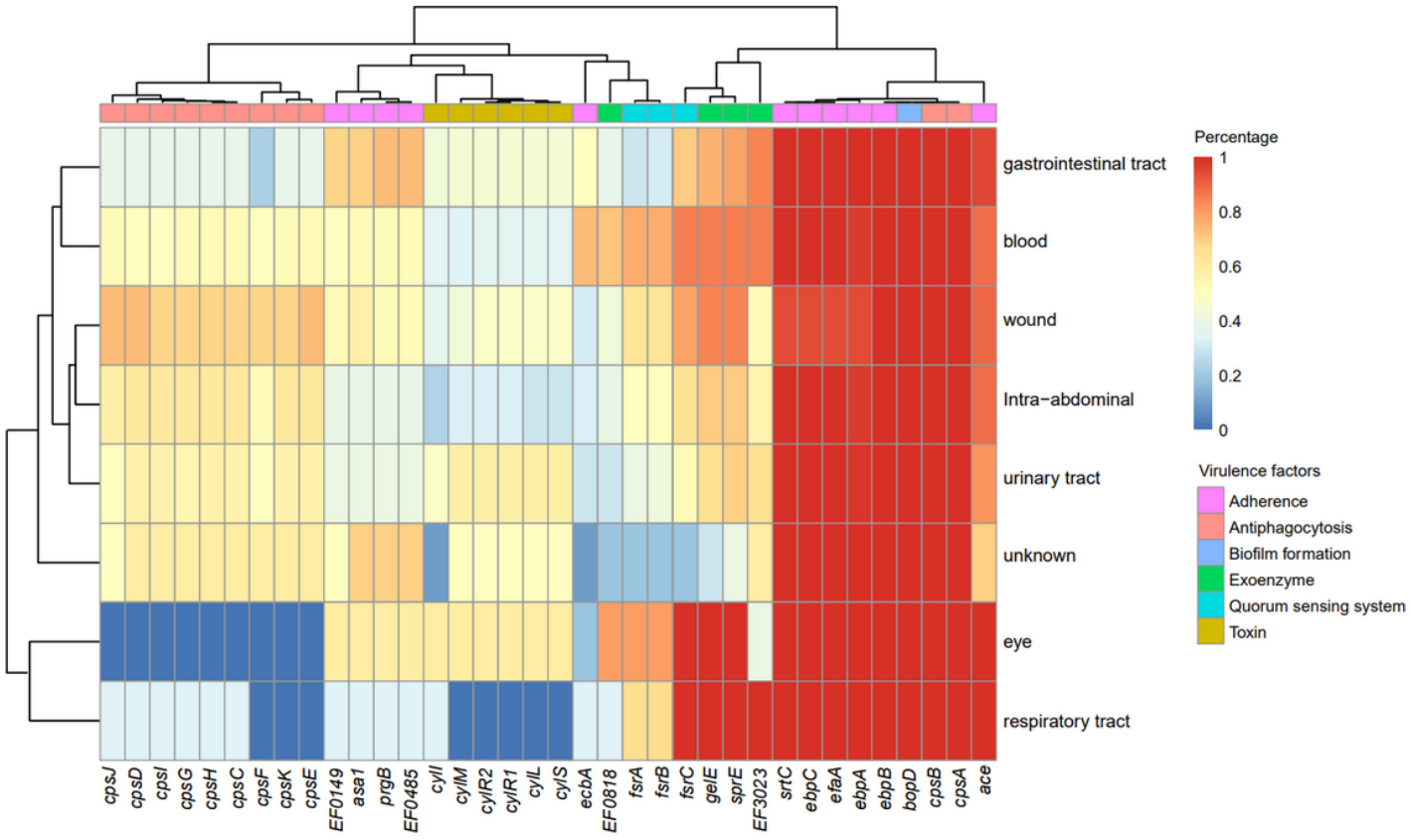

Figure 6

Distribution of virulence genes in different sources of 537 Enterococcus faecalis genomes. A box indicates that the presence percentage of a specific virulence genes.

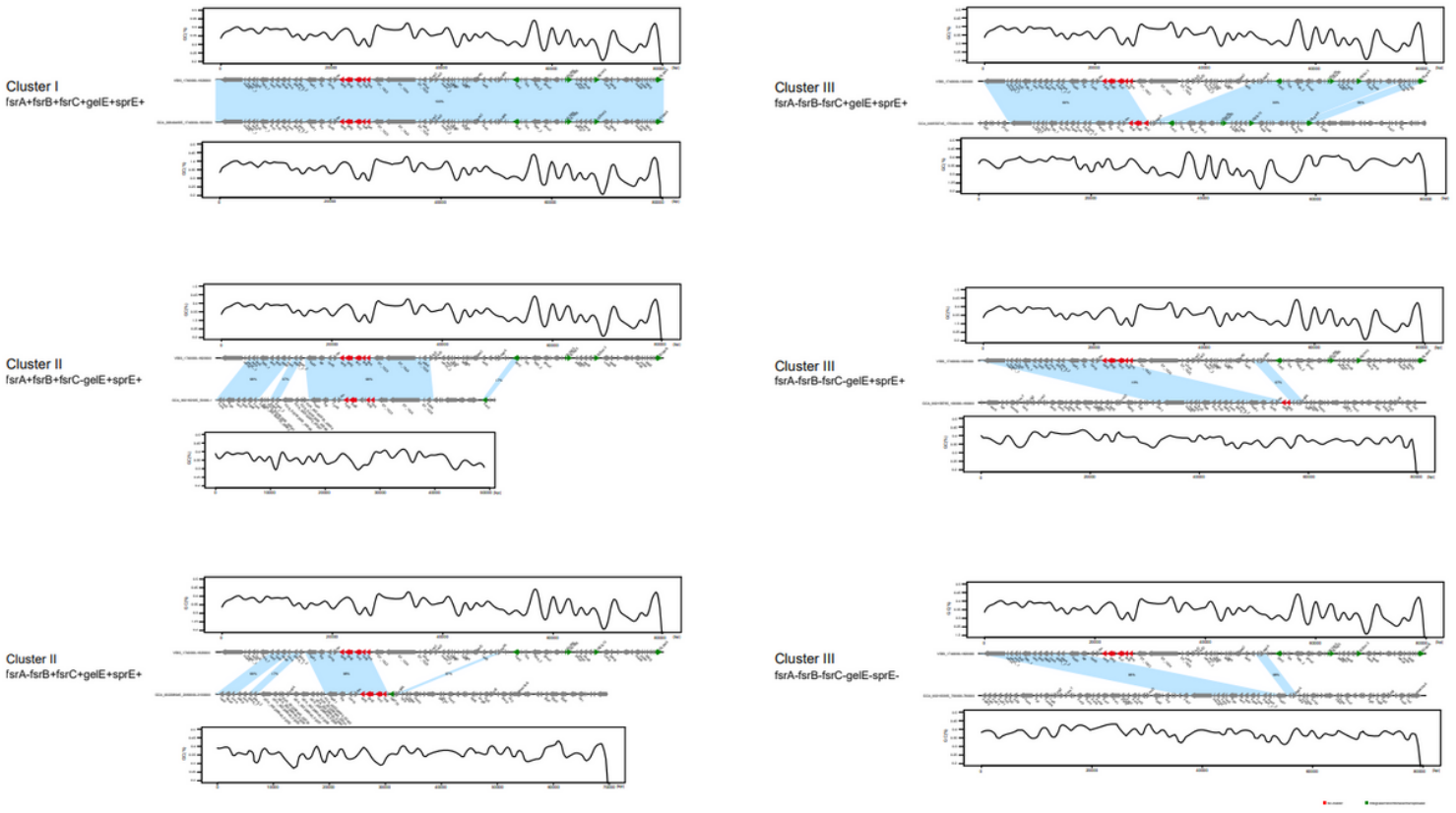

Figure 7

Structure clusters of the fsr-gelE region in Enterococcus faecalis genomes. The five genes (fsrA, fsrB, fsrC, gelE and sprE) associated fsr quorum sensing system and Integrase/recombinase/transposase are highlighted with red and green arrows, respectively.

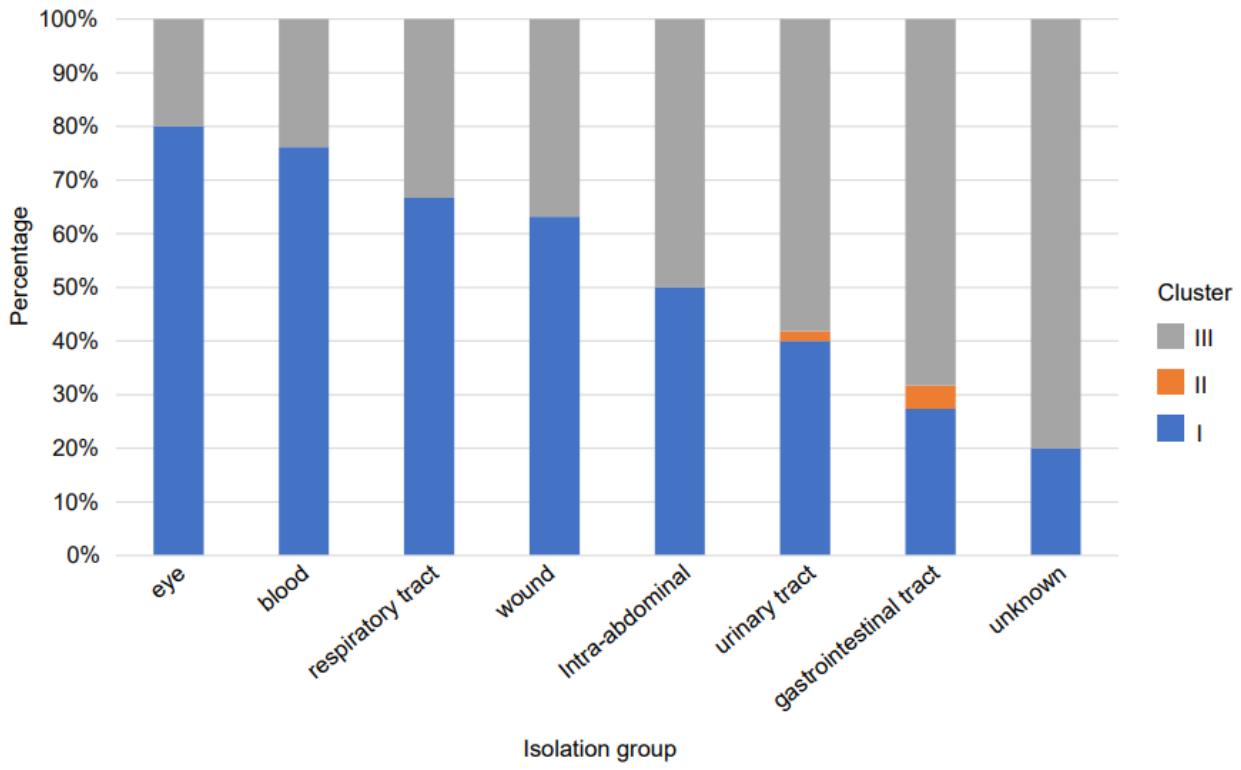

Figure 8

Distribution of structure clusters of fsr-gelE region in Enterococcus faecalis from different sources. The Cluster I/II/III are colored by blue, orange and grey respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx