RESEARCH ARTICLE

# Restricted Gene Flow among Hospital Subpopulations of *Enterococcus faecium*

**Rob J. L. Willems,[a] Janetta Top,[a] Willem van Schaik,[a] Helen Leavis,[a] Marc Bonten,[a] Jukka Sirén,[b] William P. Hanage,[c] and Jukka Corander[b]**

Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, The Netherlands[a]; Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland[b]; and Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA[c]

**ABSTRACT** ***Enterococcus faecium*** **has recently emerged as an important multiresistant nosocomial pathogen. Defining population structure in this species is required to provide insight into the existence, distribution, and dynamics of specific multiresistant or pathogenic lineages in particular environments, like the hospital. Here, we probe the population structure of *E. faecium* using Bayesian-based population genetic modeling implemented in Bayesian Analysis of Population Structure (BAPS) software. The analysis involved 1,720 isolates belonging to 519 sequence types (STs) (491 for *E. faecium* and 28 for *Enterococcus faecalis*). *E. faecium* isolates grouped into 13 BAPS (sub)groups, but the large majority (80%) of nosocomial isolates clustered in two subgroups (2-1 and 3-3). Phylogenetic and eBURST analysis of BAPS groups 2 and 3 confirmed the existence of three separate hospital lineages (17, 18, and 78), highlighting different evolutionary trajectories for BAPS 2-1 (lineage 78) and 3-3 (lineage 17 and lineage 18) isolates. Phylogenomic analysis of 29 *E. faecium* isolates showed agreement between BAPS assignment of STs and their relative positions in the phylogenetic tree. Odds ratio calculation confirmed the significant association between hospital isolates with BAPS 3-3 and lineages 17, 18, and 78. Admixture analysis showed a scarce number of recombination events between the different BAPS groups. For the *E. faecium* hospital population, we propose an evolutionary model in which strains with a high propensity to colonize and infect hospitalized patients arise through horizontal gene transfer. Once adapted to the distinct hospital niche, this subpopulation becomes isolated, and recombination with other populations declines.**

**IMPORTANCE** **Multiresistant *Enterococcus faecium* has become one of the most important nosocomial pathogens, causing increasing numbers of nosocomial infections worldwide. Here, we used Bayesian population genetic analysis to identify groups of related *E. faecium* strains and show a significant association of hospital and farm animal isolates to different genetic groups. We also found that hospital isolates could be divided into three lineages originating from sequence types (STs) 17, 18, and 78. We propose that, driven by the selective pressure in hospitals, the three hospital lineages have arisen through horizontal gene transfer, but once adapted to the distinct pathogenic niche, this population has become isolated and recombination with other populations declines. Elucidation of the population structure is a prerequisite for effective control of multiresistant *E. faecium* since it provides insight into the processes that have led to the progressive change of *E. faecium* from an innocent commensal to a multiresistant hospital-adapted pathogen.**

**Received** 20 May 2012 **Accepted** 11 June 2012 **Published** 17 July 2012

**Citation** Willems RJL, et al. 2012. Restricted gene flow among hospital subpopulations of *Enterococcus faecium*. mBio 3(4):e00151-12. doi:10.1128/mBio.00151-12.

**Editor** Paul Keim, Northern Arizona University

**Copyright** © 2012 Willems et al. This is an open-access article distributed under the terms of the Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported License, which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to Rob J. L. Willems, r.willems@umcutrecht.nl.

W.P.H. and J.C. contributed equally to this article.

In the past two decades, *Enterococcus faecium* has become recognized as an important nosocomial pathogen. Up to the 1980s, the large majority of enterococcal hospital-associated infections (HAI) were caused by *Enterococcus faecalis*, but since the beginning of the 1990s, the proportion of HAI caused by *E. faecium* has increased and has now almost reached parity with that of *E. faecalis* (1). One proposed reason for this changing epidemiology is that, in comparison with *E. faecalis*, *E. faecium* shows relatively high rates of resistance against important antibiotics, including ampicillin and vancomycin. In addition, studies of the population biology of *E. faecium* using multilocus sequence typing (MLST) data have revealed the existence of a distinct genetic subpopulation associated with nosocomial epidemics. This subpopulation has been designated lineage C1 (2) and was later renamed clonal complex 17 (CC17) on the basis of eBURST (3) analysis of MLST data (4, 5).

CC17 has been recognized as a successful hospital-associated *E. faecium* (HA *E. faecium*) clonal complex, exhibiting high-level vancomycin, ampicillin, and quinolone resistance, although in most European countries CC17 remained primarily vancomycin susceptible (4, 6, 7, 8, 9, 10, 11, 12, 13). However, in addition to this distinct resistance profile, genome-wide analyses have shown that HA *E. faecium* strains have a genetic repertoire distinct from that of *E. faecium* strains that asymptomatically colonize the gastrointestinal (GI) tract of humans and animals in the community (14, 15). This distinct genetic repertoire includes cell surface pro-

teins, of which the enterococcal surface protein, Esp, is a known virulence determinant (8, 10, 16, 17, 18, 19, 20, 21, 22, 23); genomic islands tentatively encoding novel metabolic pathways (24); and insertion sequence elements (14). It is now considered that these determinants may be adaptive elements that have improved the relative fitness of this HA *E. faecium* subpopulation in the hospital environment (5, 9, 25, 26).

Despite the clear importance of CC17 as the main genetic subpopulation, including hospital isolates, its precise phylogenetic status remains uncertain. Analyses of the population structure of *E. faecium* indicate that this species undergoes a high rate of homologous recombination (4, 27). High recombination rates can lead to error in phylogenetic analyses. This is especially true if only a small portion of the genome is interrogated. In the case of the eBURST approach used to define CC17, a consequence of large amounts of recombination is the spurious grouping of diverse and distinct lineages into a single clonal complex. It has previously been suggested (28) that phylogenetic reconstructions of *E. faecium* are vulnerable to such errors. Analyses using alternative methods to eBURST suggest this may be the case, with the constituent lineages of CC17 (sequence types [STs] 17, 18, 78, and 80) representing distinct genetic lineages of which the relationships between cannot be confidently assigned (5, 26, 27). Together with the observation that whole genome sequences of two CC17 isolates (E1162 [ST17] and U0317 [ST78]) differ substantially in gene content (15), this strongly indicates that HA *E. faecium* may not have evolved from a single founder (i.e., ST17) and that, consequently, CC17 as presently identified may not exist but is an artifact of the assumptions embedded within the eBURST algorithm.

An alternative approach to eBURST is an analysis of genetic population structure, with the power to combine the identification of deep branching lineages and recombination between them. This can be done using Bayesian Analysis of Population Structure (BAPS) software (29, 30, 31). Unlike approaches such as eBURST (3), BAPS does not attempt to retrieve phylogenetic information or implement a phylogenetic model of clustering but rather uses a statistical genetic model to partition molecular variation based on both clonal ancestry and recombination patterns as identified from DNA sequence data. This approach has recently been used to probe genetic population structure in *Streptococcus pneumoniae* (32), *Escherichia coli* (33), *Campylobacter coli*, *Campylobacter jejuni* (34, 35), and *Neisseria* spp. (36) and has been shown to be able to detect structure even in highly recombinogenic populations. Here, we used BAPS to identify groups of related *E. faecium* strains and show a significant association of hospital and farm animal isolates to different BAPS groups. We suggest that hospital-associated lineages contained in different BAPS groups have, however, acquired similar adaptive elements.

## RESULTS

**BAPS-based genetic structure in the *E. faecium* population.** The data set used for our analysis consisted of 519 distinct STs, of which 491 were found among 1,720 *E. faecium* isolates and the remaining 28 STs in 29 *E. faecalis* isolates. The eBURST analysis based on the 491 *E. faecium* STs yielded 19 eBURST groups and 101 singletons (STs not part of an eBURST group or clonal complex) (Fig. 1). The largest eBURST group included 329 (67%) of the STs and 1,459 (85%) of the isolates, indicating that eBURST might be unreliable in estimating relatedness in this large straggly *E. faecium* group.

To gain an alternative perspective on *E. faecium* population structure, we used a Bayesian population genetic model to identify clusters characterized by different allele frequencies based on multilocus DNA sequences (31). The 519 STs were partitioned into 7 groups, of which groups 1 to 6 included the *E. faecium* isolates, while BAPS group 7 was entirely composed of *E. faecalis* isolates (Table 1). BAPS groups 2 and 3 represented the majority of the sample, containing 227 (44%) and 190 (40%) of the STs and 829 (47%) and 784 (45%) of the isolates, respectively (Table 1). It is known that in Bayesian model-based analysis, fine distinctions between closely related clusters can be obscured by the presence of relatively distant clusters in the data, which produce a large amount of signal and consequently render smaller signals of difference insignificant (see, e.g., the discussion in the work of Marttinen et al. [37]). As a result, BAPS groups 2 and 3 were each individually analyzed using BAPS. This nested analysis strategy was implemented to provide greater resolution of population structure in these groups. The results showed four subgroups of BAPS group 2 (BAPS 2-1 to 2-4) and five subgroups of BAPS group 3 (BAPS 3-1 to 3-5) (Table 1). The nested analysis thus finally subdivided the *E. faecium* population into 13 distinct subpopulations and one *E. faecalis* BAPS group 7. Given the smaller sizes of the subgroups emerging in the second stage of clustering compared to the initial analysis, it is not feasible to split these subgroups even further by successively repeating the clustering procedure, as there will not be enough variation in the data to reliably infer the underlying groups.

**Distribution of *E. faecium* isolates among BAPS (sub)groups.** Two BAPS groups (2-1, 3-3) contained the majority (80%) of isolates from hospitalized patients (Table 1). However, BAPS 2-1 also contained a large number of animal-related isolates, in contrast with BAPS 3-3.

To assess significance of associations between BAPS groups and isolate sources, odds ratios (ORs) were calculated. This revealed a significant association of hospital and farm animal isolates to different BAPS groups. Strikingly, BAPS group 3-3 was positively associated with isolates retrieved from hospitalized patients, whereas hospital isolates were negatively associated with BAPS groups 1, 2-1, 2-3, 3-2, and 3-4. In turn, isolates from farm animals were significantly associated with BAPS groups 2-1, 2-3, 2-4, 3-2, and 5 and negatively associated with BAPS group 3-3 (see Table 2). To examine whether sampling bias could influence results by inflating the numbers of individual STs, ORs were calculated based on a single example of each ST. The results confirmed positive association of BAPS 3-3 with hospitalized patients and BAPS 2-1 and 2-4 with farm animals and negative association of BAPS 2-1 and 3-2 with hospitalized patients and BAPS 3-3 with farm animals (see Table S2 in the supplemental material).

STs 17, 18, and 78 are important nodes in the previously described CC17, a globally distributed clonal complex identified by eBURST analysis (Fig. 1). To correct for erroneous linkage introduced by eBURST in organisms like *E. faecium*, which have a high recombination-to-mutation ratio (28), we have divided CC17 into lineages arising from each of the subfounder STs (shown in Fig. 1), namely, STs 17, 18, and 78. Two of these (STs 17 and 18, together with descendant STs) are included in BAPS 3-3, which has the strongest and most significant association with hospital-associated isolates among all these groups. eBURST analysis of STs

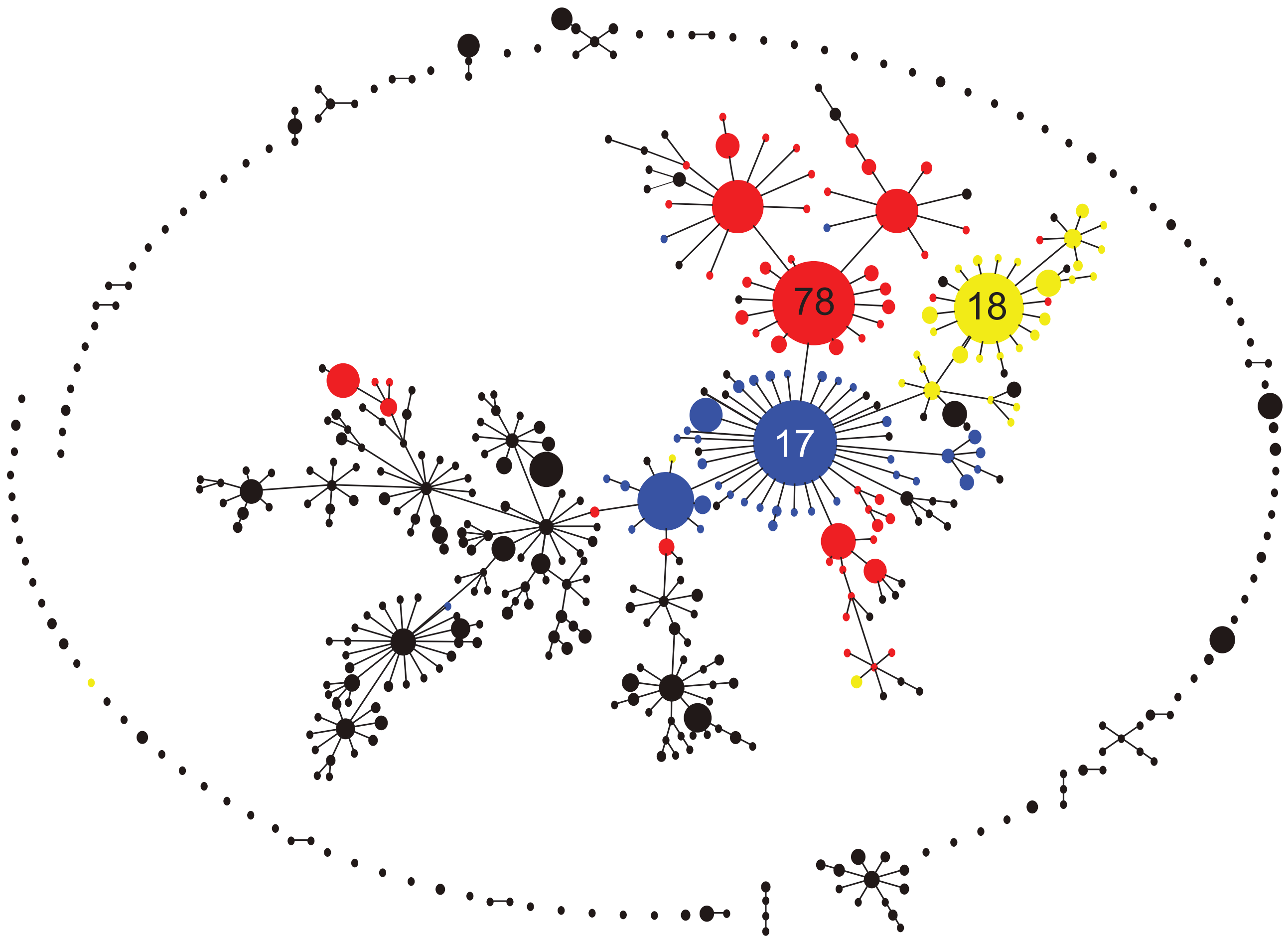


**FIG 1** eBURST-based population snapshot of *E. faecium* based on 519 STs representing 1,749 isolates contained in the online *E. faecium* MLST database (http://efaecium.mlst.net). STs belonging to lineages 17, 18, and 78 are color coded in blue, red, and yellow, respectively.

belonging to the two largest BAPS groups, BAPS 2-1 and BAPS 3-3, revealed coclustering of STs representing the majority of hospital-associated *E. faecium* isolates (881; 74%) into three major lineages: lineage 78 (364 isolates; 42 STs) in BAPS 2-1 and lineage 17 (328 isolates; 35 STs) and lineage 18 (189 isolates; 26 ST) in BAPS 3-3 (see Fig. S1 and S2 in the supplemental material). ORs

**TABLE 1** Distribution of STs and isolates among BAPS subpopulations

| BAPS group | BAPS subgroup | No. of STs | No. of *E. faecium* isolates from: Hospital patients | Nonhospital persons | Pets | Pigs | Poultry | Calves | Other[a] | Unknown | No. of *E. faecalis* isolates | Total no. of isolates |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | 53 | 36 | 12 | 0 | 10 | 0 | 0 | 11 | 14 | 0 | 83 |
| 2 | 1 | 148 | 403 | 18 | 71 | 34 | 97 | 0 | 20 | 30 | 0 | 673 |
| | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | 3 | 63 | 74 | 9 | 7 | 7 | 3 | 17 | 6 | 5 | 0 | 128 |
| | 4 | 15 | 15 | 0 | 1 | 1 | 3 | 0 | 5 | 2 | 0 | 27 |
| 3 | 1 | 44 | 53 | 6 | 7 | 8 | 1 | 0 | 8 | 3 | 0 | 86 |
| | 2 | 17 | 13 | 9 | 4 | 14 | 1 | 0 | 1 | 6 | 0 | 48 |
| | 3 | 104 | 547 | 10 | 8 | 6 | 12 | 3 | 8 | 10 | 0 | 604 |
| | 4 | 4 | 2 | 4 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 7 |
| | 5 | 21 | 21 | 1 | 9 | 0 | 4 | 0 | 2 | 2 | 0 | 39 |
| 4 | | 9 | 6 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 9 |
| 5 | | 9 | 5 | 0 | 0 | 2 | 2 | 0 | 0 | 1 | 0 | 10 |
| 6 | | 3 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 5 |
| 7 | | 28 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 29 | 29 |
| Total | | 519 | 1,180 | 70 | 107 | 83 | 123 | 20 | 62 | 75 | 29 | 1,749 |

[a] Includes animal food products ($n = 29$), other animal isolates, or isolates from nonspecified animals ($n = 8$) and environmental isolates ($n = 25$).

**TABLE 2** Associations between *E. faecium* isolates from hospitalized patients and farm animals with BAPS groups and lineages

| Source | BAPS group | Lineage | No. of isolates from source | Total no. of isolates[a] | ORs[c] | 95% CI |
|---|---|---|---|---|---|---|
| Hospitalized patients | 1 | | 36 | 69 | 0.41 | 0.254–0.669 |
| | 2-1 | allSTs | 403 | 643 | 0.49 | 0.391–0.605 |
| | 2-1 | 78 | 364 | 453 | 1.89 | 1.45–2.449 |
| | 2-2 | | 1 | 1 | ND | |
| | 2-3 | | 74 | 123 | 0.57 | 0.389–0.829 |
| | 2-4 | | 15 | 25 | 0.59 | 0.261–1.314 |
| | 3-1 | | 53 | 83 | 0.68 | 0.43–1.082 |
| | 3-2 | | 13 | 42 | 0.17 | 0.086– 0.325 |
| | 3-3 | allSTs | 547 | 594 | 7.69 | 5.567–10.61 |
| | 3-3 | 17 | 329 | 342 | 13.44 | 7.633–23.67 |
| | 3-3 | 18 | 190 | 196 | 14.69 | 6.465–33.341 |
| | 3-4 | | 2 | 7 | 0.16 | 0.03–0.808 |
| | 3-5 | | 21 | 37 | 0.51 | 0.263–0.983 |
| | 4 | | 6 | 8 | 1.18 | 0.238–5.883 |
| | 5 | | 5 | 9 | 0.49 | 0.131–1.834 |
| | 6 | | 4 | 4 | ND | |
| Farm animals[b] | 1 | | 14 | 69 | 1.54 | 0.743–2.477 |
| | 2-1 | allSTs | 144 | 643 | 2.14 | 1.64–2.795 |
| | 2-1 | 78 | 4 | 453 | 0.03 | 0.012–0.087 |
| | 2-2 | | 0 | 1 | ND | |
| | 2-3 | | 30 | 123 | 1.79 | 1.156–2.756 |
| | 2-4 | | 9 | 25 | 3.03 | 1.322–6.921 |
| | 3-1 | | 13 | 83 | 0.98 | 0.531–1.789 |
| | 3-2 | | 16 | 42 | 3.38 | 1.786–6.39 |
| | 3-3 | allSTs | 26 | 594 | 0.16 | 0.103–0.239 |
| | 3-3 | 17 | 2 | 342 | 0.02 | 0.006– 0.095 |
| | 3-3 | 18 | 2 | 196 | 0.05 | 0.012–0.19 |
| | 3-4 | | 1 | 7 | 0.88 | 0.105–7.301 |
| | 3-5 | | 6 | 37 | 1.02 | 0.42–2.463 |
| | 4 | | 0 | 8 | ND | |
| | 5 | | 4 | 9 | 4.25 | 1.135–15.945 |
| | 6 | | 0 | 4 | ND | |

[a] *E. faecium* isolates in the source categories of hospitalized patients, farm animals, and other sources. Isolates from unknown sources were not included. In total, 1,645 isolates were included in the analysis.

[b] Pigs, poultry, veal calves, meat, and milk products.

[c] ORs indicate significance of association between *E. faecium* source categories and BAPS (sub)group and lineage. ND, not done.

confirmed a significant association of lineages 17, 18, and 78 with hospital isolates and a significant negative association with isolates from farm animals and meat products (Table 2; see also Table S2 in the supplemental material).

**Comparison of BAPS-based grouping and whole-genome phylogenomics.** BAPS grouping is based on the concatenated sequences of the seven MLST housekeeping genes, which consist of 3,463 bp of DNA sequence. To examine whether the groupings we find are concordant with those based on a larger fraction of the genome, we constructed a phylogenetic tree based on 299 orthologous proteins representing 64,555 amino acids contained in the 29 *E. faecium* strains for which a genome sequence is available. The phylogenomics analysis showed agreement between BAPS assignment of STs and their relative position in the phylogenetic tree (Fig. 2). Isolates belonging to BAPS 1 clustered far from BAPS 2 and 3 isolates, and distinct groups of isolates corresponding to the BAPS 2-1 and 3-3 subgroups were observed. This suggests that the groupings identified by BAPS analysis of MLST loci reflect a deep phylogenetic structure that is also apparent in larger samples of loci and hints toward functional differences between the subgroups. The phylogenomic tree suggests one monophyletic group consisting of isolates belonging to the three hospital-adapted lineages 17, 18, and 78. However, despite the increased amounts of data available for the construction of the phylogenomic tree in Fig. 2, it must be treated with caution, because sample size and the sampling frame for Fig. 2 is markedly different from that available for MLST data. A neighbor-joining tree based on concatenated alignments of MLST gene sequences contained in the two largest BAPS (sub)groups, BAPS 2 and BAPS 3, clearly indicate phylogenetic diversity between BAPS 2 (lineage 78) and BAPS 3 (lineages 17 and 18) (see Fig. S3 in the supplemental material).

**Evolution of the hospital-associated *E. faecium* population.** The fact that animals, specifically poultry and pet animals, are significantly associated with BAPS 2-1 and hospital-associated isolates with BAPS 3-3 suggests a distinction between these two habitats. However, approximately a third (31%) of isolates from hospitalized patients, including the major hospital lineage 78, cocluster with isolates from animal sources in BAPS 2-1 (Table 1). We interpret this as evidence that this subset of hospital-associated isolates has emerged separately from hospital isolates in BAPS 3-3 and has a distinct evolutionary history.

Further examination of the constituent lineages of CC17 (i.e., lineages 17 and 18 in BAPS groups 3-3 and lineage 78 in 2-1) suggests that other important clinical features, such as ampicillin resistance and the presence of the *esp* virulence factor, which is implicated in biofilm formation, urinary tract infections, and endocarditis (17, 22, 23), are common. The *esp* gene is recorded as present in 73% and 72% of hospital isolates in lineage 17 ($n = 248$)

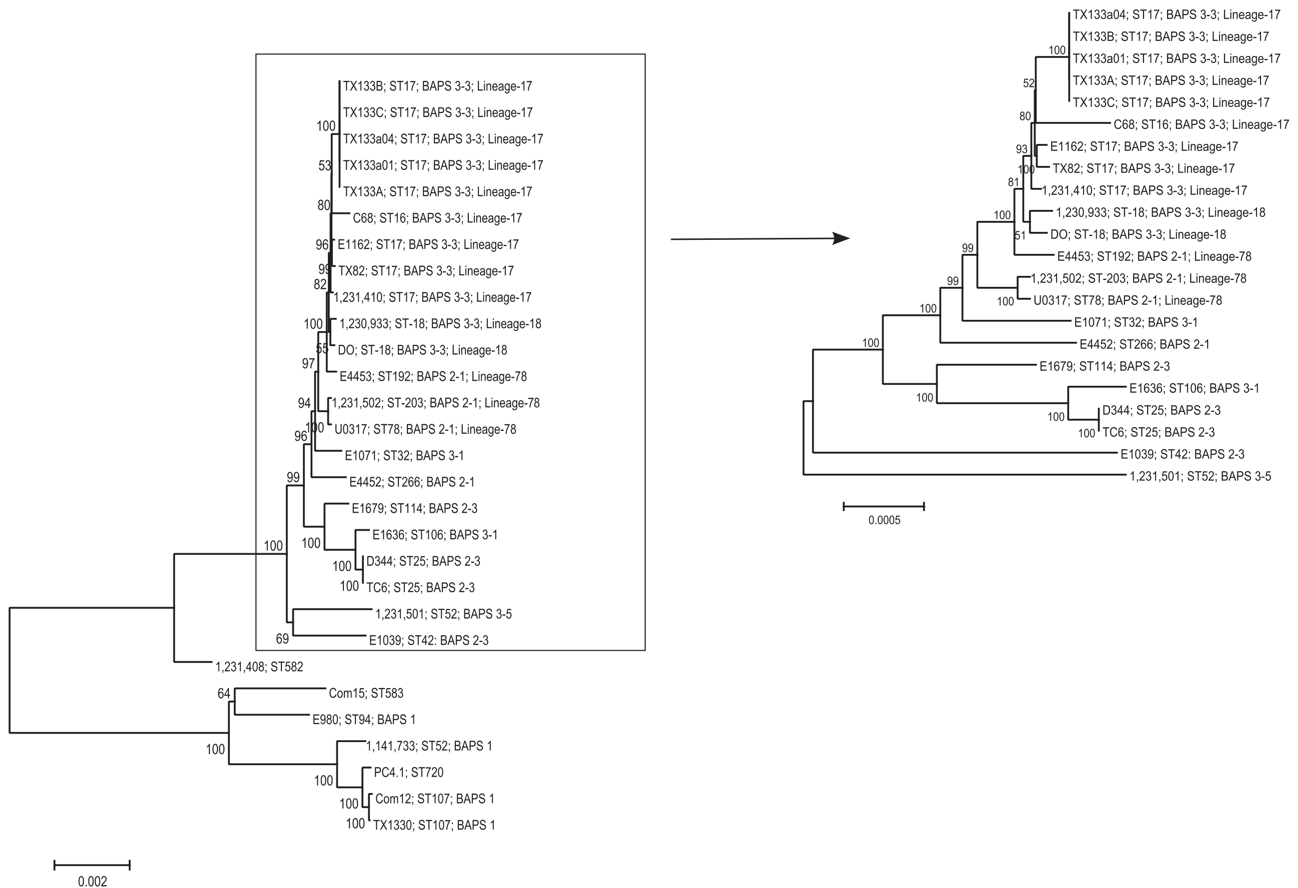


**FIG 2** Minimum evolution tree based on the concatenated alignments of 299 orthologous proteins conserved in draft genome sequences of 29 *E. faecium* isolates. Bootstrap values are indicated and are based on 1,000 permutations. Strain codes are indicated as well as ST, BAPS group, and lineage based on their ST. Accession numbers: *E. faecium* 1231408, GenBank NZ_ACBB00000000; *E. faecium* 1231501, GenBank NZ_ACAY00000000; *E. faecium* TX0133A, GenBank NZ_AECH00000000; *E. faecium* TX0082, GenBank NZ_AEBU00000000; *E. faecium* C68, GenBank NZ_ACJQ00000000; *E. faecium* 1231410, GenBank NZ_ACBA00000000; *E. faecium* 1230933, GenBank NZ_ACAS00000000; *E. faecium* E4453, GenBank AEDZ00000000; *E. faecium* 1231502, GenBank NZ_ACAX00000000; *E. faecium* E4452, GenBank AEOU00000000; *E. faecium* D344, GenBank NZ_ACZZ00000000; *E. faecium* TC6, GenBank NZ_ACOB00000000; *E. faecium* Com15, GenBank NZ_ACBD00000000; *E. faecium* 1141733, GenBank NZ_ACAZ00000000; *E. faecium* PC4.1, GenBank NZ_ADMM00000000; *E. faecium* Com12, GenBank NZ_ACBC00000000; *E. faecium* TX1330, GenBank NZ_ACHL00000000; *E. faecium* E980, GenBank ABQA00000000; *E. faecium* E1039, GenBank ACOS00000000; *E. faecium* E1071, GenBank ABQI00000000; *E. faecium* E1162, GenBank ABQJ00000000; *E. faecium* E1636, GenBank ABRY00000000; *E. faecium* E1679, GenBank ABSC00000000; *E. faecium* U0317, GenBank ABSW00000000; *E. faecium* DO, GenBank NZ_ACIY00000000.1; *E. faecium* TX133a01, GenBank NZ_AECJ00000000.1; *E. faecium* TX133a04, GenBank NZ_AEBC00000000.1; *E. faecium* TX133B, GenBank NZ_AECI00000000.1; *E. faecium* TX133C, GenBank NZ_AEBG00000000.1. Strains 1231408, PC4.1, and Com15 lack a BAPS (sub)group designation because STs extracted from the genome sequences of these strains were not assigned yet at the time the BAPS analysis was performed. To improve resolution of the upper part of the tree, the top 22 non-BAPS 1 strains were separately clustered using the minimum evolution method.

and 78 ($n = 363$) isolates, respectively. In contrast, only 31% of lineage 18 isolates ($n = 154$) are recorded as having this gene present, which compares with 18% in the remainder of the population for which *esp* presence or absence was recorded ($n = 404$ isolates). The other feature that is present at increased frequency in all three hospital-associated lineages is ampicillin resistance, which is practically ubiquitous among lineage 17, 18, and 78 hospital isolates (more than 98% of isolates belonging to one of these three lineages are resistant) but, in contrast, is less frequent among the rest of the data set (35%). In a previous publication, using a mixed whole-genome array, we identified additional genes that were enriched among hospital isolates (14). Five genes that ranked highest as genes predictive for hospital isolates in a character evolution analysis in this previous study were also enriched among lineage 17, 18, and 78 hospital isolates (63 to 100%), while being relatively rare among the remainder of the *E. faecium* population (10 to 30%) (see Table S3 in the supplemental material). Also, SgrA, a nidogen-binding LPXTG surface adhesion implicated in biofilm formation, and EcbA, a collagen binding MSCRAMM (21), are significantly enriched among isolates belonging to the three hospital lineages as well as a genomic island tentatively encoding a metabolic pathway involved in carbohydrate transport and metabolism (24) (see Table S3). We propose that the enrichment of *esp* in two of the hospital-adapted lineages, which include BAPS groups 3-3 and 2-1, is a result of separate acquisition leading to selection in hospitals and that the same applies to ampicillin resistance and the seven additional genes and the genomic island.

**Admixture analysis and gene flow networks in *E. faecium* and between *E. faecium* and *E. faecalis*.** The BAPS program was used to identify cases of admixture in which genotypes (STs) contain sequences characteristic of more than one subpopulation. For BAPS groups 2 and 3, the vast majority of genotypes have se-

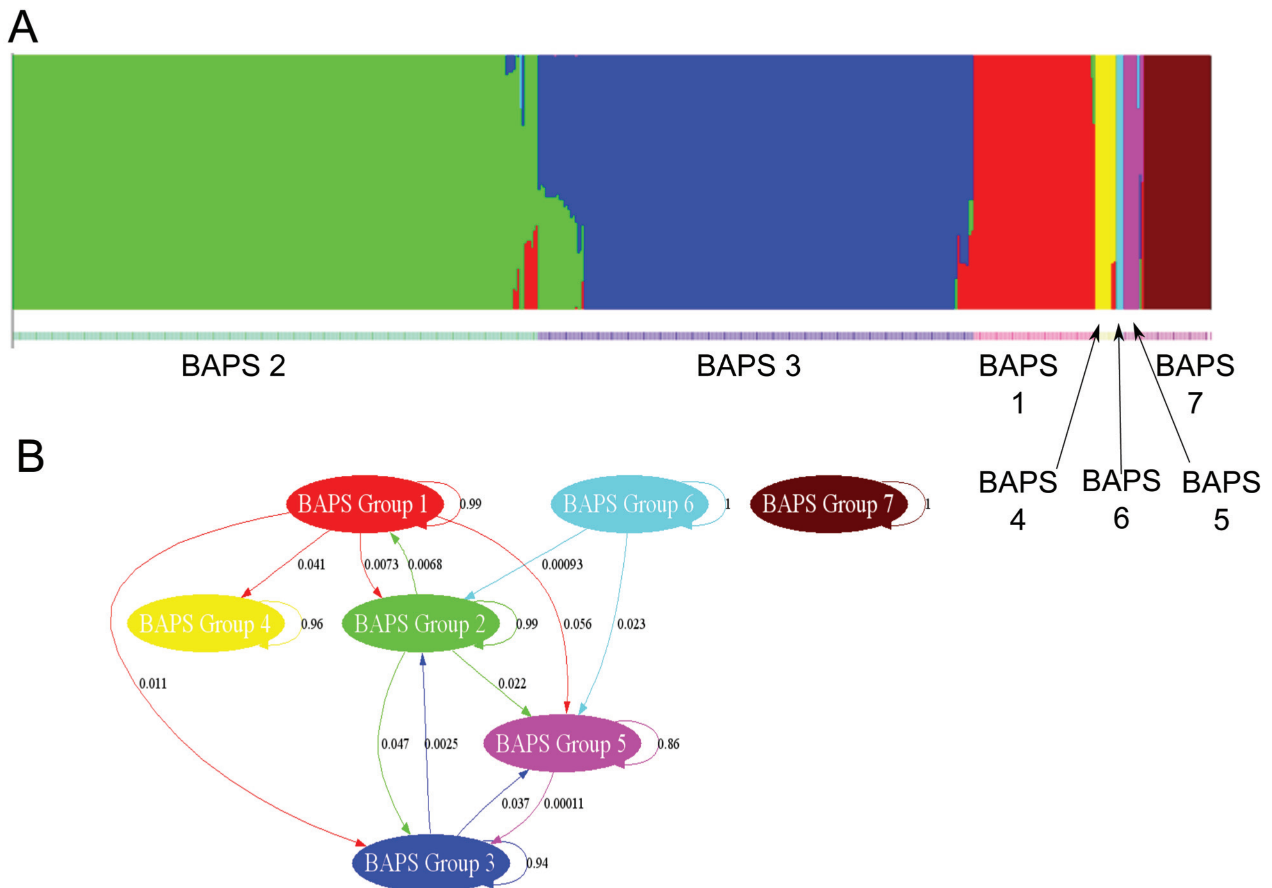


**FIG 3** Admixture analysis in the *Enterococcus* population. (Top) Admixture analysis of 519 distinct enterococcal genotypes. Each column represents a single MLST and is colored according to the proportion of genetic variation assigned to each BAPS group. The final cluster assignment is shown by the color underneath. (Bottom) Gene flow network identified in the *Enterococcus* population. Arrows indicate the average fraction of sequence variation obtained from the source cluster by clones assigned to the target cluster. Circular loops indicate the fraction of variation estimated as not arising from outside the BAPS group.

quence signatures of only one BAPS group, which is indicative of fairly restricted recombination between these groups (Fig. 3A and B; see also Fig. S4 in the supplemental material). Nevertheless, the gene flow diagram shows admixture between BAPS 2 and BAPS 3, with almost 5% of the sequences in BAPS 3 having characteristics of BAPS 2, while for genotypes in BAPS 2, only 0.25% of the sequences have characteristics of BAPS 3. In BAPS groups 1, 4, and 6, at least 94% of the sequences are characteristic of that group, pointing toward restricted recombination also in these BAPS groups (significance of admixture was determined for each ST using the threshold of $P$ values of $<0.05$). In contrast, BAPS 5 shows more mosaicism, with substantial gene flow from BAPS 1, 2, 3, and 6.

Although levels of admixture seem to be relatively low in BAPS groups 2 and 3, it is apparent from Fig. 3 that some mosaic genotypes do exist. More STs in BAPS 3 (15%) showed admixture than in BAPS 2 (6%). Within BAPS 3, admixture was most explicit in subgroups 3-2 and 3-4, which were entirely composed of mosaics (see Fig. S4 in the supplemental material). In BAPS 3-2, the most common donor is BAPS 2 (49% of STs), while in BAPS 3 to 4, 67% of STs contain sequences characteristic of BAPS 2, and the remaining 33% of BAPS 1. As such, BAPS 3-2 and 3-4 represent clear examples of populations identified by the BAPS software from traces left by recombination events (31). This suggests that the extent of recombination is unevenly distributed among the lineages of *E. faecium*, similar to the earlier reported findings for the *Pneumococcus* and *Meningococcus* based on comparable data and analysis (32, 36). However, given the relatively small number of isolates found in these populations, the potential importance of this finding for the *E. faecium* population biology remains open for further examination. Relatively high levels of admixture in BAPS 3 were found in STs from nonhospitalized persons (30% of STs) and pigs (40% of STs), while among STs from hospital isolates in BAPS 3, most notably in BAPS 3-3, the number of significantly admixed STs was low (3% of STs) (see Table S4 in the supplemental material). There was no significant association (estimated by ORs and confidence intervals [CIs] as previously described) between ampicillin resistance or the presence of *esp* and STs with significant admixture. Detailed analysis of admixture in MLST gene sequences by characterizing molecular variation at all sites indicates that traces of admixture are present in all MLST loci, and thus these results are not the result of anomalous ancestry at a single gene (data not shown).

**Mixed cultures can produce erroneous evidence for recombination between named species.** Three BAPS groups contain divergent alleles, apparently reflecting recombination between

named species. BAPS 4, involving *gdh* and *gyd*, and BAPS 6, involving *atpA*, *gdh*, *gyd*, and *adk*, contain a mix of *E. faecalis* and *E. faecium* alleles, while BAPS 5 contains mainly individual divergent *atpA E. faecalis*-like alleles and a single ST harboring an allele typical of *Enterococcus hirae* as identified by BLAST. The small number of isolates (24; and 21 STs) contained in these BAPS groups represent isolates from hospitalized patients (15), nonhospitalized persons (1), farm animals (4), the environment (2), and unknown origin (2), from 9 countries (Brazil, China, Finland, France, Netherlands, Portugal, Spain, Taiwan, and United Kingdom).

Such mosaic STs can be generated either by genuine recombination between named species or by mixed cultures that yield sequences from both species present. Repeated culture and sequencing of eight BAPS 4 isolates with ST105, ST164, ST325, ST326, ST331, ST332, ST353, and ST357 and three BAPS 6 isolates with ST419 revealed allelic profiles that were not identical to the ones reported for these STs in the MLST database. The "new" allelic profiles for these STs did no longer include typical *E. faecalis* alleles, indicating that these STs in fact do not represent mosaic STs but most probably were the result of mixed cultures. This means that these STs most probably do not exist. BAPS 5 isolates with ST366, ST367, ST369, ST394, ST405, ST428, ST436, ST444, and ST484 as well as BAPS 6 isolates with ST335 and ST362 were not available for testing, and we propose that the evidence of recombination in these STs should be assumed to be an experimental artifact, as in the case of BAPS 4 and BAPS 6. These results indicate that the utmost care must be taken when inferring rare recombination events over large distances.

## DISCUSSION

Multiresistant *E. faecium* has become one of the most important nosocomial pathogens. Emergence of multiresistant *E. faecium* strains is a problem at multiple levels. From a clinical perspective, they are among the most resistant opportunistic nosocomial pathogens, with an increasing impact on patients receiving health care. Moreover, in terms of the emergence of resistance and their considerable capacity of genetic exchange linked to high recombination rates, *E. faecium* is the perfect hub for resistance genes facilitating horizontal gene transfer among bacterial species. A high relative recombination rate means that eBURST, a popular cluster algorithm for MLST data, cannot reliably delineate the patterns of recent evolutionary descent of *E. faecium* (28). Here, we have used BAPS software to probe the genetic structure and evolution of *E. faecium*. The BAPS-based partition revealed a nonrandom distribution of animal isolates among BAPS populations and a significant association of isolates derived from hospitalized patients with specific groups that are negatively associated with isolates from farm animals. This is consistent with previous findings that demonstrated host specificity and distinct clustering of animal and human community isolates from clinical isolates based on clustering of amplified fragment length polymorphism (AFLP) profiles (38) and on comparative genomic hybridizations using an *E. faecium* mixed whole-genome array (14). The observation is also consistent with previous analysis of MLST data at the level of individual STs, confirming that the majority of hospital outbreak and infectious isolates are genotypically distinct from the majority of human commensal and animal isolates. Consequently, antibiotic-resistant clones originating from farm animals appear not to be responsible for the emergence of antibiotic-resistant *E. faecium* in hospitalized patients (27). However, the previous finding of indistinguishable vancomycin resistance transposons in pigs, nonhospitalized persons, and hospitalized patients indicates that while animal-derived *E. faecium* clones containing antibiotic resistance genes may not be responsible for infections in hospitalized patients, the resistance genes themselves may be laterally transferred from animal isolates to human clinical isolates (16).

Three *E. faecium* STs, ST17, ST18, and ST78, and STs representing their recent evolutionary descendants are significantly enriched among clinical and outbreak-associated isolates of hospitalized patients and represented major subgroup founders of the previously designated CC17, a presumed hospital-derived subpopulation of *E. faecium* that has spread globally (4, 7, 8, 39, 40). BAPS, however, resolved CC17 into two different subgroups, BAPS 2-1 and BAPS 3-3, with ST78 and descendants belonging to BAPS 2-1, separated from ST17 and ST18, both belonging to BAPS 3-3. This is consistent with the suggestion that CC17, as a monophyletic entity containing the majority of hospital isolates, is probably an artifact of documented problems with eBURST-based clustering (28), which have led to erroneous linkage of the three main hospital lineages (lineages 17, 18, and 78) into CC17.

Two recent comparative genomic studies of *E. faecium*, including 8 to 21 isolates, for which draft whole genome sequences were available, identified a deep phylogenetic split between two *E. faecium* clades that were designated clade A and clade B (using the terminology of Palmer and coworkers [41]) or commensal (CA) and hospital (HA) clades (using the terminology of Galloway-Peña and coworkers [42]). We found, using essentially the same genome sequence data, the same ancient split with isolates belonging to BAPS 1 clustering in a separate clade (B or CA), while those belonging to all the other BAPS groups clustered in clade A (or HA). Also, the topology of the phylogenetic tree described in a recent publication by Lam and coworkers (43) that includes the first complete genome sequence of an *E. faecium* strain is highly similar to the subtree we show in Fig. 2, representing only the upper 22 non-BAPS 1 strains. Our study, however, revealed that BAPS 1 isolates do not solely represent community isolates but that 43% of BAPS 1 isolates also represent isolates recovered from hospitalized patients, indicating that clade B or the CA clade is not representative only for community isolates. Similarly, isolates belonging to the other BAPS groups do not only include hospital isolates but also represent the vast majority of farm and pet animals, which indicates that clade A or the HA clade is not exclusively representative for hospital isolates. More importantly, our BAPS analysis demonstrated that hospital isolates belong to different evolutionary clusters and thus do not share a recent common ancestor—recent in this sense meaning since the establishment of modern hospitals and the dawn of the antibiotic era. The diversification of hospital-associated clusters is something that has happened relatively recently compared to the deep split between the two clades mentioned above, because appearance of large-scale hospitals and the use of antibiotics represent very recent events in evolutionary time. Using whole genome sequence information of 21 human *E. faecium* isolates, the split between the CA clade and HA clade was calculated to have occurred between 300,000 and 3 million years ago, while strains in the HA clade were estimated to have diverged from each other ~100,000 to 300,000 years ago (42). However, with a poor sampling of strains (a relatively limited number of strains, which were all human de-

rived) and a relatively simple analysis taking into account mutation rates of *Escherichia coli* and *Bacillus anthracis*, it is nearly impossible to say anything serious about divergence times for a very recombinogenic organism like *E. faecium*. One would need to carefully purge the data and be very certain that almost all variation that is left is due to mutation. Even then it is not easy to calibrate the clock and estimate divergence times, since mutation and recombination frequency may differ between bacterial populations that reside under different selective pressure.

The finding that ST78 is part of a different BAPS group (BAPS 2-1) than ST17 and ST18 (BAPS 3-3) suggests a distinct evolutionary history for hospital lineage 78. This is supported by the neighbor-joining tree based on concatenated MLST gene sequences of BAPS 2 and BAPS 3 (sub)groups (see Fig. S3), as well as a ClonalFrame analysis of the population presented in reference 5. While STs 17 and 18 (and descendant STs) are grouped in BAPS 3-3, which is significantly associated with hospital-derived isolates, lineage 78 coclusters with the majority of animal, specifically poultry, isolates in BAPS 2-1. Based on this observation, we hypothesize that the genetic evolution of hospital clones belonging to lineage 78 possibly involved animals (poultry or pet animals) as the ancestral origin since poultry isolates constitute the largest proportion of animal isolates in BAPS 2-1 and because lineage 78 was also significantly associated with STs from pet animals (data not shown). In fact, it is not implausible to speculate that the hospital-associated lineages 17, 18, and 78 all arose by connection to animals. BAPS 3-2, which is also significantly associated with animals and evolutionarily closely related to BAPS 3-3, that includes lineage 17 and lineage 18, contains a relatively high proportion of pig isolates (29% of all isolates in this BAPS group).

The observed coclustering, based on gene content, of *E. faecium* hospital isolates belonging to lineages 17, 18, and 78 (14) indicates cumulative acquisition of adaptive elements, such as ampicillin resistance and the *esp* virulence gene, by specific genotypes multiple times during the evolution of *E. faecium*. Future whole-genome-based phylogenomics analysis will provide more insights into the evolutionary history and gene content of isolates belonging to the lineages 17, 18, and 78 and the order with which particular adaptive loci and phenotypes, such as ampicillin resistance and *esp*, were acquired. If true, this suggests that the continuous rise of nosocomial *E. faecium* infections is not the result of clonal expansion of a single successful clone or lineage that emerged in hospitals 20 years ago but of consecutive waves of different clones/lineages that have evolved and were subsequently selected in hospitals.

Despite high estimated levels of recombination in *E. faecium* (4, 27), admixture and gene flow analysis indicated limited amounts of admixture between BAPS groups. Of the three largest BAPS groups (1, 2, 3), *E. faecium* isolates in BAPS 3 show higher levels of admixture, with mosaic genotypes concentrated among isolates from nonhospitalized persons and pigs (see Table S4 in the supplemental material). This may reflect an increased ability to accept foreign DNA or greater ecological opportunity for recombination. It remains to be investigated which mechanisms are responsible for the observed higher admixture level in BAPS 3. Recently, Manson and coworkers described chromosome-chromosome transfer of resistance and virulence genes as well as MLST markers between *E. faecalis* strains (44). This indicates that plasmid-mediated mobilization of chromosomal DNA contributes to MLST diversity in *E. faecalis* (44), and it is not unlikely that similar mechanisms may exist in *E. faecium*. Hospital isolates, either contained in BAPS 2 or BAPS 3, display only low levels of admixture, which may point to genetic isolation of hospital-derived *E. faecium*.

In conclusion, BAPS analysis provided new insights into the population structure of *E. faecium*, suggesting that CC17 should be divided into constituent groups descending from STs 17, 18, and 78. This analysis, as well as previous typing data, indicates a certain level of host specificity and suggests ecological isolation for some *E. faecium* populations. For the hospital population, we propose a model of enterococcal evolution in which strains with high invasive potential arise through horizontal gene transfer, but once adapted to the distinct pathogenic niche the population becomes isolated and recombination with other populations declines. This corroborates previous observations that hospital isolates carry a number of resistance and putative virulence genes not found among community/animal isolates. Analysis of the composition of the *E. faecium* hospital population over time from literature references suggests successive waves of successful *E. faecium* STs from lineages 17 and 18 in the years 1990 to 2004 (8) to lineage 78 from 2005 (7, 39, 45, 46, 47, 48, 49). The recently successful hospital lineage 78 (BAPS 2-1) seems to have an evolutionary history which is distinct from lineages 17 and 18 (BAPS 3-3) that dominated in the 1990s and may have evolved from farm animals, most probably poultry, or pet animals. The emergence of *E. faecium* as a leading nosocomial pathogen has paralleled the emergence of these three genetically distinct hospital lineages with increased potential of hospital spread. These lineages are enriched in proven and putative virulence genes, like the *esp* gene and other genes encoding surface proteins and surface appendages like pili (20) that have enabled specific hospital-adapted clones belonging to these lineages to colonize and invade hospitalized patients. The finding that successful hospital-adapted *E. faecium* strains may evolve from different genetic backgrounds, including those that prevail in animal reservoirs, has consequences for the potential flow of genes conferring resistance or virulence through the *E. faecium* population contained in various human and nonhuman ecological niches. Improved understanding of population structure can assist effective control by defining those parts of the population most associated with particular settings, such as health care or agriculture. The finding of distinct health care and agricultural populations of *E. faecium* will also facilitate future research in disclosing genetic differences between these populations. This will improve our understanding of the pathophysiological processes that have led to adaptation of the three major hospital lineages to the hospital environment. Increased insights in genes or genetic elements implicated in hospital adaptation may lead to the identification of novel targets for antibiotics and immunotherapy to combat *E. faecium* infections.

## MATERIALS AND METHODS

**Bacterial isolates.** MLST data from 1,749 enterococcal isolates (1,720 *E. faecium* and 29 *E. faecalis* isolates), representing in total 519 different STs (491 in *E. faecium* and 28 in *E. faecalis*), were included in this study. The 491 STs in *E. faecium* as well as the concatenated sequences of the seven MLST genes are accessible through and can be downloaded from the *E. faecium* MLST website (http://efaecium.mlst.net/). The 29 *E. faecalis* isolates represented all the isolates for which whole genome sequences were available at the time of analysis (January 2011). Alleles and STs of the *E. faecalis* isolates were assigned *in silico* using the whole genome sequence

data following the *E. faecium* MLST scheme. Metadata of the included isolates are shown in Table S1 in the supplemental material.

**Population genetic analysis.** Population genetic analyses were performed using BAPS software (30) with the second-order Markov model and the standard MLST data input option, similarly as described in references 31, 32, 34, and 36. The optimal number of clusters was calculated using 10 runs of the estimation algorithm, with the prior upper bound of the number of clusters varying in the range of 10 to 30 over the 10 replicates. All estimation runs yielded an identical partition of the ST data with 7 clusters (estimated *P* value of 1.000). Admixture analysis was done using 100 Monte Carlo replicates for allele frequencies and by generating 100 reference genotypes to calculate *P* values. For reference cases, we used 10 iterations in estimation according to the guidelines of reference 50. Mosaicism is defined as STs composed of sequence characteristic of more than one BAPS group. Significance of admixture or mosaicism was determined for each ST using the threshold of *P* values of $<0.05$. Phylogenetic analysis of *E. faecium* BAPS 2 and 3 (sub)groups was performed using MEGA version 4 (51). A neighbor-joining (NJ) tree was constructed with the maximum composite likelihood model, assuming rate uniformity and pattern homogeneity.

The presence of *E. faecalis* STs in the data set leads to a substantial number of sites where molecular variation is fixed between the two species, reducing the resolution of clustering within the *E. faecium* population. We therefore performed a consecutive clustering of the *E. faecium* BAPS groups 2 and 3 using only the data from each respective group. Also, in this analysis, 10 replicate runs of the estimation algorithm were used, with the prior upper bound of the number of clusters varying in the range of 2 to 10. Again, the resulting estimates of the partition of STs were strongly supported by the posterior distribution (estimated *P* value of 1.000). This second run finally yielded 13 BAPS clusters within *E. faecium* and one cluster including all *E. faecalis* isolates. Although the *E. faecalis* STs were all assigned to a single cluster, one cannot rule out the presence of multiple lineages among them: the limited number of *E. faecalis* STs available for this study means it is not possible to make detailed inferences about the *E. faecalis* population. Relationships between STs were estimated by eBURST (3).

**Calculation of odds ratios.** ORs and 95% confidence intervals (CIs) were calculated relative to all other BAPS groups in the database. Associations were calculated between BAPS groups and lineage and source as *ad/bc*, where *a* is the number of isolates from hospitalized patients, farm animal related, or other sources (surveillance isolates from nonhospitalized persons, isolates from pet animals, and environmental isolates), falling into a given BAPS group, *b* is the number of isolates falling into the same BAPS groups but lacking the particular feature in question, and *c* and *d* are calculated analogously to *a* and *b*, respectively, but using all isolates in the remainder of the *E. faecium* database. The total number of isolates from a source category falling into a given BAPS group, as well as the total number of isolates from that source category, is indicated in Table 2. The total number of isolates included in this analysis was 1,645, because strains from unknown sources and *E. faecalis* were excluded here.

To correct for any bias arising from repeated records in the MLST database with the same genotype and from the same source, ORs and 95% confidence intervals for associations of individual STs with source were also calculated as indicated above. This "clone correction" will mitigate the bias that could be introduced by multiple testing of the same genotype due to outbreaks or so-called microepidemics. The total number of STs from a source category falling into a given BAPS group is indicated in Table S2 in the supplemental material as well as the total number of STs from that source category. The total number of STs included in this analysis was 511, leaving out the *E. faecalis* STs and STs from isolates from unknown sources. STs found in multiple source categories were counted multiple times.

**Phylogenomics analysis.** Phylogenomic analysis of *E. faecium* was performed using 29 previously sequenced *E. faecium* genomes. A set of 299 protein sequences were identified in all 29 *E. faecium* genomes using BLAT software version 33×5 (52) out of an initial set of 649 orthologous protein sequences selected previously (15). Protein sequences were aligned and concatenated using Geneious Pro 4.8.4, and subsequently phylogenetic reconstruction was inferred using the minimum evolution method, including bootstrapping with 1,000 iterations, using MEGA 4 (51).

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00151-12/-/DCSupplemental.

Figure S1, EPS file, 0.4 MB.
Figure S2, EPS file, 0.3 MB.
Figure S3, EPS file, 0.3 MB.
Figure S4, TIF file, 0.3 MB.
Table S1, PDF file, 0.1 MB.
Table S2, PDF file, 0.1 MB.
Table S3, PDF file, 0.1 MB.
Table S4, PDF file, 0.1 MB.

## ACKNOWLEDGMENTS

We are grateful to Miranda van Luit-Asbroek for technical assistance. We also thank the expert referees, whose comments considerably improved the manuscript.

This publication made use of the multilocus sequence typing website (http://www.mlst.net) at Imperial College London developed by David Aanensen and funded by the Wellcome Trust.

R.J.L.W.'s and W.V.S.'s research leading to these results has received funding from the European Union Seventh Framework Programme (FP7-HEALTH-2011-single-stage) under grant agreement no. 282004, EvoTAR. W.V.S. was also funded through the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO-VENI grant 916.86.044). J.C.'s work was supported by ERC grant no. 239784, Academy of Finland grant no. 251179, and a grant from Sigrid Juselius Foundation. W.P.H. was supported by funding from NIH/NIGMS GM088558-01 for the MIDAS Center for Communicable Disease Dynamics at HSPH.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

The authors have declared that no competing interests exist.

## REFERENCES

1. **Hidron AI, et al.** 2008. NHSN annual update: antimicrobial-resistant pathogens associated with healthcare-associated infections: annual summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2006–2007. Infect. Control Hosp. Epidemiol. **29**:996–1011.
2. **Homan WL, et al.** 2002. Multilocus sequence typing scheme for *Enterococcus faecium*. J. Clin. Microbiol. **40**:1963–1971.
3. **Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG.** 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J. Bacteriol. **186**: 1518–1530.
4. **Willems RJ, et al.** 2005. Global spread of vancomycin-resistant *Enterococcus faecium* from distinct nosocomial genetic complex. Emerg. Infect. Dis. **11**:821–828.
5. **Willems RJ, Hanage WP, Bessen DE, Feil EJ.** 2011. Population biology of Gram-positive pathogens: high-risk clones for dissemination of antibiotic resistance. FEMS Microbiol. Rev. **35**:872–900.
6. **Woodford N.** 2008. Successful, multiresistant bacterial clones. J. Antimicrob. Chemother. **61**:233–234.
7. **Werner G, et al.** 2008. Emergence and spread of vancomycin resistance among enterococci in Europe. Euro Surveill. **13**:pii=19046.
8. **Galloway-Peña JR, Nallapareddy SR, Arias CA, Eliopoulos GM, Murray BE.** 2009. Analysis of clonality and antibiotic resistance among early clinical isolates of *Enterococcus faecium* in the United States. J. Infect. Dis. **200**:1566–1573.
9. **Leavis HL, Bonten MJ, Willems RJ.** 2006. Identification of high-risk enterococcal clonal complexes: global dispersion and antibiotic resistance. Curr. Opin. Microbiol. **9**:454–460.
10. **Coque TM, Willems R, Cantón R, Del Campo R, Baquero F.** 2002. High

occurrence of *esp* among ampicillin-resistant and vancomycin-susceptible *Enterococcus faecium* clones from hospitalized patients. J. Antimicrob. Chemother. **50**:1035–1038.
11. **Coque TM, et al.** 2005. Population structure of *Enterococcus faecium* causing bacteremia in a Spanish university hospital: setting the scene for a future increase in vancomycin resistance? Antimicrob. Agents Chemother. **49**:2693–2700.
12. **Leavis HL, et al.** 2003. Epidemic and nonepidemic multidrug-resistant *Enterococcus faecium*. Emerg. Infect. Dis. **9**:1108–1115.
13. **Leavis HL, Willems RJ, Top J, Bonten MJ.** 2006. High-level ciprofloxacin resistance from point mutations in *gyrA* and *parC* confined to global hospital-adapted clonal lineage CC17 of *Enterococcus faecium*. J. Clin. Microbiol. **44**:1059–1064.
14. **Leavis HL, et al.** 2007. Insertion sequence-driven diversification creates a globally dispersed emerging multiresistant subspecies of *E. faecium*. PLoS Pathog. **3**:75–96.
15. **van Schaik W, et al.** 2010. Pyrosequencing-based comparative genome analysis of the nosocomial pathogen *Enterococcus faecium* and identification of a large transferable pathogenicity island. BMC Genomics **11**:239.
16. **Bonten MJ, Willems R, Weinstein RA.** 2001. Vancomycin-resistant enterococci: why are they here, and where do they come from? Lancet Infect. Dis. **1**:314–325.
17. **Heikens E, Bonten MJ, Willems RJ.** 2007. Enterococcal surface protein Esp is important for biofilm formation of *Enterococcus faecium* E1162. J. Bacteriol. **189**:8233–8240.
18. **Leavis H, et al.** 2004. A novel putative enterococcal pathogenicity island linked to the *esp* virulence gene of *Enterococcus faecium* and associated with epidemicity. J. Bacteriol. **186**:672–682.
19. **Hendrickx AP, van Wamel WJ, Posthuma G, Bonten MJ, Willems RJ.** 2007. Five genes encoding surface-exposed LPXTG proteins are enriched in hospital-adapted *Enterococcus faecium* clonal complex 17 isolates. J. Bacteriol. **189**:8321–8332.
20. **Hendrickx AP, et al.** 2008. Expression of two distinct types of pili by a hospital-acquired *Enterococcus faecium* isolate. Microbiology **154**: 3212–3223.
21. **Hendrickx APA, et al.** 2009. SgrA, a nidogen-binding LPXTG surface adhesin implicated in biofilm formation, and EcbA, a collagen binding MSCRAMM of hospital-acquired *Enterococcus faecium*. Infect. Immun.**77**:5097–5106.
22. **Leendertse M, et al.** 2009. Enterococcal surface protein transiently aggravates *Enterococcus faecium* induced urinary tract infection in mice. J. Infect. Dis. **200**:1162–1165.
23. **Heikens E, et al.** 2011. Contribution of the enterococcal surface protein Esp to pathogenesis of *Enterococcus faecium* endocarditis. Microbes Infect. **13**:1185–1190.
24. **Heikens E, van Schaik W, Leavis HL, Bonten MJ, Willems RJ.** 2008. Identification of a novel genomic island specific to hospital-acquired clonal complex 17 *Enterococcus faecium* isolates. Appl. Environ. Microbiol. **74**:7094–7097.
25. **Willems RJ, Bonten MJ.** 2007. Glycopeptide-resistant enterococci: deciphering virulence, resistance and epidemicity. Curr. Opin. Infect. Dis. **20**:384–390.
26. **Willems RJ, van Schaik W.** 2009. Transition of *Enterococcus faecium* from commensal organism to nosocomial pathogen. Future Microbiol. **4**:1125–1135.
27. **Willems RJ.** 2010. Population genetics of *Enterococcus*, p 195–216. *In* Robinson DA, Falush D, Feil EJ (ed), Bacterial population genetics in infectious diseases. John Wiley & Sons, Hoboken, NJ.
28. **Turner KM, Hanage WP, Fraser C, Connor TR, Spratt BG.** 2007. Assessing the reliability of eBURST using simulated populations with known ancestry. BMC Microbiol. **7**:30.
29. **Corander J, Tang J.** 2007. Bayesian analysis of population structure based on linked molecular information. Math. Biosci. **205**:19–31.
30. **Corander J, Marttinen P, Sirén J, Tang J.** 2008. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. BMC Bioinformatics **9**:539.
31. **Tang J, Hanage WP, Fraser C, Corander J.** 2009. Identifying currents in the gene pool for bacterial populations using an integrative approach. PLoS Comput. Biol. **5**:e1000455.
32. **Hanage WP, Fraser C, Tang J, Connor TR, Corander J.** 2009. Hyper-recombination, diversity, and antibiotic resistance in *pneumococcus*. Science **324**:1454–1457.
33. **Gordon DM, Clermont O, Tolley H, Denamur E.** 2008. Assigning *Escherichia coli* strains to phylogenetic groups: multi-locus sequence typing versus the PCR triplex method. Environ. Microbiol. **10**:2484–2496.
34. **de Haan CP, Kivistö RI, Hakkinen M, Corander J, Hänninen ML.** 2010. Multilocus sequence types of Finnish bovine *Campylobacter jejuni* isolates and their attribution to human infections. BMC Microbiol. **10**:200.
35. **Sheppard SK, et al.** 2011. Niche segregation and genetic structure of *Campylobacter jejuni* populations from wild and agricultural host species. Mol. Ecol. **20**:3484–3490.
36. **Corander J, Connor RR, O'Dwyer CA, Kroll JS, Hanage WP.** 2012. Population structure in the *Neisseria*, and the biological significance of fuzzy species. J. R. Soc. Interface **9**:1208–1215.
37. **Marttinen P, Myllykangas S, Corander J.** 2009. Bayesian clustering and feature selection for cancer tissue samples. BMC Bioinformatics **10**:90.
38. **Willems RJL, et al.** 2000. Host specificity of vancomycin-resistant *Enterococcus faecium*. J. Infect. Dis. **182**:816–823.
39. **Panesso D, et al.** 2010. Molecular epidemiology of vancomycin-resistant *Enterococcus faecium*: a prospective, multicenter study in South American hospitals. J. Clin. Microbiol. **48**:1562–1569.
40. **Zheng B, et al.** 2007. Molecular characterization of vancomycin-resistant *Enterococcus faecium* isolates from mainland China. J. Clin. Microbiol. **45**:2813–2818.
41. **Palmer KL, et al.** 2012. Comparative genomics of enterococci: variation in *Enterococcus faecalis*, clade structure in *E. faecium*, and defining characteristics of *E. gallinarum* and *E. casseliflavus*. MBio **3**:e00318-11.
42. **Galloway-Peña J, Roh JH, Latorre M, Qin X, Murray BE.** 2012. Genomic and SNP analyses demonstrate a distant separation of the hospital and community-associated clades of *Enterococcus faecium*. PLoS One **7**:e30187.
43. **Lam MMC, et al.** 24 February 2012. Comparative analysis of the first complete *Enterococcus faecium* genome. J. Bacteriol. [Epub ahead of print.] doi:10.1128/JB.00259-12.
44. **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.
45. **Johnson PD, et al.** 2010. A sustained hospital outbreak of vancomycin-resistant *Enterococcus faecium* bacteremia due to emergence of *vanB E. faecium* sequence type 203. J. Infect. Dis. **202**:1278–1286.
46. **Khan MA, et al.** 2010. High prevalence of ST-78 infection-associated vancomycin-resistant *Enterococcus faecium* from hospitals in Asuncion, Paraguay. Clin. Microbiol. Infect. **16**:624–627.
47. **Hsieh YC, Lee WS, Ou TY, Hsueh PR.** 2010. Clonal spread of CC17 vancomycin-resistant *Enterococcus faecium* with multilocus sequence type 78 (ST78) and a novel ST444 in Taiwan. Eur. J. Clin. Microbiol. Infect. Dis. **29**:25–30.
48. **Zhu X, et al.** 2009. Molecular characterisation of outbreak-related strains of vancomycin-resistant *Enterococcus faecium* from an intensive care unit in Beijing, China. J. Hosp. Infect. **72**:147–154.
49. **Palazzo IC, Pitondo-Silva A, Levy CE, da Costa Darini AL.** 2011. Changes in vancomycin-resistant *Enterococcus faecium* causing outbreaks in Brazil. J. Hosp. Infect. **79**:70–74.
50. **Corander J, Marttinen P.** 2006. Bayesian identification of admixture events using multilocus molecular markers. Mol. Ecol. **15**:2833–2843.
51. **Tamura K, Dudley J, Nei M, Kumar S.** 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24**:1596–1599.
52. **Kent WJ.** 2002. BLAT—the blast-like alignment tool. Genome Res. **12**: 656–664.