AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Spectrum

| Computational Biology | Research Article

The population-level impact of *Enterococcus faecalis* genetics on intestinal colonization and extraintestinal infection

Chrispin Chaguza,[1,2] Anna K. Pöntinen,[3,4] Janetta Top,[5] Sergio Arredondo-Alonso,[2,3] Ana R. Freitas,[6,7,8] Carla Novais,[6,7] Carmen Torres,[9] Stephen D. Bentley,[2] Luisa Peixe,[6,7] Teresa M. Coque,[10,11] Rob J. L. Willems,[5] Jukka Corander[2,3,12]

AUTHOR AFFILIATIONS See affiliation list on p. 17.

ABSTRACT *Enterococcus faecalis* is a commensal bacterium of the human gastrointestinal tract that causes opportunistic infections. The *E. faecalis* genetic changes associated with pathogenicity, particularly gut-to-bloodstream translocation, remain poorly understood. Here, we performed a genome-wide association study (GWAS) of 736 whole-genome sequences of fecal and bloodstream *E. faecalis* isolates from hospitalized and nonhospitalized individuals, respectively, to identify *E. faecalis* genetic signatures associated with the patient's hospitalization status and body isolation source. We found that infection by hospitalization status and extraintestinal infection are heritable traits, with ~40% and ~30% of their variation explained by *E. faecalis* genetics, respectively. Furthermore, a GWAS using linear mixed models did not pinpoint any clear overrepresentation of individual genetic changes by hospitalization status or body isolation source after controlling for the population structure. However, we observed elevated signals in a genomic region containing a prophage element. However, the lineages themselves and their associated virulence factors and antibiotic resistance genes showed variable frequency among blood and fecal isolates and in hospitalized and nonhospitalized individuals. Altogether, our findings indicate that *E. faecalis* infection by hospitalization status and body sites is partially influenced by the overall genetic background of the isolates and antibiotic resistance patterns rather than genetic variation at individual loci, which suggests a greater role of other host and environmental factors and ultimately the opportunistic pathogenic lifestyle of this versatile host generalist bacterium.

IMPORTANCE *Enterococcus faecalis* causes life-threatening invasive hospital- and community-associated infections that are usually associated with multidrug resistance globally. Although *E. faecalis* infections cause opportunistic infections typically associated with antibiotic use, immunocompromised immune status, and other factors, they also possess an arsenal of virulence factors crucial for their pathogenicity. Despite this, the relative contribution of these virulence factors and other genetic changes to the pathogenicity of *E. faecalis* strains remain poorly understood. Here, we investigated whether specific genomic changes in the genome of *E. faecalis* isolates influence its pathogenicity—infection of hospitalized and nonhospitalized individuals and the propensity to cause extraintestinal infection and intestinal colonization. Our findings indicate that *E. faecalis* genetics partially influence the infection of hospitalized and nonhospitalized individuals and the propensity to cause extraintestinal infection, possibly due to gut-to-bloodstream translocation, highlighting the potential substantial role of host and environmental factors, including gut microbiota, on the opportunistic pathogenic lifestyle of this bacterium.

KEYWORDS microbial genomics, infectious disease, genome-wide association study, bacteria

Editor Rosario Gil, University of Valencia, Paterna, Valencia, Spain

Address correspondence to Chrispin Chaguza, Chrispin.Chaguza@yale.edu, or Jukka Corander, Jukka.Corander@medisin.uio.no.

The authors declare no conflict of interest.

See the funding table on p. 17.

Received 13 January 2023
Accepted 29 August 2023
Published 9 October 2023

Enterococcus faecalis is a versatile generalist commensal bacterium that colonizes the gastrointestinal tract and other niches in humans and animals and survives in the environment, including nosocomial settings (1). *E. faecalis* is a subdominant core member of the human gut microbiota, usually acquired early after birth, and its origin dates to the Paleozoic era ~400 to 500 million years ago (2). Although *E. faecalis* predominantly exhibits a commensal lifestyle, it is a conditional or opportunistic pathogen (3, 4). It causes life-threatening opportunistic infections, including bacteremia, endocarditis, intra-abdominal infection, pneumonia, and meningitis infections typically associated with high mortality (5, 6). Since the 1970s, *E. faecalis* has emerged as a leading cause of community-acquired and nosocomial infections, most of which have become increasingly difficult to treat due to intrinsic and acquired antibiotic resistance, making it a major threat to public health globally (4, 6–9). Such increasing antibiotic resistance has reignited calls to develop enterococcal vaccines.

The commensal-to-pathogenic switch of *E. faecalis* is marked by its overgrowth in the gut and subsequently translocation into the bloodstream via the intestinal epithelium (10). Such extraintestinal translocation can lead to bacteremia, infective endocarditis, and infections in other distal tissues from the intestines. However, the specific mechanisms driving *E. faecalis* bloodstream invasion, survival, and virulence are still being uncovered (3, 5, 11, 12). Observational studies have shown that antibiotics, such as cephalosporins, promote overgrowth and extraintestinal translocation of *E. faecalis* into the bloodstream (13, 14), an observation supported by *in vivo* murine experimental models (14–16). Such overgrowth of *E. faecalis* reflects the impact of ecological side effects of broad-spectrum antibiotics in driving dysbiosis of the gut microbiota, a phenomenon similarly observed with *Clostridioides difficile* (formerly known as *Clostridium difficile*) (17, 18). *E. faecalis* also harbors a diverse arsenal of putative virulence factors (19–21), which foster its adaptation and survival in the harsh clinical and midgut environments and potentially promote extraintestinal translocation into the bloodstream. These virulence factors appear to be enriched in the dominant epidemic *E. faecalis* lineages (22, 23), highlighting their importance to the success of these clones. For example, the gelatinase (*gelE*) gene encodes a metalloprotease exoenzyme commonly associated with epidemic clones (22) and is important for infective endocarditis (24) and extraintestinal translocation into the bloodstream (25). Other exotoxins, namely, hemolysin and enterococcus surface protein, are also important for virulence in endocarditis (26) and biofilm formation (27), respectively, although the role of the former on intestinal colonization and translocation has been questioned (28, 29). Acquisition of extrachromosomal elements, including pathogenicity islands (30, 31) and plasmids (32), has also been associated with virulence and survival in nosocomial settings (33). Understanding the distribution of these known and novel *E. faecalis* virulence factors in strains sampled from different tissues and individuals with contrasting pathogenicity could potentially reveal mechanisms for enterococcal pathogenicity and uncover therapeutic targets.

Remarkable advances in whole-genome sequencing and computational biology have revolutionized population genomics since the sequencing of the first enterococcal genome (34). To date, the feasibility of large-scale whole-genome sequencing and analysis has facilitated detailed population-level studies to uncover the genetic basis of bacterial phenotypes (35). For example, the application of genome-wide association studies (GWAS) to bacteria has revealed genetic variants associated with diverse phenotypes, including antimicrobial resistance (36), host adaptation (37), and pathogenicity (38). A key feature of the GWAS approach is that it can identify novel genetic variants associated with phenotypes through systematic genome-wide screening, which does not bias the analysis toward "favorite" genes and mutations commonly studied in different laboratories. Although previous studies have attempted to compare the genetic and phenotypic differences between *E. faecalis* isolates causing intestinal colonization and invasive disease (39), clinical and nonclinical strains (40), and isolates of diverse origins (41), these studies were limited by the small sample sizes and use of low-resolution molecular typing methods such as pulsed-field gel electrophoresis. Recent studies

of *E. faecalis* and *Enterococcus faecium* species identified unique mutations associated with outbreak strains, highlighting the potential effects of specific genetic changes on pathogenicity (12, 42). Despite the increasing affordability of population-scale microbial sequencing, the genetic basis of *E. faecalis* infection in individuals with different hospitalization statuses, i.e., pathogenicity and extraintestinal infection, including those due to extraintestinal translocation, remains poorly understood. The application of GWAS approaches to discover the genetic changes driving the pathogenicity and virulence of *E. faecalis* could expedite antibiotic and vaccine development.

Here, we leveraged a collection of 736 whole-genome sequenced *E. faecalis* isolates sampled from the feces and blood specimens of hospitalized and nonhospitalized individuals (43). We undertook a GWAS of the isolates to investigate if specific genomic variations, including single-nucleotide polymorphisms (SNPs) and insertions/deletions, were associated with infection by hospitalization status and body isolation source. We show a predominantly higher differential abundance of virulence factors and antibiotic resistance in *E. faecalis* isolates from hospitalized than from nonhospitalized individuals, as well as isolates from blood than from feces. This largely reflects the effects of the genetic background or lineages, as no specific individual genetic changes showed population-wide effects on the infection of individuals by hospitalization status or isolation source. Additionally, we found that infection in individuals depends on their hospitalization status and extraintestinal infection, which are heritable traits partially explained by *E. faecalis* genetics. Altogether, our findings provide evidence suggesting that the collective effects of several genetic variants, genetic background or lineages, and gut ecological factors drive the pathogenicity and extraintestinal infection of *E. faecalis* rather than the population-wide effects of individual bacterial genetic changes. These findings have broader implications for *E. faecalis* disease prevention strategies, specifically the need to target all genetic backgrounds when designing vaccines to achieve optimal protection against severe enterococcal invasive diseases.

RESULTS

Clinical and genomic characteristics of *E. faecalis* isolates

To investigate the population genomics of *E. faecalis* pathogenicity, marked by infection of individuals by hospitalization status and body isolation source, we compiled a data set of 736 whole genome sequences of *E. faecalis* isolates sampled from blood and fecal specimens of hospitalized and nonhospitalized individuals between 1996 and 2016 (43) (Fig. 1A; Data Set S1). We included isolates from countries where both fecal and bloodstream isolates were collected, but not necessarily from the same individual. In total, our final data set comprised isolates from Europe: the Netherlands (n = 300) and Spain (n = 436) (Fig. 1B). By infection of individuals, 485 isolates were obtained from hospitalized patients, while 251 isolates were sampled from nonhospitalized individuals (Data Set S1). Regarding the isolation of *E. faecalis* from human body sites, 440 isolates were sampled from the blood, while 296 isolates were from feces. Nearly all the isolates from nonhospitalized individuals were collected from feces, while those from blood were from hospitalized individuals. Such a discrepancy in sampling of the *E. faecalis* isolates by hospitalization status and body isolation source reflected the fact that invasive sampling, such as collecting blood samples, was less likely to be performed for the nonhospitalized than the hospitalized patients. In addition, considering that *E. faecalis* is a major cause of nosocomial infections, there is a greater likelihood that the isolation of *E. faecalis* in hospitalized individuals may be a consequence of acquisition in the hospital environment by already hospitalized individuals with weaker immunity rather than only a reflection of its intrinsic pathogenicity (Fig. 1C).

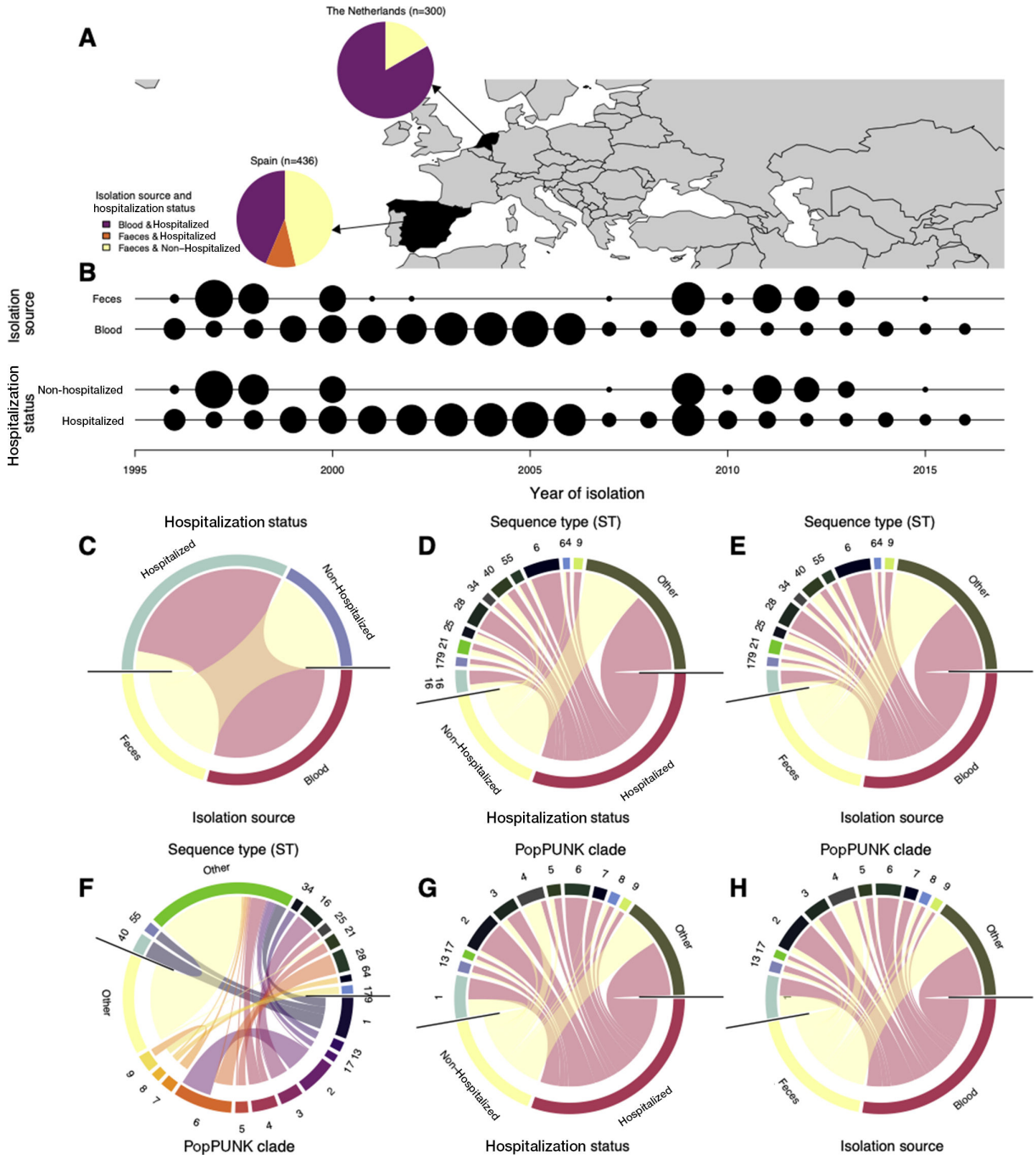

FIG 1 Characteristics of *E. faecalis* isolates included in this study. (A) Summary of the convenient sample of *E. faecalis* isolates collected from individuals in the Netherlands and Spain, showing the frequency of isolates from hospitalized and nonhospitalized individuals, and blood and feces. The map was generated using the R package rworldmap. (B) Temporal distribution of the *E. faecalis* isolates in each country. The radius of each black circle represents the square root of the number of isolates selected per year for whole-genome sequencing. (C) Association between *E. faecalis* isolates by the sampling site and pathogenicity or hospitalization status. (D) Association of the *E. faecalis* isolates by hospitalization status and sequence type (ST) based on the multi-locus sequence typing scheme approach. (E) Association of the *E. faecalis* isolates by ST and isolation source. (F) Association of the *E. faecalis* isolates by STs and clades or lineages defined using Population Partitioning Using Nucleotide K-mers (PopPUNK) by Pöntinen et al. (43). (G) Association of the *E. faecalis* isolates by hospitalization status and PopPUNK clades or lineages. (H) Association of the *E. faecalis* isolates by body isolation source and the PopPUNK clades or lineages.

E. faecalis infections by hospitalization status and body isolation source are heritable phenotypes predominantly explained by genetic background or lineages

To assess the overall genetic basis of the infection in individuals with different hospitalization statuses, we quantified the proportion of the variability in the phenotypes explained by *E. faecalis* genetics. We calculated the narrow-sense heritability (h^2) based on the kinship matrix generated using unitig sequences (44). After adjusting for the geographical origin and year of isolation of the isolates, we found a heritability of h^2 = 0.40 [95% confidence interval (CI): 0.23 to 0.57] and h^2 = 0.30 (95% CI: 0.15 to 0.45) for infection by hospitalization status and body isolation source, respectively. Next, we calculated the heritability for infection of individuals by hospitalization status and body isolation source using only the Spanish cohort, which had an even number of isolates from hospitalized and nonhospitalized individuals as well as from blood and feces. We found consistent, but slightly higher, estimates of heritability for both infection of individuals by hospitalization status (h^2 = 0.43, 95% CI: 0.23 to 0.63) and body isolation source (h^2 = 0.28, 95% CI: 0.12 to 0.45) than estimated based on the combined data set. These findings suggest that *E. faecalis* infections by hospitalization status and body isolation source are moderately heritable traits partially explained by genetics.

Infections of individuals with *E. faecalis* by hospitalization status and body isolation source vary across lineages

We sought to investigate the distribution of the hospitalization and body isolation source phenotypes in the context of the *E. faecalis* population structure. We generated a maximum-likelihood phylogenetic tree using 251,983 core genome SNPs, exclusively containing nonambiguous nucleotide and deletion characters, and annotated it with the hospitalization status and body isolation source phenotypes. The isolates were widely distributed across different genetic backgrounds based on the country of origin as well as body isolation source and hospitalization status, a finding consistent with the literature that the severity of *E. faecalis* infections is not restricted to specific lineages, in contrast to the genetic separation between commensal and hospital-adapted lineages observed in *E. faecium* (45, 46) (Fig. 2). We then performed an in-depth analysis of the *E. faecalis* population structure using lineage definitions based on the Population Partitioning Using Nucleotide K-mers (PopPUNK) genomic sequence clustering framework (47) by Pöntinen et al. (43). Our isolates clustered into 96 clades, which corresponded to 121 sequence types (STs) or clones defined by the *E. faecalis* multi-locus sequence typing scheme (MLST) (48) (Fig. 2). There was no single dominant ST present at a frequency of >50% compared to the others among the isolates sampled from hospitalized and nonhospitalized patients and isolates from feces and blood ($P > 0.05$) (Fig. 1D and E). As expected, the clusters defined by the MLST scheme were concordant with the PopPUNK clades or lineages, although the latter were less granular than the former as they are defined based on genome-wide variation and therefore are robust to subtle genomic variation (Fig. 1F). Therefore, as similarly observed with the STs, there was no single dominant clade present at a frequency of >50% compared to the rest associated with the *E. faecalis* isolates from hospitalized and nonhospitalized patients and isolates from feces and blood ($P > 0.05$) (Fig. 1G and H).

We then compared the relative frequency of individual STs and PopPUNK clades between isolates collected from hospitalized patients and nonhospitalized individuals. We found three clades more common in hospitalized patients than in nonhospitalized individuals, namely, clades 2 (adjusted $P = 1.20 \times 10^{-05}$), 6 (adjusted $P = 4.80 \times 10^{-08}$), and 7 (adjusted P = 0.0003). In contrast, two clades, clade 1 (adjusted P = 0.027) and clade 4 (adjusted $P = 3.40 \times 10^{-10}$), were more common in nonhospitalized individuals than in hospitalized patients (Fig. 3A; Table S1). Due to the correlation between the hospitalization status of the individuals and the isolation source, we found similar patterns in the relative abundance of the clades between blood and fecal isolates (Fig. 3B; Table S2). We found a higher abundance of ST6 (clade 2; adjusted $P = 1.50 \times 10^{-05}$), ST9

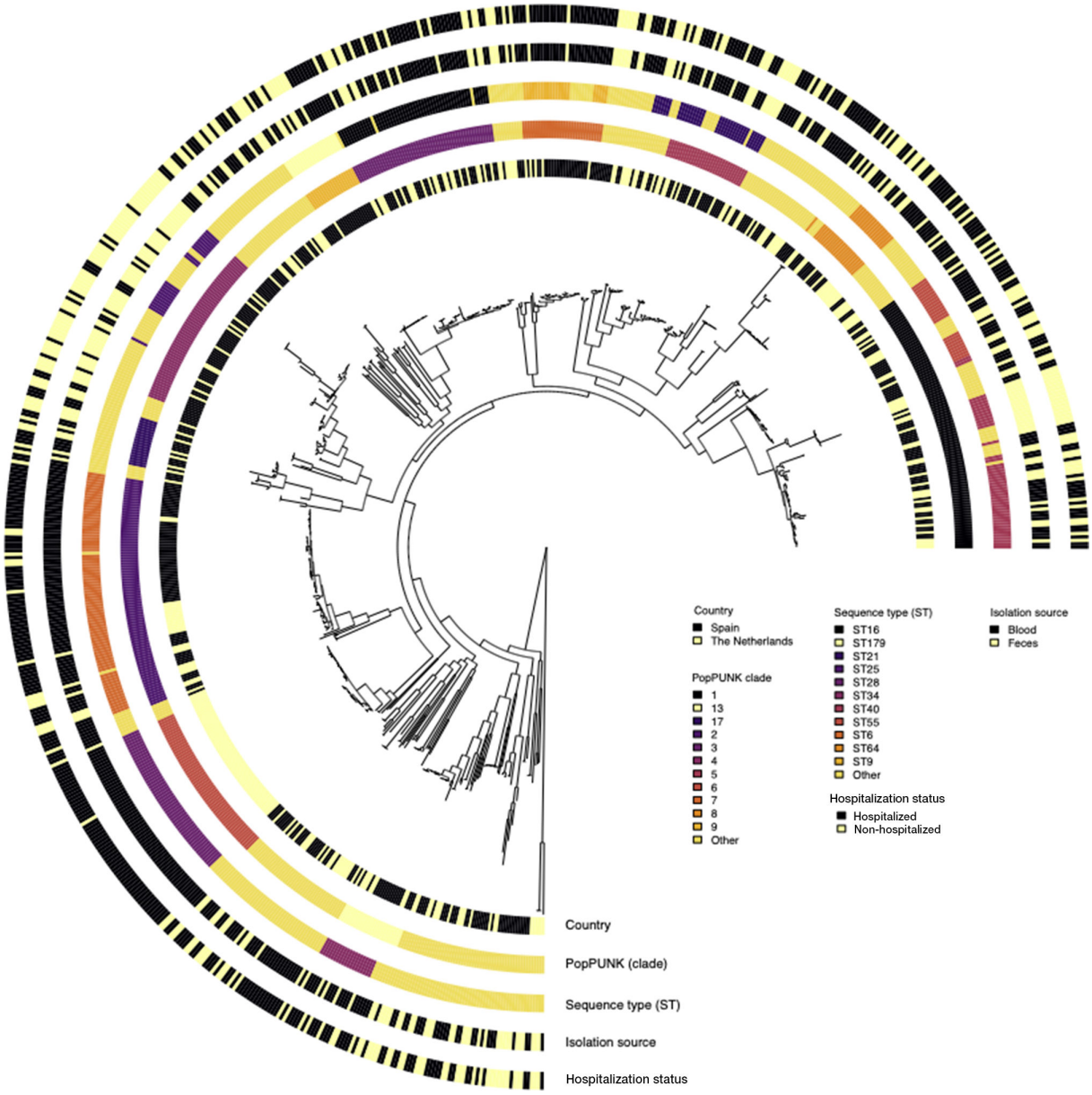

FIG 2 Maximum-likelihood phylogenetic tree of 736 *E. faecalis* isolates from the Netherlands and Spain. Each circular ring at the tip of the phylogenetic tree, from innermost to outermost, represents the country of origin for each *E. faecalis* isolate (the Netherlands and Spain), clade or lineage defined by the PopPUNK genomic sequence clustering framework (47) by Pöntinen et al. (43), ST based on the *E. faecalis* MLST scheme (48), body isolation source (blood and feces), and pathogenicity defined based on hospitalization status (hospitalized and nonhospitalized). The phylogeny was rooted at the midpoint of the longest branch between the two most divergent *E. faecalis* isolates.

(clade 7; adjusted P = 0.0082), and ST28 (clade 6; adjusted $P = 1.20 \times 10^{-08}$) among hospitalized patients than among nonhospitalized individuals (Fig. 3C and D; Table S2). Similar patterns were observed among isolates sampled from blood compared to feces. Conversely, we found that ST25 (clade 4) was enriched in nonhospitalized patients compared to nonhospitalized individuals (adjusted P = 0.014) as well as in isolates sampled from blood compared to feces (adjusted $P = 7.80 \times 10^{-05}$) (Fig. 3C and D; Table S2). Together, these findings suggest that certain *E. faecalis* genetic backgrounds are overrepresented in patients by hospitalization status and isolation source, suggesting that the lineage, which correlates with the presence of virulence factors and antibiotic resistance determinants, partially influences extraintestinal infection of *E. faecalis*.

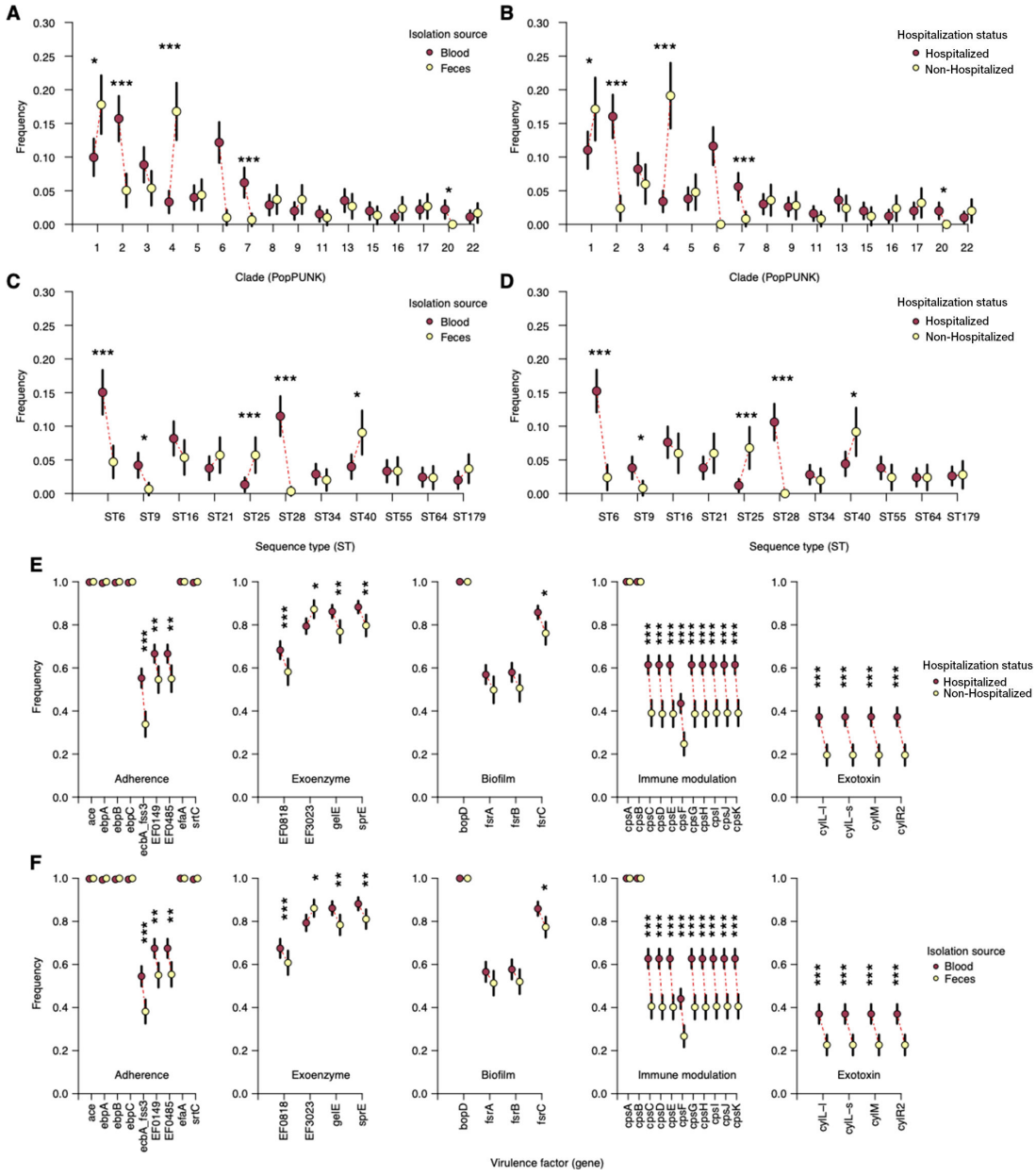

FIG 3 Relative abundance of *E. faecalis* lineages and virulence factors by hospitalization status and body isolation source. (A) Relative frequency of *E. faecalis* clades or lineages among isolates collected from blood and feces. (B) Relative frequency of *E. faecalis* clades or lineages among isolates collected from hospitalized and nonhospitalized individuals. (C) Relative frequency of *E. faecalis* ST among isolates collected from blood and feces. (D) Relative frequency of *E. faecalis* STs among isolates collected from hospitalized and nonhospitalized individuals. (E) Relative frequency of a catalog of known *E. faecalis* virulence factors from the virulence factor database (VFDB) (49) among hospitalized and nonhospitalized individuals. (F) Relative frequency of *E. faecalis* virulence factors from VFDB among isolates collected from blood and feces. All the error bars in each plot represent 95% binomial proportional confidence intervals. The asterisks above the frequency of some genes show the statistical significance of the difference in proportions based on the test for the equality of two proportions defined as follows: $P < 0.001$ (***), $P < 0.01$ (**), and $P < 0.05$ (*).

Only a few virulence factors show variable prevalence in individuals with different hospitalization statuses and isolation sources

As a generalist host species, *E. faecalis* exhibits high levels of recombination (48), which may facilitate the acquisition of genes promoting colonization and virulence, driving the success of its clones (23) We hypothesized that certain known virulence factors would be enriched among *E. faecalis* isolates from hospitalized patients, especially those with

bloodstream infection, compared to nonhospitalized individuals without bloodstream infection. We used a candidate gene approach to compare the enrichment of a catalog of *E. faecalis* virulence factors obtained from the virulence factor database (VFDB) (49) by individuals' hospitalization status and body isolation source. We found that three genes, namely, ecbpA (adjusted $P = 5.70 \times 10^{-08}$), EF0149 (adjusted $P = 0.0019$), and EF0485 (adjusted $P = 0.0026$), which play a role in epithelial surface adherence, were more common in extraintestinal infection than in intestinal colonization (Fig. 3E and F; Table S3). No genes encoding known exoenzyme and biofilm-associated proteins showed differential enrichment in either hospitalized patients relative to nonhospitalized individuals or extraintestinal infection compared to intestinal colonization (Fig. 3E and F; Table S3). However, all four exotoxin-encoding genes were enriched in hospitalized compared to nonhospitalized individuals, namely, *cylL-l* (adjusted $P = 1.20 \times 10^{-06}$), *cylL-s* (adjusted $P = 1.20 \times 10^{-06}$), *cylM* (adjusted $P = 1.20 \times 10^{-06}$), and *cylR2* (adjusted $P = 1.20 \times 10^{-06}$) (Fig. 3E; Table S3). Similar patterns were observed among the isolates sampled from extraintestinal infection and intestinal colonization (Fig. 3F; Table S3). Additionally, nine capsule biosynthesis genes (*cpsC* to *cpsK*) were more common among hospitalized than among nonhospitalized individuals as well as isolates from the extraintestinal infection than from the intestinal colonization (Fig. 3E and F; Table S3). These findings are partly consistent with previous studies (22, 23), although the present study investigated a larger catalog of virulence factors. Therefore, we conclude that certain virulence factors are associated with individuals with different hospitalization statuses and possibly promote extraintestinal translocation of *E. faecalis* into the bloodstream in hospitalized individuals.

Distribution of antibiotic resistance genes in *E. faecalis* isolates by hospitalization status and body isolation source

Hospitalized patients are more exposed to antibiotics in hospitals than nonhospitalized individuals, as more antibiotics are used in hospital settings than outside. Therefore, it is likely that *E. faecalis* isolates from hospitalized patients are more likely to have acquired resistance than isolates from nonhospitalized individuals. Because most patients were probably hospitalized because of other complaints and developed the *E. faecalis* infection during hospitalization, we hypothesized that *E. faecalis* isolates sampled from hospitalized individuals and extraintestinal infection would show a higher frequency of antibiotic resistance traits than isolates from nonhospitalized individuals and intestinal colonization. The rationale behind this hypothesis was that antibiotic-susceptible *E. faecalis* strains are more likely to be cleared from the gut following antibiotic use, leaving more space for the surviving antibiotic-resistant strains to cause extraintestinal infection and subsequently cause severe disease (Fig. 1C; Table S4). This would be due to the surviving antibiotic-resistant strains. We investigated this hypothesis by comparing the abundance of antibiotic resistance genes for seven antibiotic classes, namely, glycopeptides (vancomycin), aminoglycosides, macrolides, tetracyclines, phenicols, and oxazolidinones (linezolid), in *E. faecalis* isolates from hospitalized and nonhospitalized individuals, blood, and feces. Regressing the number of antibiotic classes susceptible to the hospitalization status while adjusting for the country of origin showed resistance to more antibiotic classes among isolates from hospitalized than from nonhospitalized individuals (effect size $\beta = 1.43$, $P < 3.63 \times 10^{-12}$) (Fig. 4A; Table S4). As expected, due to the correlation between hospitalization status and body isolation source (Fig. 1C; Table S4), we found a similar pattern for isolation source, i.e., isolates from blood harboring resistance traits to a higher number of antibiotic classes than isolates from feces (effect size $\beta = 1.37$, $P = 2.89 \times 10^{-10}$) (Fig. 4B; Table S4). Next, we compared the relative abundance of genotypically antibiotic-resistant isolates for each antibiotic class among *E. faecalis* isolates from hospitalized and nonhospitalized individuals. We found a higher relative abundance of genotypically inferred antibiotic-resistant isolates in hospitalized than in nonhospitalized individuals for aminoglycosides (adjusted $P = 5.51 \times 10^{-09}$), macrolides (adjusted $P = 3.77 \times 10^{-09}$), phenicols (adjusted $P = 8.59 \times 10^{-05}$),

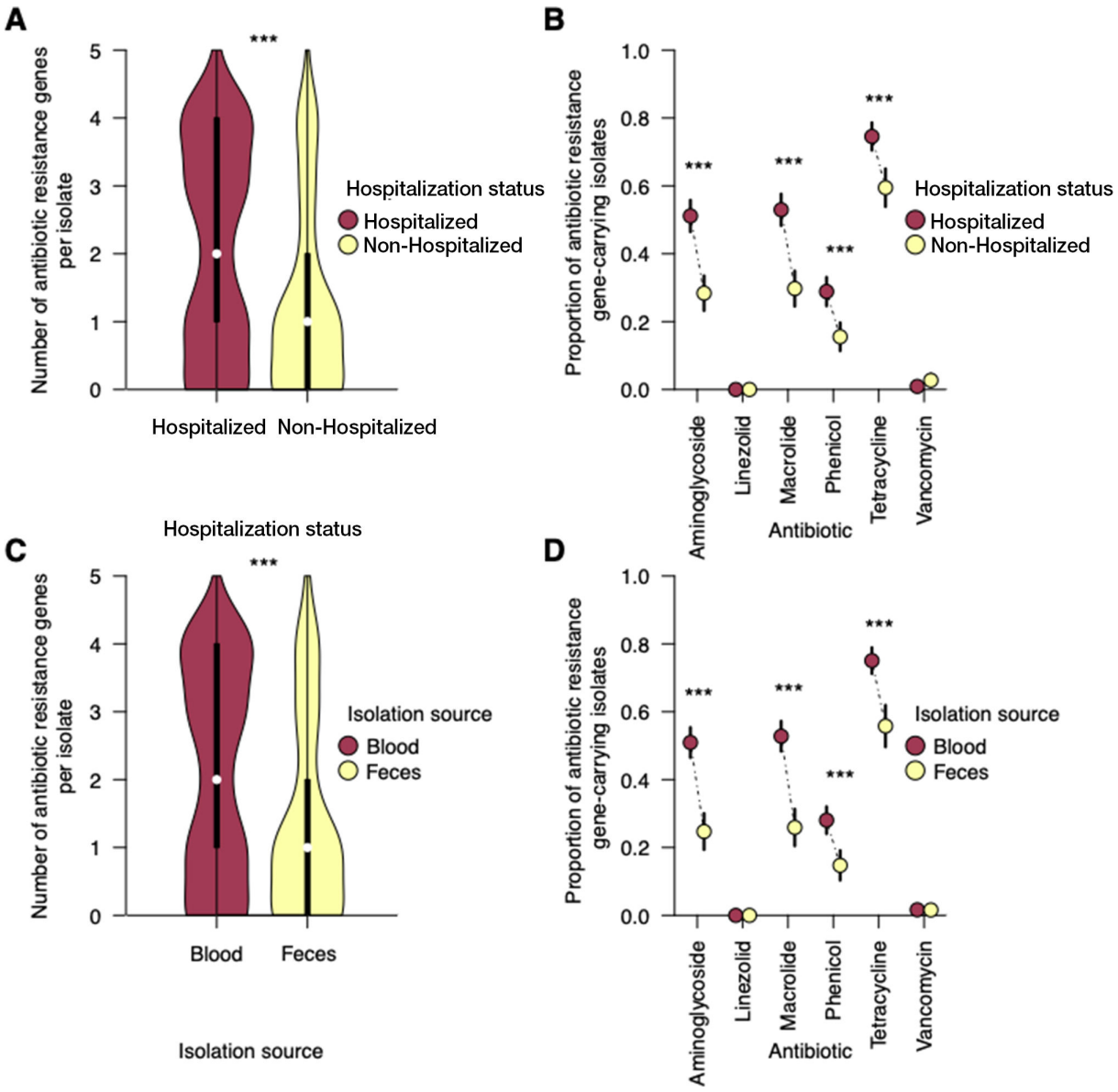

FIG 4 The abundance of *E. faecalis* antibiotic resistance genes by hospitalization status and body isolation source. (A) Distribution of the number of antibiotic resistance genes (see Materials and Methods) per *E. faecalis* isolates from hospitalized and nonhospitalized individuals. (B) Relative abundance or frequency of genotypically resistant *E. faecalis* isolates from hospitalized and nonhospitalized individuals. (C) Distribution of the number of antibiotic resistance genes per *E. faecalis* isolates collected from blood and feces. (D) Relative abundance or frequency of genotypically resistant *E. faecalis* isolates collected from blood and feces. All the error bars in each plot represent 95% binomial proportional confidence intervals.

and tetracyclines (adjusted $P = 6.75 \times 10^{-05}$). However, we observed no differences for glycopeptides, i.e., vancomycin (adjusted $P = 1$), which had almost negligible resistance in the isolates (Fig. 4C; Table S4). Again, we observed similar patterns in blood and fecal isolates (Fig. 4D; Table S4).

There is no evidence of population-wide effects of individual *E. faecalis* genetic changes on infection of individuals by hospitalization status and body isolation source

Having demonstrated differences in the prevalence of virulence factors, likely driven by lineage or strains' genetic background effects, we next undertook a GWAS using linear mixed models to identify individual *E. faecalis* genetic changes with population-wide

events on infection of individuals with varying hospitalization status. We hypothesized that genetic variation in known and unknown virulence factors would be disproportionately distributed among *E. faecalis* isolates from hospitalized and nonhospitalized individuals. In total, we selected 99,730 out of 252,278 SNP variants and 462,374 out of 1,089,909 unitig sequences, which capture variation in both the core and accessory genomes and are present at a frequency of 5% to 95% of the isolates for the GWAS. Contrary to our hypothesis, we found no statistically significant differences in the distribution of SNPs and unitigs between isolates from hospitalized and nonhospitalized individuals independent of the strain genetic background in the GWAS using linear mixed models (50) (Fig. 5A and B). Inspection of the quantile-quantile (Q-Q) plots revealed no issues with population structure (Fig. S1A and B). Altogether, these findings demonstrated that the infection of individuals with varying hospitalization status with *E. faecalis* is not driven by individual genetic changes independently of their genetic

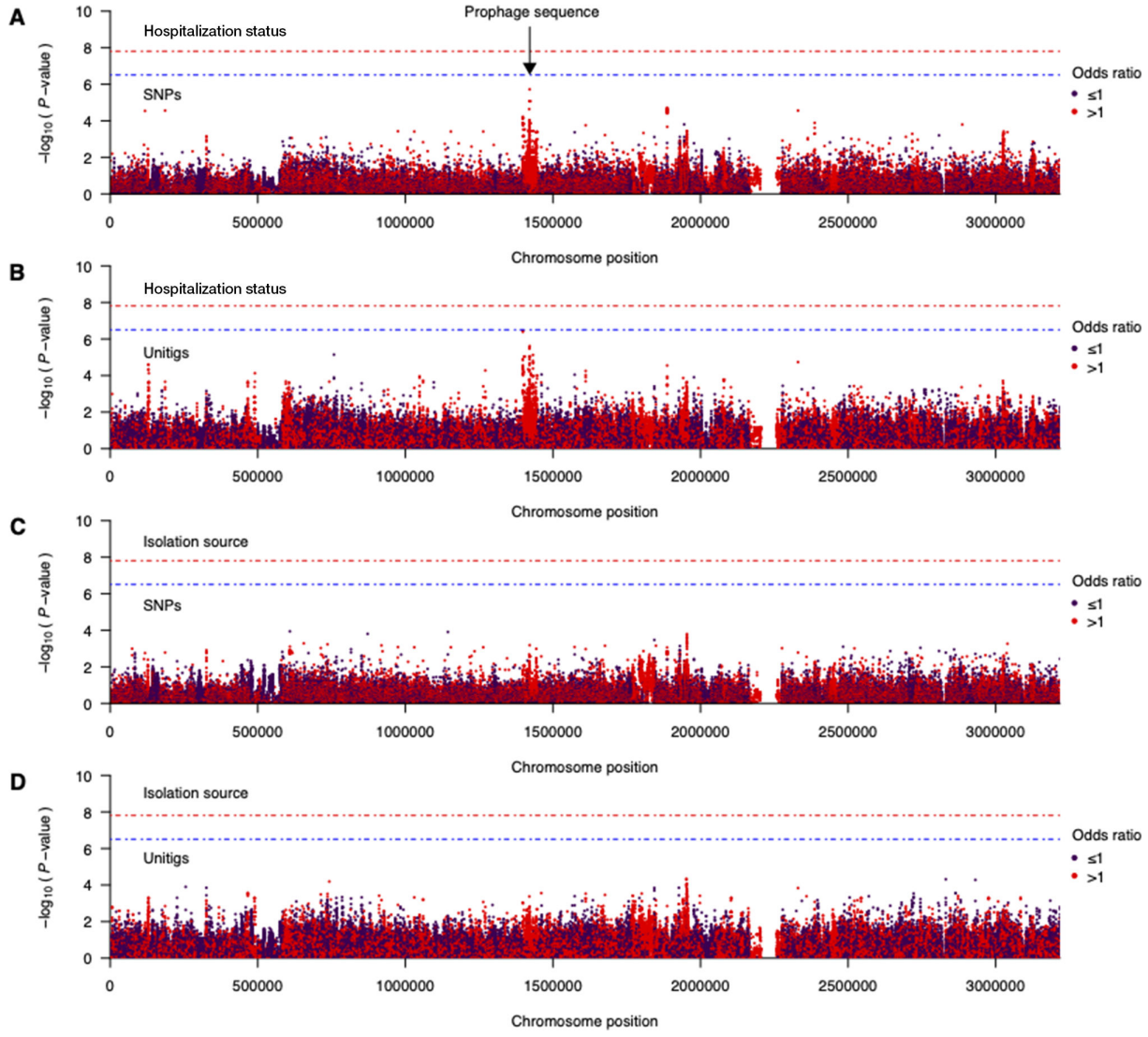

FIG 5 Association of *E. faecalis* genomic variants, hospitalization status, and body isolation source. (A) Manhattan plot summarizing the statistical association of SNP with pathogenicity or hospitalization status. The statistical significance of each SNP is log-transformed [$-\log_{10}$(*P*-value)] and plotted against its position in the V583 *E. faecalis* reference genome (34). (B) Manhattan plot summarizing the statistical association of unitigs with pathogenicity. (C) Manhattan plot summarizing the statistical association of SNPs with extraintestinal infection or body isolation source. (D) Manhattan plot summarizing the statistical association of unitigs with extraintestinal infection or body isolation source. The red and blue dotted lines represent the genome-wide significance and suggestive threshold, respectively.

background, suggesting that all *E. faecalis* strains are intrinsically adapted for extraintestinal infection, partly through translocation into the bloodstream.

We then carried out an additional GWAS to identify genetic changes associated with extraintestinal infection with *E. faecalis* strains by comparing fecal and bloodstream isolates. Like the GWAS based on the hospitalization status, we found no SNPs and unitigs associated with the isolation source independent of the strains' genetic background (Fig. 5C and D). However, we found the strongest signal in an ~48.1 Kb genomic region from positions ~1,390,000 to 1,450,000 bp in the V583 *E. faecalis* genome (34). Since horizontal gene transfer is a critical process in the mobilization of pathogenicity-associated genes (31, 51), we hypothesized that this region may represent a pathogenicity island. Re-annotation of the nucleotide sequence for this region revealed several phage-associated genes, which suggested the potential integration of a bacteriophage. Similarly, the Q-Q plots showed no issues with adjusting for the population structure (Fig. S1AB). We then performed phage prediction using the entire V583 *E. faecalis* genome sequence to annotate the SNPs and unitig sequences identified in the GWAS. We found a total of nine prophage sequences in the genome, including one with intact *attL*and *attR* attachment sites and integrase sequences located at genomic positions 1,398,051 to 1,446,151 bp. This prophage showed high genetic similarity to prophages including PHAGE_Entero_phiFL3A_NC_013648, PHAGE_Lister_B054_NC_009813 (27), and PHAGE_Lactob_LBR48_NC_027990. Furthermore, most of the phage-associated genes and protein sequences showed high genetic similarity to those found on prophages associated with several bacterial genera, including *Enterococcus*, *Lactobacillus*, *Bacillus*, *Listeria*, and *Staphylococcus*. These findings highlighted a potential virulence locus that should be prioritized for further investigation to understand its role in *E. faecalis* pathogenicity.

DISCUSSION

Tremendous advances in sequencing technology and analytical approaches have occurred over the past two decades since the sequencing of the first enterococcal genome—*E. faecalis* strain V583 (34). However, despite the increasing availability of population-level *E. faecalis* genomic data sets, no systematic studies have investigated the population-wide effects of individual genetic changes on infection in individuals with varying hospitalization status and extraintestinal infection and the overall contribution of *E. faecalis* genetics to these phenotypes (5). Such studies could reveal critical pathways for *E. faecalis* virulence, including survival in the bloodstream through evasion of innate host immune defenses, and inform the development of therapeutics (12). Here, we address this knowledge gap by investigating the effects of known and novel virulence factors, lineages, and the entire repertoire of *E. faecalis* genomic changes in a large collection of human fecal isolates, representing a snapshot of the *E. faecalis* diversity in the gut, and isolates sampled from blood specimens of individuals with different hospitalization statuses. Our findings demonstrate that the abundance of certain virulence and antibiotic resistance determinants is higher in *E. faecalis* isolates associated with severe disease and extraintestinal infection, largely driven by the effects of the strains, lineages, or genetic background effects but not the population-wide effects of individual genetic changes.

E. faecalis is a versatile pathogen that survives in a wide range of challenging niches, including the human gut, blood, and environment, such as in clinical settings. Such adaptation and survival of *E. faecalis* in these diverse environments are modulated by several mechanisms, including antimicrobial resistance (52), intracellular survival (53–56), and biofilm formation (27). Although several virulence factors of *E. faecalis* have been described (24–27), how (and if this happens) these factors contribute to infection of individuals with varying hospitalization status and extraintestinal infection, especially through gut-to-bloodstream translocation, remains poorly understood. Previous genetic studies shed light on how the distribution of virulence factors shapes the adaptation of *E. faecalis* clones to different environments despite the limitation of small sample sizes

(39, 41). In this study, we demonstrate enrichment of known virulence genes in isolates associated with different hospitalization statuses using a larger collection of isolates. These include genes encoding for aggregation substance adherence factors (EF0485 and EF0149) (32); lantipeptide cytolysin subunits CylL-L and CylL-S (*cylL-l* and *cylL-s*), cytolysin subunit modifier (*cylM*), and cytolysin regulator *R2* (*cylR2*) exotoxins (57); and polysaccharide capsule biosynthesis genes (*cpsC* to *cpsK*) involved in immune modulation or antiphagocytosis (58). These findings suggest that the variable abundance of these virulence genes in hospitalized and nonhospitalized individuals could influence *E. faecalis* pathogenicity, possibly because they primarily contribute to intestinal colonization, survival, and fitness or competitiveness in different intestinal compartments in the dysbiotic gut of hospitalized patients. Once the strains harboring these genes are established in higher numbers in the gastrointestinal tract, this promotes transmission, which in turn promotes the evolution and fixation of these virulence genes in the population. Interestingly, the observed higher antibiotic resistance, especially aminoglycosides, in isolates from blood and hospitalized individuals than from feces and nonhospitalized individuals suggests that antibiotic-resistant *E. faecalis* strains are more likely to survive and overgrow after the use of these antibiotics, consistent with findings reported elsewhere (14–16, 39, 59, 60). Conversely, while the distribution of the virulence factors and clades, or STs, was observed, the observation from the GWAS of *E. faecalis* pathogenicity, after adjusting for the genetic background of the isolates, implied that no individual genetic changes influence the severity of diseases at the population level. These findings are consistent with the notion that genetic traits influencing virulence are less likely to be selected than those promoting colonization as similarly seen in other pathogens (61). Altogether, these findings suggest that the distribution of the *E. faecalis* virulence factors may largely depend on the genetic background, implying that the lineage effects on pathogenicity may be more pronounced than the population-wide effects of individual genetic changes. Alternatively, there may be a predominance of certain lineages in some individuals, as seen with other opportunistic pathogens (62), whose risk factors for infection, including hospital exposure history, antibiotic treatment, and other underlying conditions, make them favorable for the selection of *E. faecalis* strains enriched in antibiotic resistance genes and other adaptive traits.

Likewise, the distribution of known *E. faecalis* virulence factors by isolation source mirrored the patterns observed for infection in individuals with varying hospitalization status due to the correlation between these phenotypes. These findings suggested that no individual genetic changes are overrepresented in blood and gut niches independent of the genetic background, which implied that while individual genetic changes may have an impact on extraintestinal infection, their effect at the population level is likely minimal. However, some genetic changes could be linked to specific lineages, making disentangling their effects from the genetic background a challenge. However, the absence of genetic changes statistically associated with the body isolation source, after adjusting for the population structure, suggests that these variants are not likely under positive selection because extraintestinal infection represents an evolutionary dead-end for *E. faecalis* (63). Therefore, even if such genetic changes exist, they may be rare and likely exhibit small effect sizes, making their detection challenging without analyzing large data sets with thousands of genomes. We speculate that the observed strong but nonstatistically significant signals in a single prophage, integrated at chromosome coordinates 1,398,051 to 1,446,151 bp in the V583 *E. faecalis* genome (34), could exemplify a potential locus with small population-wide effects on virulence. Indeed, prophages play a critical role in the pathogenicity of *E. faecalis* (64–67) and other bacterial pathogens, such as *Staphylococcus aureus* (37, 68). Therefore, further studies using even larger genomic data sets than the present study and adjusting for other important covariates, such as prior antibiotic usage and immune status, are required to fully investigate the impact of the identified *E. faecalis* prophage in modulating extraintestinal infection. Crucially, such studies should prospectively collect samples to minimize confounding effects due to cohort and temporal variability between the

number of cases and controls for a robust GWAS, which was one of the limitations of this study. Furthermore, definitive *E. faecalis* genetic signals for extraintestinal infection may be identified by comparing isolates obtained from the blood of patients with feces from individuals with confirmed negative blood cultures as controls. Inclusion of *E. faecalis* strains from community-acquired infections could also overcome the confounding effects due to factors related to hospitalization, such as *E. faecalis* from individuals with community-acquired bacteremia who are at a higher risk of developing infective endocarditis (69). Altogether, our findings demonstrate that no individual *E. faecalis* genetic changes exhibit a population-wide statistical association with extraintestinal infection, implying that all *E. faecalis* strains are capable of translocating into the bloodstream and causing severe diseases, consistent with their known opportunistic pathogenic lifestyle. Although *E. faecalis* genetic changes that are important for survival in the blood may exist, these would not be fixed in the population, especially if they had no impact on colonization, as individual strains would have to accidentally "re-discover" them repeatedly. Therefore, vaccination strategies targeting all rather than specific genetic backgrounds would lead to increased protection from severe *E. faecalis* diseases.

The estimated heritability based on unitig sequence variation of ~40% for infection in individuals with different hospitalization statuses and ~30% for body isolation source suggests that the contribution of *E. faecalis* genetics to these phenotypes is not negligible but relatively modest compared to that observed for other phenotypes, such as antimicrobial resistance (70). Our findings are consistent with findings from a recent bacterial GWAS of pathogenicity in *Streptococcus pneumoniae* (71) and Group B Streptococcus (*Streptococcus agalactiae*) (72). However, other studies have found negligible heritability for pathogenicity in *Neisseria meningitidis* (61), which suggests that the evolution of the pathogenicity trait is neutral. Previous studies have suggested that antibiotic resistance plays a major role in bloodstream invasion (14–16, 59, 60). Indeed, broad-spectrum antibiotic use disrupts the stable gut microbial community by removing typically antibiotic-susceptible competitor species, leading to the overgrowth and dissemination of *E. faecalis* into the bloodstream (59, 60). Therefore, follow-up studies of *E. faecalis* isolates sampled from feces of healthy individuals and bloodstream of patients, adjusting for other important variables such as antibiotic use, are required to determine specific genetic changes modulating pathogenicity and virulence and account for potential missing heritability. These studies will be better placed to assess the relative effect of host and gut environmental factors, such as microbiota perturbations due to antibiotic use, compared to the population-wide impact of individual genetic changes in modulating *E. faecalis* virulence and pathogenicity (73).

We acknowledge the limitations of this study, which primarily stem from the sampling biases due to the use of preexisting sequencing data sets. Firstly, there was uneven distribution of blood and fecal isolates from hospitalized and nonhospitalized individuals. Secondly, due to the retrospective nature of the study, we did not have access to detailed clinical information, including comorbidities, previous antibiotic use, and the individual's age. Adjusting for these factors would further strengthen our findings. Thirdly, our sample size is modest as it is based on a collection of *E. faecalis* isolates from only two countries in Europe. However, our data set size is similar to or larger than those described in previous studies (68, 74), which demonstrated sufficient power to detect statistically significant associations between specific individual loci and phenotypes. We recommend follow-up studies with larger sample sizes, balanced data sets by hospitalization status and body isolation source, and most importantly, including detailed clinical information, especially antibiotic use, comorbidities, and an individual's age, to adjust for potential confounding effects in the GWAS analysis.

Our exploratory findings derived from a geographically and temporally diverse whole-genome data set of *E. faecalis* isolates suggest that the pathogenicity of *E. faecalis* infections may not be primarily driven by the specific population-wide effects of individual genetic changes. These results may further illustrate the opportunistic

pathogenic lifestyle of *E. faecalis*, whereby infection of individuals with different hospitalization statuses and body isolation sources could be an accidental consequence of gut colonization dynamics as seen in other gut commensals (63). Due to the absence of specific individual genetic variants associated with body isolation source and hospitalization status, ultimately, the commensal-to-pathogen switch and virulence of *E. faecalis* may be predominantly modulated by multiple genetic variants, i.e., polygenic, genetic background or lineages, epigenetic mechanisms, host factors, and the gut milieu, including the ecological side effects of broad-spectrum antibiotics on the gastrointestinal microbiota.

MATERIALS AND METHODS

Sample characteristics and microbiological processing

For this study, we selected a total of 736 human *E. faecalis* isolates from a collection of whole-genome sequences from isolates collected from several European countries described by Pöntinen et al. (43). We included isolates from countries where both fecal and blood specimens were collected, namely, the Netherlands (n = 300) and Spain (n = 436). The isolates represent collections from the University Medical Center Utrecht, Utrecht, The Netherlands (n = 300); the European Network for Antibiotic Resistance and Epidemiology at the University Medical Center Utrecht, Utrecht, The Netherlands (n = 6); the Hospital Ramòn y Cajal, Madrid, Spain (n = 375); and Spain (n = 55). By isolation source, 296 isolates were sampled from feces, while 440 were from blood. Of these, 485 were collected from hospitalized patients, while 251 were from nonhospitalized individuals. The isolates were collected over a 21-year period (1996 to 2016); therefore, our data set was both geographically and temporally diverse. We did not use clinical metadata related to the patients, and all isolate identifiers were de-identified; therefore, additional institutional review board approval was not required.

Genome sequencing, molecular typing, assembly, and annotation

Short-read sequencing was done at the Wellcome Sanger Institute using the Illumina HiSeq X paired-end sequencing platform. As part of our quality control procedures, we used Kraken (version 0.10.66) (75) to check for potential species contamination. We assembled sequence reads that passed quality control using Velvet *de novo* assembler (version 1.2.10) (76) and annotated the resultant draft assemblies using Prokka (version 1.14.6) (77). To generate multiple sequence alignments for the whole genome sequences, we mapped the reads against the V583 *E. faecalis* reference genome (34) using the Snippy (version 4.6.0) haploid variant calling and core genome pipeline (https://github.com/tseemann/snippy). We performed *in silico* genome-based typing of the isolates using MLST, using ST or clone definitions in the MLST database (https://pubmlst.org/efaecalis) (48, 78), implemented in SRST2 (79).

Phylogenetic reconstruction and population structure analysis

To generate a phylogeny of the *E. faecalis* isolates, we first identified genomic positions containing SNPs using SNP-sites (version 2.3.2) (80). Next, we used the SNPs to construct a maximum-likelihood phylogenetic tree using IQ-TREE (version 2.1.2) (81). We selected the general time reversible and Gamma substitution models. We processed and rooted the generated phylogeny at the midpoint of the longest branch using the APE package (version 4.3) (82) and phytools (version 0.7.70) (83). We annotated and visualized the rooted phylogeny using the "gridplot" and "phylo4d" functions implemented in phylosignal (version 1.3) (84) and phylobase (version 0.8.6) packages (https://cran.r-project.org/package=phylobase), respectively. We used PopPUNK (version 1.2.2) to define the population structure of the isolates (47).

Antibiotic resistance and virulence gene profiles

We identified genotypic antibiotic resistance for seven major antibiotic classes, namely, glycopeptides (vancomycin), aminoglycosides, macrolides, tetracyclines, phenicols, and oxazolidinones, as described by Pöntinen et al. (43). We screened the sequencing reads for the presence and absence of antibiotic resistance genes using ARIBA (version 2.14.4) (85) and the ResFinder 3.2 database (86). We included additional genes conferring resistance to vancomycin, namely, vanA [European Nucleotide Archive (ENA): accession: AAA65956.1], vanB (ENA accession: AAO82021.1), vanC (ENA accession number: AAA24786.1), vanD (ENA accession: AAD42184.1), vanE (ENA accession: AAL27442.1), and vanG (ENA accession: NG_048369.1), and linezolid, namely, cfrD (ENA accession: PHLC01000011). We compared the abundance of antibiotic resistance genes per isolate using a generalized linear regression model with a Poisson log link function with pathogenicity or hospitalization status and country of origin as covariates, the latter to adjust for geographical differences. We used the test of equal proportions to compare the relative abundance of genotypic antibiotic resistance for each antibiotic class among hospitalized and nonhospitalized individuals, as well as blood and feces.

We also assessed the presence and absence of *E. faecalis* virulence genes obtained from the VFDB (49). These included genes encoding proteins involved in adherence to the epithelial surfaces (*ace*, *ebpA*, *ebpB*, *ebpC*, *ecbA*, EF0149, EF0485, *efaA*, and *srtC*), exoenzymes (EF0818, EF3023, *gelE*, and *sprE*), biofilm formations (*bopD*, *fsrA*, *fsrB*, and *fsrC*), immune modulation or antiphagocytosis (*cpsA-K*), and exotoxins (*cylL-l*, *cylL-s*, *cylM*, and *cylR2*) between isolates from hospitalized and nonhospitalized individuals and those associated with intestinal colonization and extraintestinal infection. We used BLASTN (version 2.9.0+) (87) to determine the presence and absence of the virulence genes. To avoid incorrectly missing genes potentially split between multiple contigs during *de novo* genome assembly, we considered all the highest scoring pairs with a minimum length of 100 bp using BioPython (88). We used the test of equal proportions to compare the relative abundance of genotypic antibiotic resistance for each antibiotic class among hospitalized and nonhospitalized individuals, as well as blood and feces.

Genome-wide association study

To generate the input SNP data for the GWAS, we used VCFtools (version 0.1.16) (89) to convert bi-allelic SNPs into the pedigree file accepted by PLINK software (90). We filtered out genomic positions with SNPs with a minor allele frequency of <5% or missing variant calls in >10% of the isolates using PLINK (version 1.90b4) (90). Next, we identified unitig sequences, variable-length *k*-mer sequences generated from nonbranching paths in a compacted De Bruijn graph. First, we build a De Bruijn graph using assemblies of all the isolates based on 31 bp *k*-mer sequences using Bifrost (version 1.0.1) (91). We then queried the generated De Bruijn graph using the query option in Bifrost to generate the presence and absence patterns of each identified unitig in the assemblies of each isolate. We then combined the presence and absence patterns of all the isolates into a single file and then merged them with the phenotype data (isolation source or hospitalization status) to generate PLINK-formatted pedigree files, which were used for the downstream GWAS analysis. We used the same threshold for variant frequency to filter out rare unitigs before the GWAS.

We undertook GWAS analyses using SNPs and unitigs to identify genetic variants associated with pathogenicity (hospitalization) and extraintestinal infection of *E. faecalis*. We used FaST-LMM (FastLmmC, version 2.07.20140723) (50), which uses a linear mixed model for the GWAS. For both methods, we specified a kinship matrix based on the unitig presence and absence data to adjust for the clonal population structure of the isolates, which is a major confounder in bacterial GWAS analyses (35). Since the GWAS tools used in this study were originally developed to mostly handle human diploid DNA data, we coded the variants as human mitochondrial DNA (which is haploid) by specifying the chromosome number as 26 (92, 93). To control the false discovery rate,

we used the Bonferroni correction method to adjust the statistical significance (*P*-values) inferred by each GWAS method based on the likelihood ratio test. We specified the genome length of the *E. faecalis* V583 reference genome (3,218,031 bp) as the maximum possible number of genomic variants possible, assuming that variants can independently occur at each genomic position. Since this assumption may not necessarily be true, our approach is likely to be more conservative than the Bonferroni correction based on the number of tested variants; therefore, it may minimize false positives but may slightly increase false negatives. The advantage of our approach is that by using the same number of possible variants based on the genome length, a consistent *P*-value threshold can be used to adjust different types of genetic variation, i.e., SNPs, accessory genes, *k*-mers, and unitigs, to simplify interpretation and comparison of statistical significance across different studies.

We visualized the GWAS results using Manhattan plots generated using standard plotting functions in R (version 4.0.3) (https://www.R-project.org/). Specific genomic features associated with each SNP and unitig were analyzed further by comparing the genomic sequences to the V583 *E. faecalis* reference genome (34) using BLASTN (version 2.5.0+) (94) and BioPython (version 1.78) (88). To identify potential issues arising due to the population structure, we generated Q-Q plots to compare the observed and expected statistical significance using qqman (version 0.1.7) (95). We calculated the overall proportion of the variance of the phenotype explained by *E. faecalis* genetics, i.e., narrow-sense heritability, using GCTA (version 1.93.2) (44).

Statistical analysis

We compared the number of *E. faecalis* antibiotic resistance genes per isolate among hospitalized and nonhospitalized patients and blood and fecal isolates using a Poisson generalized linear regression model with a log link. We used the rest for equality of proportions to assess whether a single dominant lineage is present at a frequency of >50%. We compared the frequency of STs and lineages in isolates from hospitalized and nonhospitalized patients and blood and fecal isolates using the chi-squared test.

ACKNOWLEDGMENTS

The authors would like to thank the study participants and guardians, the clinical and laboratory staff who collected and processed the samples at various laboratories in the Netherlands, and the sequencing, core, and pathogen teams at the Wellcome Sanger Institute for their support.

A.K.P., S.A.-A., and J.C. were funded by the Trond Mohn Foundation (grant number: TMS2019TMT04); R.J.L.W. and T.M.C. by the Joint Programming Initiative in Antimicrobial Resistance (grant number: JPIAMR2016-AC16/00039); A.R.F. by the FCT/MCTES Individual Call to Scientific Employment Stimulus (grant number: CEECIND/02268/2017); A.R.F., C.N., and L.P. by the Applied Molecular Biosciences Unit-UCIBIO that is financed by national funds from FCT (grant numbers: UIDP/04378/2020 and UIDB/04378/2020); J.C. also by ERC (grant number: 742158); and A.K.P. also by Marie Skłodowska-Curie Actions (grant number: 801133). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript, and the findings do not necessarily reflect the official views and policies of the author's institutions and funders. For the purposes of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

C.C., A.K.P., and J.C. conceived and designed the study. C.C., A.K.P., and J.C. performed the data curation. C.C. performed the formal data analysis. J.C. acquired the funding. S.D.B. and J.C. provided the resources for the study. C.C. and A.K.P. analyzed the data. C.C., A.K.P., and J.C. wrote the first draft. All authors edited and revised the manuscript.

The authors declare no competing financial or non-financial interests.

AUTHOR AFFILIATIONS

[1]Department of Epidemiology of Microbial Diseases, Yale School of Public Health, Yale University, New Haven, Connecticut, USA
[2]Parasites and Microbes Programme, Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, United Kingdom
[3]Department of Biostatistics, Faculty of Medicine, University of Oslo, Oslo, Norway
[4]Norwegian National Advisory Unit on Detection of Antimicrobial Resistance, Department of Microbiology and Infection Control, University Hospital of North Norway, Tromsø, Norway
[5]Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, the Netherlands
[6]UCIBIO-Applied Molecular Biosciences Unit, Laboratory of Microbiology, Department of Biological Sciences, REQUIMTE Faculty of Pharmacy, University of Porto, Porto, Portugal
[7]Associate Laboratory i4HB, Institute for Health and Bioeconomy, Faculty of Pharmacy, University of Porto, Porto, Portugal
[8]TOXRUN, Toxicology Research Unit, University Institute of Health Sciences, CESPU, CRL, Gandra, Portugal
[9]Department of Food and Agriculture, Area of Biochemistry and Molecular Biology, University of La Rioja, Logroño, Spain
[10]Department of Microbiology, Ramón y Cajal University Hospital, Ramón y Cajal Institute for Health Research (IRYCIS), Madrid, Spain
[11]CIBER in Infectious Diseases (CIBERINFEC), Madrid, Spain
[12]Department of Mathematics and Statistics, Helsinki Institute of Information Technology, University of Helsinki, Helsinki, Finland

AUTHOR ORCIDs

Chrispin Chaguza http://orcid.org/0000-0002-2108-1757
Anna K. Pöntinen http://orcid.org/0000-0002-3160-2042
Janetta Top http://orcid.org/0000-0002-4620-8128
Carla Novais http://orcid.org/0000-0002-3826-6731
Carmen Torres http://orcid.org/0000-0003-3709-1690
Jukka Corander http://orcid.org/0000-0002-7752-1942

FUNDING

Funder	Grant(s)	Author(s)
Trond Mohn stiftelse (Trond Mohn Foundation)	TMS2019TMT04	Anna K. Pöntinen Sergio Arredondo-Alonso Jukka Corander
JPIAMR	JPIAMR2016-AC16/00039	Teresa M. Coque Rob J. L. Willems
Applied Molecular Biosciences Unit-UCIBIO	CEECIND/02268/2017	Ana R. Freitas
FCT	UIDP/04378/2020 and UIDB/04378/2020	Ana R. Freitas Carla Novais Luisa Peixe
ERC	742158	Jukka Corander
EC | H2020 | PRIORITY "Excellent science" | H2020 Marie Skłodowska-Curie Actions (MSCA)	801133	Anna K. Pöntinen

DATA AVAILABILITY

The whole-genome sequencing data used in this study are publicly available in the European Nucleotide Archive (ENA) under the accession numbers provided in Data Set S1 in the supplemental material.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Data Set S1 (Spectrum00201-23-S0001.xlsx). Characteristics of the whole-genome sequenced *E. faecalis* isolates used in the present study.
Supplemental material (Spectrum00201-23-S0002.pdf). Tables S1 to S4 and Fig. S1.

REFERENCES

1. Kao PHN, Kline KA. 2019. Dr. Jekyll and Mr. Hide: how *Enterococcus faecalis* subverts the host immune response to cause infection. J Mol Biol 431:2932–2945. https://doi.org/10.1016/j.jmb.2019.05.030
2. Lebreton F, Manson AL, Saavedra JT, Straub TJ, Earl AM, Gilmore MS. 2017. Tracing the enterococci from paleozoic origins to the hospital. Cell 169:849–861. https://doi.org/10.1016/j.cell.2017.04.027
3. Arias CA, Murray BE. 2012. The rise of the enterococcus: beyond vancomycin resistance. Nat Rev Microbiol 10:266–278. https://doi.org/10.1038/nrmicro2761
4. Murray BE. 1990. The life and times of the enterococcus. Clin Microbiol Rev 3:46–65. https://doi.org/10.1128/CMR.3.1.46
5. Fischetti VA, Novick RP, Ferretti JJ, Portnoy DA, Braunstein M, Rood JI. 2019. Pathogenicity of enterococci. In Gram-positive pathogens. Washington, DC, USA. https://doi.org/10.1128/9781683670131
6. Pinholt M, Ostergaard C, Arpi M, Bruun NE, Schønheyder HC, Gradel KO, Søgaard M, Knudsen JD, Danish Collaborative Bacteraemia Network (DACOBAN). 2014. Incidence, clinical characteristics and 30-day mortality of enterococcal bacteraemia in Denmark 2006-2009: a population-based cohort study. Clin Microbiol Infect 20:145–151. https://doi.org/10.1111/1469-0691.12236
7. Sievert DM, Ricks P, Edwards JR, Schneider A, Patel J, Srinivasan A, Kallen A, Limbago B, Fridkin S, National Healthcare Safety Network (NHSN) Team and Participating NHSN Facilities. 2013. Antimicrobial-resistant pathogens associated with healthcare-associated infections: summary of data reported to the national healthcare safety network at the centers for disease control and prevention, 2009-2010. Infect Control Hosp Epidemiol 34:1–14. https://doi.org/10.1086/668770
8. Lebreton F, van Schaik W, McGuire AM, Godfrey P, Griggs A, Mazumdar V, Corander J, Cheng L, Saif S, Young S, Zeng Q, Wortman J, Birren B, Willems RJL, Earl AM, Gilmore MS. 2013. Emergence of epidemic multidrug-resistant *Enterococcus faecium* from animal and commensal strains. mBio 4:e00534-13. https://doi.org/10.1128/mBio.00534-13
9. Gilmore MS, Lebreton F, van Schaik W. 2013. Genomic transition of enterococci from gut commensals to leading causes of multidrug-resistant hospital infection in the antibiotic era. Curr Opin Microbiol 16:10–16. https://doi.org/10.1016/j.mib.2013.01.006
10. Berg RD. 1995. Bacterial translocation from the gastrointestinal tract. Trends Microbiol 3:149–154. https://doi.org/10.1016/s0966--842x(00)88906-4
11. Palmer KL, va Schaik W, Willems RJL, Gilmore MS. 2014. Enterococcal genomics. In Gilmore MS, DB Clewell, Y Ike, N Shankar (ed), Enterococci: from commensals to leading causes of drug resistant infection. Massachusetts Eye and Ear Infirmary, Boston.
12. Van Tyne D, Manson AL, Huycke MM, Karanicolas J, Earl AM, Gilmore MS. 2019. Impact of antibiotic treatment and host innate immune pressure on enterococcal adaptation in the human bloodstream. Sci Transl Med 11:eaat8418. https://doi.org/10.1126/scitranslmed.aat8418
13. Rani A, Ranjan R, McGee HS, Andropolis KE, Panchal DV, Hajjiri Z, Brennan DC, Finn PW, Perkins DL. 2017. Urinary microbiome of kidney transplant patients reveals dysbiosis with potential for antibiotic resistance. Transl Res 181:59–70. https://doi.org/10.1016/j.trsl.2016.08.008
14. Archambaud C, Derré-Bobillot A, Lapaque N, Rigottier-Gois L, Serror P. 2019. Intestinal translocation of enterococci requires a threshold level of enterococcal overgrowth in the lumen. Sci Rep 9:8926. https://doi.org/10.1038/s41598-019-45441-3
15. Krueger WA, Krueger-Rameck S, Koch S, Carey V, Pier GB, Huebner J. 2004. Assessment of the role of antibiotics and enterococcal virulence factors in a mouse model of extraintestinal translocation. Crit Care Med 32:467–471. https://doi.org/10.1097/01.CCM.0000109447.04893.48
16. Reyman M, van Houten MA, Watson RL, Chu M, Arp K, de Waal WJ, Schiering I, Plötz FB, Willems RJL, van Schaik W, Sanders EAM, Bogaert D. 2022. Effects of early-life antibiotics on the developing infant gut microbiome and resistome: a randomized trial. Nat Commun 13:893. https://doi.org/10.1038/s41467-022-28525-z
17. McFarland LV, Surawicz CM, Stamm WE. 1990. Risk factors for clostridium difficile carriage and C. difficile-associated diarrhea in a cohort of hospitalized patients. J Infect Dis 162:678–684. https://doi.org/10.1093/infdis/162.3.678
18. Brown KA, Khanafer N, Daneman N, Fisman DN. 2013. Meta-analysis of antibiotics and the risk of community-associated clostridium difficile infection. Antimicrob Agents Chemother 57:2326–2332. https://doi.org/10.1128/AAC.02176-12
19. Jett BD, Atkuri RV, Gilmore MS. 1998. *Enterococcus faecalis* localization in experimental endophthalmitis: role of plasmid-encoded aggregation substance. Infect Immun 66:843–848. https://doi.org/10.1128/IAI.66.2.843-848.1998
20. Jett BD, Jensen HG, Nordquist RE, Gilmore MS. 1992. Contribution of the pAD1-encoded cytolysin to the severity of experimental *Enterococcus faecalis* endophthalmitis. Infect Immun 60:2445–2452. https://doi.org/10.1128/iai.60.6.2445-2452.1992
21. Huycke MM, Spiegel CA, Gilmore MS. 1991. Bacteremia caused by hemolytic, high-level gentamicin-resistant *Enterococcus faecalis*. Antimicrob Agents Chemother 35:1626–1634. https://doi.org/10.1128/AAC.35.8.1626
22. Kawalec M, Pietras Z, Daniłowicz E, Jakubczak A, Gniadkowski M, Hryniewicz W, Willems RJL. 2007. Clonal structure of *Enterococcus faecalis* isolated from Polish hospitals: characterization of epidemic clones. J Clin Microbiol 45:147–153. https://doi.org/10.1128/JCM.01704-06
23. Raven KE, Reuter S, Gouliouris T, Reynolds R, Russell JE, Brown NM, Török ME, Parkhill J, Peacock SJ. 2016. Genome-based characterization of hospital-adapted *Enterococcus faecalis* lineages. Nat Microbiol 1:15033. https://doi.org/10.1038/nmicrobiol.2015.33
24. Thurlow LR, Thomas VC, Narayanan S, Olson S, Fleming SD, Hancock LE. 2010. Gelatinase contributes to the pathogenesis of endocarditis caused by *Enterococcus faecalis*. Infect Immun 78:4936–4943. https://doi.org/10.1128/IAI.01118-09
25. Zeng J, Teng F, Murray BE. 2005. Gelatinase is important for translocation of *Enterococcus faecalis* across polarized human enterocyte-like T84 cells.

Infect Immun 73:1606–1612. https://doi.org/10.1128/IAI.73.3.1606-1612.2005
26. Chow JW, Thal LA, Perri MB, Vazquez JA, Donabedian SM, Clewell DB, Zervos MJ. 1993. Plasmid-associated hemolysin and aggregation substance production contribute to virulence in experimental enterococcal endocarditis. Antimicrob Agents Chemother 37:2474–2477. https://doi.org/10.1128/AAC.37.11.2474
27. Tendolkar PM, Baghdayan AS, Shankar N. 2005. The N-terminal domain of enterococcal surface protein, esp, is sufficient for esp-mediated biofilm enhancement in *Enterococcus faecalis*. J Bacteriol 187:6213–6222. https://doi.org/10.1128/JB.187.17.6213-6222.2005
28. Pultz NJ, Shankar N, Baghdayan AS, Donskey CJ. 2005. Enterococcal surface protein esp does not facilitate intestinal colonization or translocation of *Enterococcus faecalis* in clindamycin-treated mice. FEMS Microbiol Lett 242:217–219. https://doi.org/10.1016/j.femsle.2004.11.006
29. Heikens E, Leendertse M, Wijnands LM, van Luit-Asbroek M, Bonten MJM, van der Poll T, Willems RJL. 2009. Enterococcal surface protein esp is not essential for cell adhesion and intestinal colonization of *Enterococcus faecium* in mice. BMC Microbiol. 9:19. https://doi.org/10.1186/1471-2180-9-19
30. Shankar N, Baghdayan AS, Gilmore MS. 2002. Modulation of virulence within a pathogenicity island in vancomycin-resistant *Enterococcus faecalis*. Nature 417:746–750. https://doi.org/10.1038/nature00802
31. McBride SM, Coburn PS, Baghdayan AS, Willems RJL, Grande MJ, Shankar N, Gilmore MS. 2009. Genetic variation and evolution of the pathogenicity island of *Enterococcus faecalis*. J Bacteriol 191:3392–3402. https://doi.org/10.1128/JB.00031-09
32. Olmsted SB, Dunny GM, Erlandsen SL, Wells CL. 1994. A plasmid-encoded surface protein on *Enterococcus faecalis* augments its Internalization by cultured intestinal epithelial cells. J Infect Dis 170:1549–1556. https://doi.org/10.1093/infdis/170.6.1549
33. McBride SM, Fischetti VA, Leblanc DJ, Moellering RC, Gilmore MS. 2007. Genetic diversity among *Enterococcus faecalis*. PLoS One 2:e582. https://doi.org/10.1371/journal.pone.0000582
34. Paulsen IT, Banerjei L, Myers GSA, Nelson KE, Seshadri R, Read TD, Fouts DE, Eisen JA, Gill SR, Heidelberg JF, Tettelin H, Dodson RJ, Umayam L, Brinkac L, Beanan M, Daugherty S, DeBoy RT, Durkin S, Kolonay J, Madupu R, Nelson W, Vamathevan J, Tran B, Upton J, Hansen T, Shetty J, Khouri H, Utterback T, Radune D, Ketchum KA, Dougherty BA, Fraser CM. 2003. Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. Science 299:2071–2074. https://doi.org/10.1126/science.1080613
35. Power RA, Parkhill J, de Oliveira T. 2017. Microbial genome-wide association studies: lessons from human GWAS. Nat Rev Genet 18:41–50. https://doi.org/10.1038/nrg.2016.132
36. Mortimer TD, Zhang JJ, Ma KC, Grad YH. 2022. Loci for prediction of penicillin and tetracycline susceptibility in neisseria gonorrhoeae: a genome-wide association study. Lancet Microbe 3:e376–e381. https://doi.org/10.1016/S2666-5247(22)00034-9
37. Richardson EJ, Bacigalupe R, Harrison EM, Weinert LA, Lycett S, Vrieling M, Robb K, Hoskisson PA, Holden MTG, Feil EJ, Paterson GK, Tong SYC, Shittu A, van Wamel W, Aanensen DM, Parkhill J, Peacock SJ, Corander J, Holmes M, Fitzgerald JR. 2018. Gene exchange drives the ecological success of a multi-host bacterial pathogen. Nat Ecol Evol 2:1468–1478. https://doi.org/10.1038/s41559-018-0617-0
38. Wee BA, Alves J, Lindsay DSJ, Klatt A-B, Sargison FA, Cameron RL, Pickering A, Gorzynski J, Corander J, Marttinen P, Opitz B, Smith AJ, Fitzgerald JR. 2021. Population analysis of *Legionella pneumophila* reveals a basis for resistance to complement-mediated killing. Nat Commun 12:7165. https://doi.org/10.1038/s41467-021-27478-z
39. Ruiz-Garbajosa P, Cantón R, Pintado V, Coque TM, Willems R, Baquero F, del Campo R. 2006. Genetic and Phenotypic differences among *Enterococcus faecalis* clones from intestinal colonisation and invasive disease. Clin Microbiol Infect 12:1193–1198. https://doi.org/10.1111/j.1469-0691.2006.01533.x
40. Kim EB, Marco ML. 2014. Nonclinical and clinical *Enterococcus faecium* strains, but not *Enterococcus faecalis* strains, have distinct structural and functional genomic features. Appl Environ Microbiol 80:154–165. https://doi.org/10.1128/AEM.03108-13
41. He Q, Hou Q, Wang Y, Li J, Li W, Kwok L-Y, Sun Z, Zhang H, Zhong Z. 2018. Comparative genomic analysis of *Enterococcus faecalis*: insights into their environmental adaptations. BMC Genomics 19:527. https://doi.org/10.1186/s12864-018-4887-3
42. Boumasmoud M, Dengler Haunreiter V, Schweizer TA, Meyer L, Chakrakodi B, Schreiber PW, Seidl K, Kühnert D, Kouyos RD, Zinkernagel AS. 2022. Genomic surveillance of vancomycin-resistant *Enterococcus faecium* reveals spread of a linear plasmid conferring a nutrient utilization advantage. mBio 13:e0377121. https://doi.org/10.1128/mbio.03771-21
43. Pöntinen AK, Top J, Arredondo-Alonso S, Tonkin-Hill G, Freitas AR, Novais C, Gladstone RA, Pesonen M, Meneses R, Pesonen H, Lees JA, Jamrozy D, Bentley SD, Lanza VF, Torres C, Peixe L, Coque TM, Parkhill J, Schürch AC, Willems RJL, Corander J. 2021. Apparent nosocomial adaptation of *Enterococcus faecalis* predates the modern hospital era. Nat Commun 12:1523. https://doi.org/10.1038/s41467-021-21749-5
44. Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88:76–82. https://doi.org/10.1016/j.ajhg.2010.11.011
45. Guzman Prieto AM, van Schaik W, Rogers MRC, Coque TM, Baquero F, Corander J, Willems RJL. 2016. Global emergence and dissemination of enterococci as nosocomial pathogens: attack of the clones Front Microbiol 7:788. https://doi.org/10.3389/fmicb.2016.00788
46. Tedim AP, Ruiz-Garbajosa P, Corander J, Rodríguez CM, Cantón R, Willems RJ, Baquero F, Coque TM. 2015. Population biology of intestinal enterococcus isolates from hospitalized and nonhospitalized individuals in different age groups. Appl Environ Microbiol 81:1820–1831. https://doi.org/10.1128/AEM.03661-14
47. Lees JA, Harris SR, Tonkin-Hill G, Gladstone RA, Lo SW, Weiser JN, Corander J, Bentley SD, Croucher NJ. 2019. Fast and flexible bacterial genomic epidemiology with poppunk. Genome Res 29:304–316. https://doi.org/10.1101/gr.241455.118
48. Ruiz-Garbajosa P, Bonten MJM, Robinson DA, Top J, Nallapareddy SR, Torres C, Coque TM, Cantón R, Baquero F, Murray BE, del Campo R, Willems RJL. 2006. Multilocus sequence typing scheme for *Enterococcus faecalis* reveals hospital-adapted genetic complexes in a background of high rates of recombination . J Clin Microbiol 44:2220–2228. https://doi.org/10.1128/JCM.02596-05
49. Chen L. 2004. VFDB: A reference database for bacterial virulence factors. Nucleic Acids Research 33:D325–D328. https://doi.org/10.1093/nar/gki008
50. Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. 2011. Fast linear mixed models for genome-wide association studies. Nat Methods 8:833–835. https://doi.org/10.1038/nmeth.1681
51. Manson JM, Hancock LE, Gilmore MS. 2010. Mechanism of chromosomal transfer of *Enterococcus faecalis* pathogenicity island, capsule, antimicrobial resistance, and other traits. Proc Natl Acad Sci U S A 107:12269–12274. https://doi.org/10.1073/pnas.1000139107
52. Miller WR, Munita JM, Arias CA. 2014. Mechanisms of antibiotic resistance in enterococci. Expert Rev Anti Infect Ther 12:1221–1236. https://doi.org/10.1586/14787210.2014.956092
53. da Silva RAG, Tay WH, Ho FK, Tanoto FR, Chong KKL, Choo PY, Ludwig A, Kline KA. 2022. *Enterococcus faecalis* alters endo-lysosomal trafficking to replicate and persist within mammalian cells. PLoS Pathog 18:e1010434. https://doi.org/10.1371/journal.ppat.1010434
54. Waters CM, Wells CL, Dunny GM. 2003. The aggregation domain of aggregation substance, not the RGD motifs, is critical for efficient internalization by HT-29 enterocytes. Infect Immun 71:5682–5689. https://doi.org/10.1128/IAI.71.10.5682-5689.2003
55. Gentry-Weeks CR, Karkhoff-Schweizer R, Pikis A, Estay M, Keith JM. 1999. Survival of *Enterococcus faecalis* in mouse peritoneal macrophages. Infect Immun 67:2160–2165. https://doi.org/10.1128/IAI.67.5.2160-2165.1999
56. Nunez N, Derré-Bobillot A, Trainel N, Lakisic G, Lecomte A, Mercier-Nomé F, Cassard A-M, Bierne H, Serror P, Archambaud C. 2022. The unforeseen intracellular lifestyle of in hepatocytes. Gut Microbes 14:2058851.
57. Shankar N, Coburn P, Pillar C, Haas W, Gilmore M. 2004. Enterococcal cytolysin: activities and association with other virulence traits in a pathogenicity island. Int J Med Microbiol 293:609–618. https://doi.org/10.1078/1438-4221-00301

58. Thurlow LR, Thomas VC, Fleming SD, Hancock LE. 2009. *Enterococcus faecalis* capsular polysaccharide serotypes C and D and their contributions to host innate immune evasion. Infect Immun 77:5551–5557. https://doi.org/10.1128/IAI.00576-09
59. Taur Y, Xavier JB, Lipuma L, Ubeda C, Goldberg J, Gobourne A, Lee YJ, Dubin KA, Socci ND, Viale A, Perales M-A, Jenq RR, van den Brink MRM, Pamer EG. 2012. Intestinal domination and the risk of bacteremia in patients undergoing allogeneic hematopoietic stem cell transplantation. Clin Infect Dis 55:905–914. https://doi.org/10.1093/cid/cis580
60. Ubeda C, Taur Y, Jenq RR, Equinda MJ, Son T, Samstein M, Viale A, Socci ND, van den Brink MRM, Kamboj M, Pamer EG. 2010. Vancomycin-resistant enterococcus domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans. J Clin Invest 120:4332–4341. https://doi.org/10.1172/JCI43918
61. Kremer PHC, Lees JA, Ferwerda B, van de Ende A, Brouwer MC, Bentley SD, van de Beek D. 2020. Genetic variation in neisseria meningitidis does not influence disease severity in meningococcal meningitis. Front Med (Lausanne) 7:594769. https://doi.org/10.3389/fmed.2020.594769
62. Rodríguez I, Figueiredo AS, Sousa M, Aracil-Gisbert S, Fernández-de-Bobadilla MD, Lanza VF, Rodríguez C, Zamora J, Loza E, Mingo P, Brooks CJ, Cantón R, Baquero F, Coque TM. 2021. A 21-year survey of *Escherichia coli* from bloodstream infections (BSI) in a tertiary hospital reveals how community-hospital dynamics of B2 phylogroup clones influence local BSI rates. mSphere 6:e0086821. https://doi.org/10.1128/msphere.00868-21
63. Le Gall T, Clermont O, Gouriou S, Picard B, Nassif X, Denamur E, Tenaillon O. 2007. Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic group *Escherichia coli* strains. Mol Biol Evol 24:2373–2384. https://doi.org/10.1093/molbev/msm172
64. Royer G, Roisin L, Demontant V, Lo S, Coutte L, Lim P, Pawlotsky JM, Jacquier H, Lepeule R, Rodriguez C, Woerther PL. 2021. Microdiversity of *Enterococcus faecalis* isolates in cases of infective endocarditis: selection of non-synonymous mutations and large deletions is associated with phenotypic modifications. Emerg Microbes Infect 10:929–938. https://doi.org/10.1080/22221751.2021.1924865
65. Askora A, El-Telbany M, El-Didamony G, Ariny E, Askoura M. 2020. Characterization of Φef-Vb1 prophage infecting oral *Enterococcus faecalis* and enhancing bacterial biofilm formation. J Med Microbiol 69:1151–1168. https://doi.org/10.1099/jmm.0.001246
66. La Rosa SL, Snipen L-G, Murray BE, Willems RJL, Gilmore MS, Diep DB, Nes IF, Brede DA. 2015. A genomic virulence reference map of *Enterococcus faecalis* reveals an important contribution of phage03-like elements in nosocomial genetic lineages to pathogenicity in a caenorhabditis elegans infection model. Infect Immun 83:2156–2167. https://doi.org/10.1128/IAI.02801-14
67. Matos RC, Lapaque N, Rigottier-Gois L, Debarbieux L, Meylheuc T, Gonzalez-Zorn B, Repoila F, Lopes M de F, Serror P, Hughes D. 2013. *Enterococcus faecalis* prophage dynamics and contributions to pathogenic traits. PLoS Genet 9:e1003539. https://doi.org/10.1371/journal.pgen.1003539
68. Young BC, Earle SG, Soeng S, Sar P, Kumar V, Hor S, Sar V, Bousfield R, Sanderson ND, Barker L, Stoesser N, Emary KRW, Parry CM, Nickerson EK, Turner P, Bowden R, Crook DW, Wyllie DH, Day NPJ, Wilson DJ, Moore CE. 2019. Panton-valentine leucocidin is the key determinant of *Staphylococcus aureus* pyomyositis in a bacterial GWAS. Elife 8:e42486. https://doi.org/10.7554/eLife.42486
69. Dahl A, Lauridsen TK, Arpi M, Sørensen LL, Østergaard C, Sogaard P, Bruun NE. 2016. Risk factors of endocarditis in patients with *Enterococcus faecalis* bacteremia: external validation of the NOVA score. Clin Infect Dis 63:771–775. https://doi.org/10.1093/cid/ciw383
70. Farhat MR, Freschi L, Calderon R, Ioerger T, Snyder M, Meehan CJ, de Jong B, Rigouts L, Sloutsky A, Kaur D, Sunyaev S, van Soolingen D, Shendure J, Sacchettini J, Murray M. 2019. GWAS for quantitative resistance phenotypes in mycobacterium tuberculosis reveals resistance genes and regulatory regions. Nat Commun 10:2128. https://doi.org/10.1038/s41467-019-10110-6
71. Lees JA, Ferwerda B, Kremer PHC, Wheeler NE, Serón MV, Croucher NJ, Gladstone RA, Bootsma HJ, Rots NY, Wijmega-Monsuur AJ, Sanders EAM, Trzciński K, Wyllie AL, Zwinderman AH, van den Berg LH, van Rheenen W, Veldink JH, Harboe ZB, Lundbo LF, de Groot LCPGM, van Schoor NM, van der Velde N, Ängquist LH, Sørensen TIA, Nohr EA, Mentzer AJ, Mills TC, Knight JC, du Plessis M, Nzenze S, Weiser JN, Parkhill J, Madhi S, Benfield T, von Gottberg A, van der Ende A, Brouwer MC, Barrett JC, Bentley SD, van de Beek D. 2019. Joint sequencing of human and pathogen genomes reveals the genetics of pneumococcal meningitis. Nat Commun 10:2176. https://doi.org/10.1038/s41467-019-09976-3
72. Chaguza Chrispin, Jamrozy D, Bijlsma MW, Kuijpers TW, van de Beek D, van der Ende A, Bentley SD. 2022. Population genomics of group B streptococcus reveals the genetics of neonatal disease onset and meningeal invasion. Nat Commun 13:4215. https://doi.org/10.1038/s41467-022-31858-4
73. Lefort A, Panhard X, Clermont O, Woerther P-L, Branger C, Mentré F, Fantin B, Wolff M, Denamur E. 2011. Host factors and portal of entry outweigh bacterial determinants to predict the severity of *Escherichia coli* bacteremia. J Clin Microbiol 49:777–783. https://doi.org/10.1128/JCM.01902-10
74. Chaguza C, Smith JT, Bruce SA, Gibson R, Martin IW, Andam CP. 2022. Prophage-encoded immune evasion factors are critical for host infection, switching, and adaptation. Cell Genomics 2:100194. https://doi.org/10.1016/j.xgen.2022.100194
75. Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15:R46. https://doi.org/10.1186/gb-2014-15-3-r46
76. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de bruijn graphs. Genome Res 18:821–829. https://doi.org/10.1101/gr.074492.107
77. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153
78. Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics: BIGSdb software, the pubMLST org website and their applications. Wellcome Open Res 3:124. https://doi.org/10.12688/wellcomeopenres.14826.1
79. Inouye M, Dashnow H, Raven L-A, Schultz MB, Pope BJ, Tomita T, Zobel J, Holt KE. 2014. SRST2: rapid genomic surveillance for public health and hospital microbiology labs. Genome Med 6:90. https://doi.org/10.1186/s13073-014-0090-6
80. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, Harris SR. 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb Genom 2:e000056. https://doi.org/10.1099/mgen.0.000056
81. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R, Teeling E. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol BiolEvol 37:1530–1534. https://doi.org/10.1093/molbev/msaa015
82. Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290. https://doi.org/10.1093/bioinformatics/btg412
83. Revell LJ. 2012. Phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol 3:217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x
84. Keck F, Rimet F, Bouchez A, Franc A. 2016. Phylosignal: an R package to measure, test, and explore the phylogenetic signal. Ecol Evol 6:2774–2780. https://doi.org/10.1002/ece3.2051
85. Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, Keane JA, Harris SR. 2017. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. Microb Genom 3:e000131. https://doi.org/10.1099/mgen.0.000131
86. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 67:2640–2644. https://doi.org/10.1093/jac/dks261
87. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
88. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL. 2009. Biopython: freely available python tools for computational molecular biology and bioinformatics. Bioinformatics 25:1422–1423. https://doi.org/10.1093/bioinformatics/btp163
89. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group. 2011. The variant call format and

VCFtools. Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330
90. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575. https://doi.org/10.1086/519795
91. Holley G, Melsted P. 2020. Bifrost: highly parallel construction and indexing of colored and compacted de bruijn graphs. Genome Biol 21:249. https://doi.org/10.1186/s13059-020-02135-8
92. Chewapreecha C, Marttinen P, Croucher NJ, Salter SJ, Harris SR, Mather AE, Hanage WP, Goldblatt D, Nosten FH, Turner C, Turner P, Bentley SD, Parkhill J, Guttman DS. 2014. Comprehensive identification of single nucleotide polymorphisms associated with beta-lactam resistance within pneumococcal mosaic genes. PLoS Genet 10:e1004547. https://doi.org/10.1371/journal.pgen.1004547
93. Li Y, Metcalf BJ, Chochua S, Li Z, Walker H, Tran T, Hawkins PA, Gierke R, Pilishvili T, McGee L, Beall BW. 2019. Genome-wide association analyses of invasive pneumococcal isolates identify a missense bacterial mutation associated with meningitis. Nat Commun 10:178. https://doi.org/10.1038/s41467-018-07997-y
94. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389
95. Turner SD. Qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. Bioinformatics. https://doi.org/10.1101/005165