JOURNAL OF BACTERIOLOGY, Nov. 2007, p. 8321–8332 Vol. 189, No. 22
0021-9193/07/$08.00+0 doi:10.1128/JB.00664-07
Copyright © 2007, American Society for Microbiology. All Rights Reserved.

Five Genes Encoding Surface-Exposed LPXTG Proteins Are Enriched in Hospital-Adapted *Enterococcus faecium* Clonal Complex 17 Isolates[▽]

Antoni P. A. Hendrickx,[1]* Willem J. B. van Wamel,[2] George Posthuma,[3] Marc J. M. Bonten,[1,4,5] and Rob J. L. Willems[1]

Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, The Netherlands,[1] Department of Medical Microbiology and Infectious Diseases, Erasmus MC-University Medical Center Rotterdam, Rotterdam, The Netherlands,[2] and Cell Microscopy Center, Department of Cell Biology, Institute of Biomembranes,[3] Department of Internal Medicine, Division of Acute Internal Medicine and Infectious Diseases,[4] and Julius Center for Health Sciences and Primary Care,[5] University Medical Center Utrecht, Utrecht, The Netherlands

Received 27 April 2007/Accepted 6 September 2007

Most *Enterococcus faecium* isolates associated with hospital outbreaks and invasive infections belong to a distinct genetic subpopulation called clonal complex 17 (CC17). It has been postulated that the genetic evolution of CC17 involves the acquisition of various genes involved in antibiotic resistance, metabolic pathways, and virulence. To gain insight into additional genes that may have favored the rapid emergence of this nosocomial pathogen, we aimed to identify surface-exposed LPXTG cell wall-anchored proteins (CWAPs) specifically enriched in CC17 *E. faecium*. Using PCR and Southern and dot blot hybridizations, 131 *E. faecium* isolates (40 CC17 and 91 non-CC17) were screened for the presence of 22 putative CWAP genes identified from the *E. faecium* TX0016 genome. Five genes encoding LPXTG surface proteins were specifically enriched in *E. faecium* CC17 isolates. These five LPXTG surface protein genes were found in 28 to 40 (70 to 100%) of CC17 and in only 7 to 24 (8 to 26%) of non-CC17 isolates ($P < 0.05$). Three of these CWAP genes clustered together on the *E. faecium* TX0016 genome, which may comprise a novel enterococcal pathogenicity island covering *E. faecium* contig 609. Expression at the mRNA level was demonstrated, and immunotransmission electron microscopy revealed an association of the five LPXTG surface proteins with the cell wall. Minimal spanning tree analysis based on the presence and absence of 22 CWAP genes revealed grouping of all 40 CC17 strains together with 18 hospital-derived but evolutionary unrelated non-CC17 isolates in a distinct CWAP-enriched cluster, suggesting horizontal transfer of CWAP genes and a role of these CWAPs in hospital adaptation.

Enterococcus faecium is a commensal organism of the mammalian gastrointestinal tract, but during the last 2 decades it has been widely recognized as an opportunistic pathogen causing serious infections in immunocompromised patients (11, 13). In these patients, *E. faecium* is responsible for urinary tract infections, surgical site infections, bacteremia, and endocarditis. The emergence of *E. faecium* infections was associated with increasing resistance towards different classes of antibiotics, e.g., penicillins, aminoglycosides, and glycopeptides (9). Recent studies have shown that *E. faecium* isolates responsible for the vast majority of clinical infections and hospital outbreaks belong to a distinct genetic subpopulation designated clonal complex 17 (CC17), which has spread globally (46). Key features of CC17 *E. faecium* are high-level resistance to ampicillin and ciprofloxacin and the presence of a putative pathogenicity island harboring the *esp* virulence gene, suggesting that CC17 not only is multiresistant to antibiotics but also may be more virulent than non-CC17 *E. faecium* isolates (6, 15). The genetic evolution of CC17 has probably been a multistep process involving the sequential acquisition of multiple adaptive mechanisms (14). These adaptive mechanisms include resistance genes as well as genes encoding novel metabolic pathways (16), putative virulence genes such as hyl_{Efm} (29), and the *esp* gene (45). Surface-exposed Esp expression in *E. faecium* CC17 isolates quantitatively correlates with initial adherence to polystyrene and biofilm formation (40). In addition, the collagen adhesin Acm, which is associated with increased collagen type I binding, is predominantly expressed at the surfaces of clinical *E. faecium* isolates (22, 24). Both Esp and Acm represent cell wall-anchored surface proteins (CWAPs) which may provide *E. faecium* with a selective advantage in the hospital setting, for instance through biofilm formation and better adherence to extracellular matrix molecules.

CWAPs typically contain an N-terminal signal sequence peptide and a C-terminal cell wall sorting signal (CWS). CWSs consist of a conserved Leu-Pro-X-Thr-Gly (LPXTG) sortase substrate motif (where X denotes any amino acid) followed by a hydrophobic domain and positively charged amino acids (31). After the translocation of the precursor CWAP across the plasma membrane, it becomes covalently anchored to the cell wall peptidoglycan by sortase-mediated transpeptidase activity (18, 19). Various CWAPs and MSCRAMM (*m*icrobial *s*urface *c*omponents *r*ecognizing *a*dhesive *m*acro*m*olecules) of *Staphylococcus aureus*, *Enterococcus faecalis*, and *E. faecium* have

* Corresponding author. Mailing address: Department of Medical Microbiology, University Medical Center Utrecht, Heidelberglaan 100, Rm G04.614, 3584 CX Utrecht, The Netherlands. Phone: 31-30-2506534. Fax: 31-30-2541770. E-mail: a.hendrickx@umcutrecht.nl.
▽ Published ahead of print on 14 September 2007.

Downloaded from http://jb.asm.org/ on September 30, 2020 by guest

been recognized as important virulence factors involved in adhesion, biofilm formation, and invasion (12, 24, 33, 37, 40, 43, 44).

To gain insight in the adaptive mechanisms that may have favored the emergence of CC17 *E. faecium*, we aimed to identify novel actively expressed LPXTG surface proteins in addition to Esp and Acm that are specifically enriched in CC17.

(Part of this study was presented as a poster at the 46th Interscience Conference on Antimicrobial Agents and Chemotherapy, San Francisco, CA, September 27 to 30, 2006 [abstr. G-0147].)

MATERIALS AND METHODS

Bacterial strains and growth conditions. A total of 131 *E. faecium* isolates collected from 20 countries worldwide, representing clinical (n = 31) and hospital outbreak (n = 18) isolates from hospitalized patients, surveillance isolates from hospitalized patients (n = 30) and from nonhospitalized persons (n = 30), 12 isolates from various animals (bison [n = 1], calves [n = 2], cat [n = 1], dogs [n = 2], pigs [n = 2], poultry [n = 2], rodent [n = 1], and ostrich [n = 1]), 2 isolates from animal food products, and 8 environmental isolates, were used in this study. The isolates were obtained from various isolation sites, i.e., bile (n = 1), blood (n = 22), carcass (n = 1), catheter (n = 2), environment (n = 6), feces (n = 72), food (n = 2), liquor (n = 1), peritoneal fluid (n = 2), urine (n = 4), and wound (n = 1), and 18 were from undetermined isolation sites. All 131 *E. faecium* isolates were typed previously by our group using multilocus sequence typing (MLST) (46). Based upon MLST, 40 isolates belonged to the hospital-adapted and multiresistant CC17, and 91 strains represented non-CC17 isolates. All bacterial strains were grown aerobically at 37°C on Trypticase soy agar II (TSA) plates supplemented with 5% sheep blood (Becton Dickinson, Alphen aan den Rijn, The Netherlands). *E. faecalis* strain V583 was used as the negative control (30).

Genome search for genes encoding LPXTG-like cell wall-associated proteins. The draft assembly of the *E. faecium* TX0016 (TEX16, DO) genome deposited at DDBJ/EMBL/GenBank under the project accession number AAAK00000000 was searched for proteins annotated as cell wall surface anchor family, von Willebrand factor type A, Cna B type, or surface proteins from gram-positive cocci (anchor region). Furthermore, CWAPs were identified by searching for the presence of an N-terminal signal peptide sequence within the first 70 amino acids or a CWS. The CWS typically consists of a conserved LPXTG sortase substrate motif (where X denotes any amino acid), a hydrophobic domain, and at least one positively charged amino acid within the last eight residues of the C terminus. The presence of an N-terminal signal sequence was analyzed using the SignalP 3.0 server (http://www.cbs.dtu.dk/services/SignalP) (SignalP score of >0.6) (2). The presence of the conserved LPXTG motif, hydrophobic amino acids, and one or more positively charged amino acids within the last 45 residues of the C-terminal end was examined visually. The LPXTG motif was initially compared to the consensus L[FV]PXT[AL]G[N] sortase cleavage motif of the closely related species *E. faecalis* V583 published at the sortase substrate database website (http://bamics3.cmbi.kun.nl/cgi-bin/jos/sortase_substrates/index.py) (3). In another approach, the LPXTG motif was compared to the [FILMPSVY][AP]X[ATS][GAKNS] and NPX[ST][DGNS] patterns representing distinct sortase substrates (10). To identify similarities, BLAST analyses on protein sequences were performed at the NCBI website (http://www.ncbi.nlm.nih.gov/BLAST/).

PCR analysis. The presence or absence of 22 genes encoding putative CWAPs in the 131 *E. faecium* isolates was determined in first instances by PCR. *E. faecium* cell lysates for PCR were prepared as described previously (38). A total of 5 μl of 1:10-diluted lysate was used in the PCR. To confirm the uniform quality of the lysates used in this study, a control PCR on the housekeeping gene *ddl* (encoding D-alanine, D-alanine ligase) was performed with primers ddl_fw and ddl_rv, which are specific for *E. faecium* prior to use of the lysate (26). In all PCRs, *E. faecium* TX0016 was used as a positive control and *E. faecalis* V583 as a negative control. Reactions were performed in 25-μl volumes with HotStar *Taq* polymerase and HotStar MasterMix buffers (QIAGEN Inc., Venlo, The Netherlands). The oligonucleotide PCR primers were designed from the region encoding the N-terminal region of the 22 putative CWAPs. The sequences of the oligonucleotides used in this study are listed in Table 1. The primers used for the detection of the CWAP genes share the same names as the CWAP genes unless otherwise stated. All oligonucleotides were purchased from Invitrogen Corporation (Breda, The Netherlands).

TABLE 1. Oligonucleotides used in this study

ORF	Primer name[a]	Oligonucleotide sequence (5′→3′)	Start in sequence (5′→3′)[b]
orf371	371_fw	TGACTTCCAATGTACCGACA	161
	371_rv	GCTGCTGCGACTAACACAC	830
orf418	418_fw	CTAACTGGTAACTATGGCTTGT	121
	418_rv	GTCCGTGCTGTCACTTGT	1230
orf773	773_fw	GCATCAGTCATTAACCAGAGTA	139
	773_rv	CCCTGTCAAAGGAATAACG	1050
orf884	884_fw	AATCAGACAGTCCACACAGAG	61
	884_rv	ATGATTCCGCTCCACAGTA	1229
orf903	903_fw	TCAACGGACATACCATACCA	311
	903_rv	TCAGTTGGATTCCATGTGAT	1400
	903_rv2	CTTACCATCAACGATCTGCC	720
orf904	904s_fw	AGACGAGGAAGAGGGCAT	360
orf905	905_fw	GTGACAGATTCATCTAATCAT	1
	905_rv	TCATTTTATTTCCCTCCTATTG	1005
	905_rv2	GTGTGAAATGTATGAATCTTGTC	281
	905s_rv	CTTTGGATACAAGTGGATGG	51
orf906	906s_fw	GACACAGCAGTTCCAGCA	271
orf907	907_fw	GTGACCGGTTTTGATGAAAAC	1
	907_rv	TTAAGCTTCTGTTTCTTGATGGC	816
	907_rv2	GTTGTATTTTCCCGATTGAC	290
orf1901	1901_fw	CCGTCAAATCCAATCCAG	121
	1901_rv	CAGTACCTTGACCAAACTTCG	940
orf1904	1904_fw	AGGCAGATTATGGTGATGTT	251
	1904_rv	GGCTGTTGGTTCTTTATCTG	870
orf1996	1996_fw	TGAAGTTGCTGCTAACGAAC	81
	1996_rv	GCAACTTCGATTCCTTCTAC	1040
orf2008	2008_fw	CTACGGATTTTGGTGAAGAA	191
	2008_rv	TCGCACTTCTATTTCCTGAT	900
orf2009	2009_fw	GCTTGGCTATCTCTGAGGA	182
	2009_rv	TCATAGGCAGAGACTGGAAC	1100
orf2010	2010_fw	GTAGCGAAGAAAATGAGATGG	80
	2010_rv	TAACTTGACTGAATCGGTGC	1101
orf2109	2109_fw	GCAGGTGCAACTTATACATTAG	61
	2109_rv	CTTGATCCGTCGTAATATTGA	811
orf2351	2351_fw	AATGAACGGGCAAATGAG	82
	2351_rv	CTTTTGTTCCTTAGTTGGTATGA	753
orf2356	2356_fw	GTTGGTCATCTTATTTGCTGTA	30
	2356_rv	TGCCCTGCTCCTTCTACTA	1370
orf2430	2430_fw	GCAGTTTACAATGGTGTGAAGCAA	217
	2430_rv	CGGCTAATGAGTATTTGTCGTTCC	1180
	2430_rv2	AGAACCCTCAGTTTCCACAT	840
orf2514	2514_fw	AGTTCCAGTTGCGAGTCAGA	131
	2514_rv	ATGTAGTCGGATTCGGTGC	1120
orf2515	2515_fw	GTGGTAGAATTGACGAAAGA	1
	2515_rv	AACAAGTAGCACACCCAATA	990
orf2569	2569_fw	GTGTTTGCAGAGGAGACAGC	61
	2569_rv	GACAGAATAATTTACTGGGTCG	1182
orf2570	2570_fw	GGTAGCAACAGAAACAGAAAC	66
	2570_rv	GAATACTGCGCCAGCTAATC	1026
orf2571	2571_fw	ATGACGACCACAGGGAAGAA	1
	2571_rv	CCGCTATCTGCTAAAGTATC	980
ddl	ddl_fw	GAGACATTGAATATGCCTTATG	334
	ddl_rv	AAAAAGAAATCGCACCG	893

[a] Annotation: fw, forward; rv, reverse.
[b] Relative to the start codon of the open reading frame.

Southern blot and dot blot analysis. Southern blot and dot blot analysis was performed on chromosomal DNA isolated from 131 *E. faecium* isolates to determine the presence or absence of CWAP genes and to confirm PCR results. Chromosomal DNA isolation was performed as described previously with some minor modifications (1). In brief, a loop of *E. faecium* cells was resuspended in 200 μl 10 mM Tris EDTA plus 10 μl lysozyme (50 mg/ml; Fluka Biochemika, Buchs, Switzerland) and incubated for 15 min at 37°C. After incubation, 30 μl 10% sodium dodecyl sulfate (SDS) and 20 μl proteinase K were added and incubated for 1 h at 65°C. Subsequently, chromosomal DNA was isolated using the DNeasy tissue kit (QIAGEN Inc.) according to the manufacturer's instructions. *E. faecium* TX0016 DNA was used as a positive control and *E. faecalis* V583 DNA as a negative control. Chromosomal DNA was digested with EcoRI (Roche Diagnostics), and fragments were separated by agarose gel electrophoresis and capillary blotted onto a Hybond N^+ nylon membrane (GE Healthcare, Diegem, Belgium). For dot blot analysis, the chromosomal DNA was denatured with 0.5 M NaOH in a 96-well microtiter plate for 15 min, transferred to a 96-well Bio-Rad Bio dot apparatus, and vacuum blotted onto a Hybond N^+ nylon membrane. For both methods, DNA was fixed onto the membrane by incubation

Downloaded from http://jb.asm.org/ on September 30, 2020 by guest

for 2 min in 0.4 M NaOH followed by a neutralization in 10× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) for 1 min. Membranes were hybridized overnight at 42°C with a 100-ng probe. For Southern blot analysis, probes were generated by PCR in 50-μl reaction volumes by use of primer pairs 903_fw and 903_rv2, 905_fw and 905_rv2, 907_fw and 907_rv2, 2351_fw and 2351_rv2, and 2430_fw and 2430_rv. For dot blot analysis, gene-specific probes for *orf371*, *orf418*, *orf773*, *orf884*, *orf1901*, *orf1904*, *orf1996*, *orf2008*, *orf2009*, *orf2010*, *orf2109*, *orf2356*, *orf2514*, *orf2515*, *orf2569*, *orf2570*, and *orf2571* were amplified by PCR using primers depicted in Table 1. Amplified DNA probes were purified with a PCR purification kit (QIAGEN Inc.) and labeled according to directions for the ECL nucleic acid labeling kit (GE Healthcare). The membranes were exposed to Hyperfilm ECL (GE Healthcare).

DNA sequencing. PCR products were sequenced by using the BigDye Terminator 3.1 reaction kit and an ABI PRISM 3100 capillary DNA sequencer (both from Applied Biosystems, Foster City, CA).

Western blot analysis. To detect 904.5 and 906.7 LPXTG protein expression, plate-grown bacteria were resuspended in phosphate-buffered saline (PBS) and harvested by centrifugation (1,560 × *g*, 5 min). To detect 904.5 protein expression, cell pellets were suspended in 70% formic acid and incubated at 65°C for 30 min as described previously (7). For the 906.7 protein, cell pellets were resuspended in 50 μl PBS plus 50 μl sample buffer (100 mM Tris-HCl, 5% dithiothreitol, 2% SDS, 0.004% bromophenol blue, and 20% glycerol) and boiled for 5 min. Equal amounts of protein samples of E135 (negative control) and E380 *E. faecium* were electrophoresed through a 10% SDS-polyacrylamide gel. Prior to electroblotting, the membranes were incubated for 10 min in blot buffer (20 mM Tris, 150 mM glycine, and 20% methanol, pH 8.3). Samples were electroblotted overnight using a Trans-Blot cell tank transfer unit at 15 V onto Trans-Blot nitrocellulose membranes (0.45 μm; Bio-Rad Laboratories Inc., Veenendaal, The Netherlands). The membranes were blocked with 4% skim milk (Campina Holland, Alkmaar, The Netherlands) in PBS-0.1% Tween 20 for 1 h at 37°C. Incubation with mouse anti-904 and mouse anti-906 immune sera (both at a 1:3,000 dilution) was carried out for 1 hour in 1% bovine serum albumin (BSA) in PBS-1% Tween 20 at 37°C, followed by two washes, each of 10 min, in PBS-0.1% Tween 20 at 37°C. As a control, blots were incubated with mouse preimmune sera with similar serum dilutions. Subsequently, membranes were incubated for 1 h with goat anti-mouse immunoglobulin G (IgG) (heavy plus light chains)-horseradish peroxidase (Bio-Rad) in 1% BSA in PBS (PBSb)-1% Tween 20 at 37°C. Membranes were washed twice with PBS-0.1%, Tween 20, and the 904.5 and 906.7 proteins were visualized using the ECL plus Western blotting detection system (GE Healthcare) and exposed to film (Hyperfilm ECL; GE Healthcare).

Peptide mouse antisera. Polyclonal mouse antisera were prepared by Eurogentec (Seraing, Belgium) according to their protocol by immunizing mice with 15-amino-acid-residue keyhole limpet hemocyanin-conjugated peptide directed against the N-terminal regions of proteins 903 (H_2N-QTTEETNSPPYSEIQ-$CONH_2$), 904 (H_2N-CRMFESDKMPSISQND-$CONH_2$), 906 (H_2N-CGTELTF PNSEPLNG-$CONH_2$), 2351 (H_2N-CERANENLSFTVKTDR-$CONH_2$), and 2430 (H_2N-EFTPGYEKNPLPDMSC-$CONH_2$) of the *E. faecium* TX0016 genome sequence.

Generation of cDNA from total mRNA. Eight *E. faecium* isolates, E135, E155, E380, E470, E745, E1165, E1172, and E1176 (4, 36, 39, 45, 46), were assayed for mRNA expression of the five CC17-enriched putative CWAP genes (*orf903*, *orf905*, *orf907*, *orf2351*, and *orf2430*). E135 is a community surveillance isolate and was used as a negative control. E155, E470, and E745 were isolates from hospital outbreaks in the United States and The Netherlands, while E380, E1165, E1172, and E1176 were all clinical isolates recovered from blood, wound, urine, and respiratory tract, respectively. Bacterial strains were grown overnight on TSA at 37°C, scraped off of the plates, resuspended into 3 ml PBS to an optical density at 660 nm of 1.0 (~1×10^9 CFU/ml), and pelleted by centrifugation (6,500 × *g* for 1 min). Total RNA was isolated according to the method of Cheung et al. (5). RNA purification and cDNA generation was done according to the method of Nallapareddy et al. (21). In brief, total RNA was treated three times with 20 U of RQ1 DNase (Promega Corp., Leiden, The Netherlands) for 30 min at 37°C, and RNA was isolated using an RNeasy minikit (QIAGEN Inc.). From 5 μg total RNA, cDNA was synthesized with the SuperScript II first-strand synthesis system (Invitrogen Corp.) using random primers according to the instruction of the manufacturer. cDNA was used as the template for PCR using primer pairs 903_fw and 903_rv2, 905_fw and 905_rv2, 907_fw and 907_rv2, 2351_fw and 2351_rv, and 2430_fw and 2430_rv2 as shown in Table 1. As an internal control, the housekeeping gene *ddl* (encoding D-alanine, D-alanine ligase) was amplified using primers shown in Table 1. RNA samples not treated with reverse transcriptase were used as a control to detect DNA contamination in the total RNA preparations.

Electron microscopy and immunogold labeling. Transmission immunoelectron microscopy was performed as described previously with some modifications (40). In brief, copper grids (mesh Formvar-carbon coated) were incubated for 30 min with carbon side on a drop of 1×10^9 CFU/ml of *E. faecium* E380 cells (CC17 isolate containing *orf904.5* allele 2 and *orf906.7* insertion A) to detect surface-exposed 903, 904.5, and 906.7 protein expression and E470 and E745 cells to detect surface-exposed 2351 and 2430 protein expression, respectively. The E135 strain was used as a negative control. Grids were washed three times for 5 min on drops of 0.02 M glycine in PBS and subsequently blocked for 30 min on drops of 1% PBSb. The five LPXTG surface proteins were labeled for 1 h on drops with 1:100 (for the 903, 904, and 906 sera)-, 1:250 (for the 2430 serum)-, or 1:300 (for the 2351 serum)-diluted specific peptide mouse immune sera in PBSb. The preimmune sera were diluted similarly. Grids were washed four times for 2 min on drops of 0.1% PBSb. Then, grids were incubated for 20 min on drops with 1:250-diluted rabbit anti-mouse IgG antibody (Dako, Glostrup, Denmark) in PBSb and washed four times for 2 min on drops of 0.1% PBSb. Antibody complexes were labeled by incubation for 20 min on drops with 1:55-diluted protein A-gold label (15 nm) in PBSb. Grids were washed four times for 2 min on drops of PBS, fixed by incubation for 5 min on drops of 1% glutaraldehyde in PBS, and washed again eight times for 2 min on drops of H_2O. Bacteria were stained by incubation of the grids for 5 min on drops containing 1.8% methylcellulose (25 centipoises; Sigma-Aldrich, St. Louis, MO) and 0.4% uranyl acetate (pH 4) and subsequently air dried for 10 min. Grids were examined using a Jeol 1010 transmission electron microscope (Jeol Europe, Amsterdam, The Netherlands) at a magnification of ×65,000.

MST analysis. The BioNumerics software, version 4.0 (Applied Maths, Sint-Martens-Latem, Belgium), was used to generate a minimum spanning tree (MST) as described previously (32). The MST was based on the presence and absence of the 22 putative CWAP genes. The categorical coefficient was used to calculate the MST. CWAP gene profile types were linked by the eBURST priority rule by first linking profiles that differ in the presence or absence of only one CWAP gene. When two types have an equal distance to a linkage position in the tree, the type with the highest number of single CWAP gene variants is linked first.

Nucleotide sequence accession numbers. The DNA sequences of the junction regions of the 904.5 and 906.7 (types A, B, and C) open reading frames (ORFs) reported in this article have been deposited in the GenBank nucleotide sequence database under accession numbers EU122157, EU122158, EU122159, EU122160, EU122161, and EU122162.

RESULTS

Genome search for ORFs encoding putative CWAPs. The partially sequenced and annotated genome of *E. faecium* TX0016 was searched for ORFs encoding putative CWAPs. Putative CWAPs were identified by the presence of a C-terminal CWS domain. The genome search identified 22 putative CWAP ORFs, of which 18 contained an N-terminal signal sequence and a CWS domain, 4 ORFs only a CWS and no N-terminal signal sequence, and 1 ORF an N-terminal signal sequence but no positively charged amino acids in the CWS (Table 2). Eleven of the 22 proteins contained the most abundant canonical LPXTG motif, while the remaining proteins had deviant cell wall anchor motifs.

Distribution of 22 putative CWAP-encoding genes among *E. faecium* isolates. The presence of the 22 identified putative CWAP-encoding genes among 131 *E. faecium* isolates was determined by PCR and Southern blot and dot blot hybridizations. Based on this, 17 of the 22 putative CWAP genes were found to be widespread and highly prevalent in the *E. faecium* population (Table 3, lower part). Three ORFs, *orf371*, *orf773*, and *orf2109*, were detected in all 131 isolates, thus presumably belonging to the *E. faecium* core genome. Four ORFs, *orf418*, *orf884*, *orf1996*, and *orf2356*, were detected in 100% of the CC17 isolates and in 91.2 to 98.9% of the non-CC17 isolates. The ORFs *orf371*, *orf773*, *orf1901*, *orf1904*, *orf2008*, *orf2009*, *orf2010*, *orf2109*, *orf2514*, *orf2515*, *orf2569*, *orf2570*, and *orf2571* were detected in 90 to

Downloaded from http://jb.asm.org/ on September 30, 2020 by guest

TABLE 2. Characteristics of the 22 putative LPXTG cell wall-associated proteins

ORF	Signal peptide (probability)	Cleavage site probability	C-terminal cell wall sorting sequence[a]	Accession no.
orf371	Yes (1.000)	0.995	**LPKTN**EKTASVSYSLAGAGVLVAAAYGIIRKKK	ZP_00604782
orf418	Yes (1.000)	0.976	**LPKTG**SQSNNWITLAGVILLVIGLRITFMSYKKSKR	ZP_00604835
orf773	Yes (0.923)	0.853	**IPLTG**ITFPFLSQGGSSLLMLSICVGFVLNISADEKRKSLGL	ZP_00604269
orf884	Yes (0.804)	0.798	**LPKSG**ESQNKVILWSGII LLSIATMLSAKRFKQNRSL	ZP_00603974
orf903	Yes (0.840)	0.787	**LPSTG**GIGLLPFVFLGLIFIFSGFFYFIHRKKKAGEQR	ZP_00604460
orf905	No	0.000	**FPKTS**EEMLGGFSILGLLLVLSTGTAWFYKKQQKGNNRREIK	ZP_00604462
orf907	No	0.000	**LPSTG**GMGIIVFILVGTALVGGAVIYFKKRHQETEA	ZP_00604464
orf1901	Yes (0.999)	0.999	**VPKTG**STHLVTISAVSLLLLVLATFSYAVLRYI	ZP_00603526
orf1904	Yes (0.999)	0.999	**VPMTG**SNGFQTYVLISCLLLGAGALSAVVYFKKKA	ZP_00603529
orf1996	Yes (1.000)	0.914	**LPKTG**EFNNPLLALAGGILLIGVVVYVMKQRKK	ZP_00603029
orf2008	Yes (1.000)	0.999	**LPATG**GNGLLAFLLIGISLMIGAYSWYRKSKMKSEV	ZP_00603041
orf2009	Yes (0.999)	0.797	**FPQTN**EARTLISLLGILLLLGAVLLIKKERGEKNE	ZP_00603042
orf2010	No	0.000	**LPATG**GSGRLFYILLAAFFMALAGAAYGLFVRQSKKEGAR	ZP_00603043
orf2109	Yes (0.993)	0.512	**QPITN**NTLLEQGQKKLKITAKDTSINITTDQEDNETD	ZP_00602865
orf2351	Yes (1.000)	0.782	**LPETG**SRTLNKLITWMGVLLILIVGASYFRSLFHRVK	ZP_00602747
orf2356	Yes (0.989)	0.938	**LPKTG**EKKHSLLIYSGGFLSVSTIITWIWRRKRG	ZP_00602752
orf2430	Yes (1.000)	1.000	**FPQTG**EKNSNVLLFIGFTLIFATAGYYFWNRRN	ZP_00603098
orf2514	Yes (1.000)	0.978	**LTPGG**VVLTKIDDQSGEILQELFLSYRTEKEKRYKPV	ZP_00602630
orf2515	No	0.000	**LPKTG**ETIISQFILSILGVLLVFISIKFRKKRNI	ZP_00602631
orf2569	Yes (1.000)	0.997	**LPSTG**GKGIYVYIGAGVVLLLIAGLYFARRKHSQI	ZP_00602687
orf2570	Yes (0.678)	0.516	**LPKTN**ETKNTLLGVVGMVFASFAIWLFIKKRTGVKK	ZP_00602688
orf2571	Yes (1.000)	0.997	**LPETG**GIGRLGIYLVGMIGCAFSIWYLFLKKERGGS	ZP_00602689

[a] The LPXTG-like motifs are depicted in bold, and the positively charged C-terminal amino acids are underlined.

100% of the CC17 *E. faecium* isolates and in 40.7 to 85.7% of the non-CC17 isolates (Table 3).

Five of the 22 putative CWAP genes were specifically enriched in CC17 *E. faecium* isolates recovered from clinical sites or associated with hospital outbreaks (Table 3, upper part). Three putative CWAP genes, *orf903*, *orf905*, and *orf907*, clustered together on contig 609 (gi 68194716) on the *E. faecium* TX0016 genome and were detected in all (40/40) of the CC17 *E. faecium* isolates (Fig. 1A). This cluster was present only in a minor subset of non-CC17 isolates, *orf903* in 24.2% (22/91) and *orf905* and *orf907* in 22.0% (20/91). The 20 non-CC17 isolates that contain this cluster of putative CWAP genes represented hospital outbreak isolates (3), clinical isolates (6), hospital surveillance isolates (5), and community surveillance isolates (6). Thus, the majority (14/20) of these non-CC17 isolates were also hospital derived. In addition, the *orf903* gene was also found in one extra hospital surveillance isolate and one community surveillance isolate. The predicted 903 protein is 773 amino acids in size, and the C-terminal part has the highest sequence similarity with a CWAP of *Listeria monocytogenes* strain 1/2a F6854 designated LMOf6854_0833, with an unknown function (Table 4). The predicted 905 protein is 338 amino acids in size and contains one Cna B-type domain in the C-terminal part of the protein. This Cna B-type domain was found in the *S. aureus* collagen binding protein but is not involved in collagen binding (34). It has the highest similarity with the endocarditis- and biofilm-associated pilin protein EbpB (previously known as EF1092), a CWAP of *E. faecalis* OG1RF (Table 4) (23). The predicted 907 protein is 371 amino acids in size and contains two Cna B-type domains, one located in the N-terminal part and one in the C-terminal part of the protein, and has the highest similarity with the biofilm enhancer in enterococcus 3 protein (Bee3) of *E. faecalis* (Table 4) (35). The presence of putative transcriptional regulators upstream of *orf903* (LuxR; *orf902*) and downstream of *orf907* (MgA-like; *orf910*, MerR; *orf912*, LysR; *orf913*, and *orf914*), direct repeats upstream of *orf901* and downstream of *orf911*, and an integrase (*orf901*) may imply that this region encompasses a distinct genomic island that is acquired through horizontal gene transfer.

TABLE 3. Incidence of 22 putative CWAP genes in hospital-adapted *E. faecium* CC17 and non-CC17 isolates

Description	ORF	Distribution [no. (%)] of the 22 putative CWAP genes among:	
		CC17 isolates ($n = 40$)	Non-CC17 isolates ($n = 91$)
CC17 enriched	*orf903*	40 (100)	22 (24.2)
	orf905	40 (100)	20 (22.0)
	orf907	40 (100)	20 (22.0)
	orf2351	40 (100)	24 (26.4)
	orf2430	28 (70)	7 (7.7)
Widespread	*orf371*	40 (100)	91 (100)
	orf418	40 (100)	84 (92.3)
	orf773	40 (100)	91 (100)
	orf884	40 (100)	91 (98.9)
	orf1901	36 (90)	38 (41.8)
	orf1904	39 (97.5)	78 (85.7)
	orf1996	40 (100)	91 (98.9)
	orf2008	38 (95)	77 (84.6)
	orf2009	37 (92.5)	72 (79.1)
	orf2010	37 (92.5)	37 (40.7)
	orf2109	40 (100)	91 (100)
	orf2356	40 (100)	83 (91.2)
	orf2514	40 (100)	68 (74.7)
	orf2515	40 (100)	70 (76.9)
	orf2569	40 (100)	70 (76.9)
	orf2570	40 (100)	67 (73.6)
	orf2571	40 (100)	70 (76.9)

Putative CWAP gene *orf2351* was detected in all (40/40) of

Downloaded from http://jb.asm.org/ on September 30, 2020 by guest

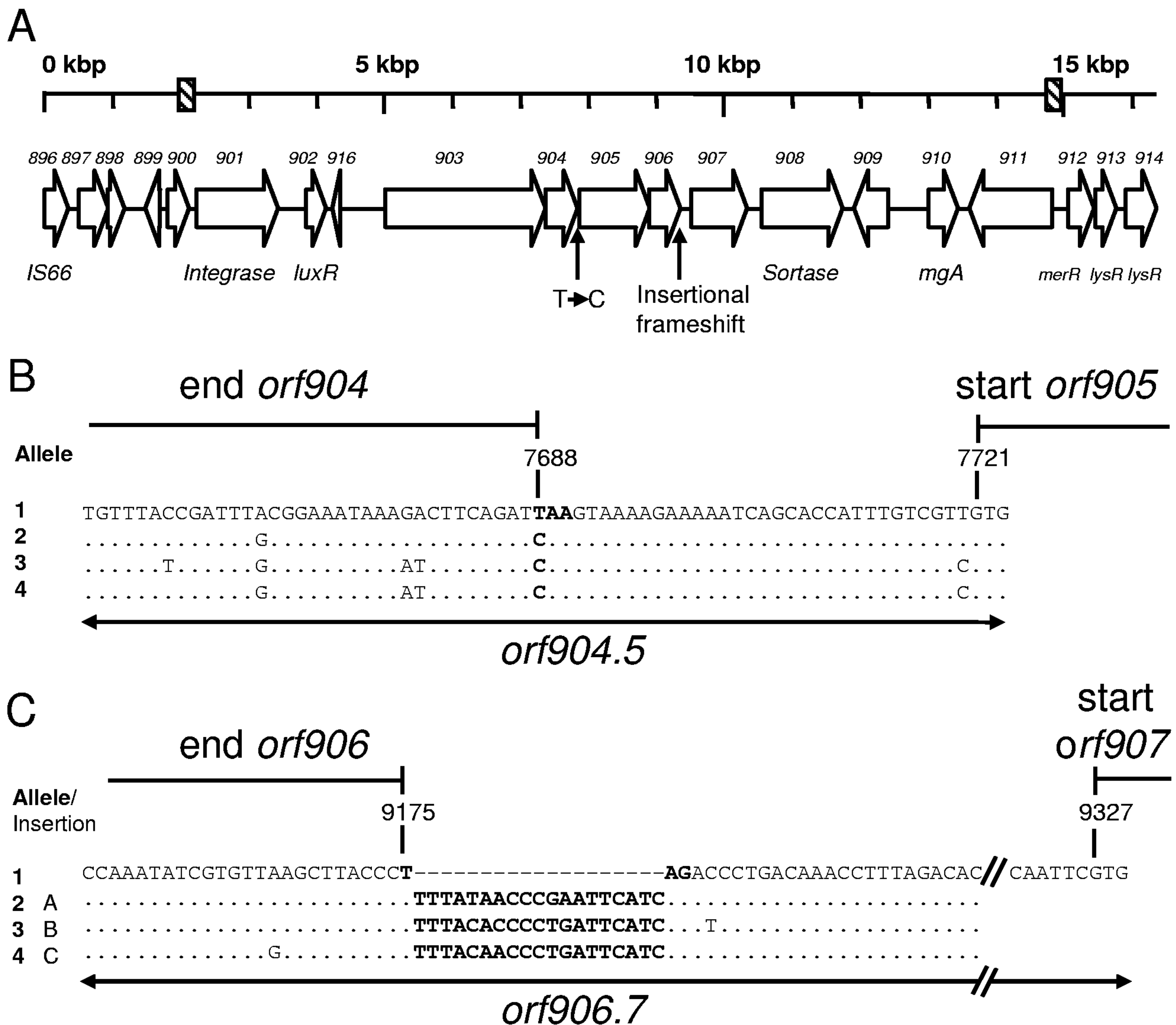

FIG. 1. The genomic cluster (*E. faecium* contig 609) enriched in hospital-adapted *E. faecium* CC17 isolates. (A) The genomic organization of contig 609 as published at DDBJ/EMBL/GenBank. Direct repeats are indicated as dashed boxes in the scale at positions 2030 and 14751 of contig 609. The thymine-to-cytosine point mutation and the 19-bp oligonucleotide insertion are indicated with black arrows. The position and presumed direction of transcription of ORFs have been indicated by open arrows. (B) Alignment of the *orf904-905* junction region. The TAA stop codon is depicted in bold for strain TX0016 at position 7688 (allele 1). In alleles 2, 3, and 4, a thymine-to-cytosine point mutation leads to a merged *orf904.5*. (C) Different alleles in the *orf906-907* junction region. Allele 1 is found in 29 *E. faecium* isolates, including TX0016. Alleles 2, 3, and 4 resulted from three different 19-bp insertions (A, B, and C) and were found in 26, 3, and 2 isolates, respectively. The TAG stop codon is depicted in bold for strain TX0016 at position 9175 (allele 1) and is disrupted by the 19-bp insertion to yield *orf906.7*.

the CC17 *E. faecium* isolates and in 26.4% (24/91) of the non-CC17 isolates. Of the 91 non-CC17 isolates, the 24 which harbor *orf2351* included clinically relevant (2 hospital outbreak and 7 clinical isolates) as well as less clinically relevant (9 surveillance, 3 animal, and 3 environmental *E. faecium* isolates) isolates. The predicted 2351 protein is 324 amino acids in size and contains a serine-rich repeat region close to the C terminus. It has the highest similarity to a CWAP of *E. faecalis* V583 designated EF0093 (Table 4). The serine-rich repeat region of the 2351 protein is highly similar to the serine-rich repeat domain of the clumping factor B protein of *Staphylococcus aureus* subsp. *aureus* NCTC 8325.

orf2430 was detected in 70% (28/40) of the CC17 *E. faecium* isolates and in only 7.7% (7/91) of the non-CC17 isolates. The majority of the non-CC17 *E. faecium* isolates (six of seven) harboring *orf2430* are hospital-derived isolates (two hospital outbreak isolates, three clinical isolates, and one hospital surveillance isolate). The predicted 2430 protein is 1075 amino acids in size and contains four conserved Cna B domains. The protein has the highest similarity with a CWAP of *E. faecalis* V583 designated EF1896 (Table 4) and a collagen adhesion protein of *Bacillus cereus* ATCC 14579.

Sequence heterogeneity in the *orf903-907* gene cluster. The *orf904* gene in the *orf903-907* cluster of the sequenced *E. faecium* TX0016 genome is predicted to encode a protein with an N-terminal signal peptide sequence (signal peptide probability,

Downloaded from http://jb.asm.org/ on September 30, 2020 by guest

TABLE 4. Homologies of the five LPXTG surface proteins enriched in CC17 *E. faecium* isolates as determined by BLAST

Protein	BLAST hit	Function	Organism	% Amino acid identity
903	LMOf6854_0833	Unknown	*L. monocytogenes* 1/2a F6854	40
905	EbpB	Biofilm formation	*E. faecalis* OG1RF	36
904.5	EbpB	Biofilm formation	*E. faecalis* OG1RF	40
907	Bee3	Biofilm formation	*E. faecalis*	29
906.7	Bee3	Biofilm formation	*E. faecalis*	28
2351	EF0093	Unknown	*E. faecalis* V583	57
2430	EF1896	Unknown	*E. faecalis* V583	90

1.000; cleavage site probability, 1.000) but no CWS. In contrast, *orf905* is predicted to encode a protein with no detectable signal peptide sequence but with an intact CWS (Fig. 1A). Detailed examination of the *orf904-905* junction region suggests that *orf904* in the sequenced *E. faecium* TX0016 genome contains a premature TAA stop codon at position 7688 of contig 609 (Fig. 1B, allele 1). The region encompassing this stop codon was amplified by PCR in all isolates positive for *orf905* (60 of 131) by use of primers 904s_fw and 905s_rv to yield a 191-bp DNA fragment that was subsequently sequenced. Sequencing revealed four different alleles in the *orf904-905* junction region. In 13 of the 60 isolates, a single thymine-to-cytosine point mutation changed the TAA stop codon into CAA, resulting in a merging of the *orf904-905* genes, leading to a potential functional CWAP gene designated *orf904.5*, encoding a protein with an N-terminal signal sequence and a CWS (Fig. 1A and B). The T-to-C point mutation was found in two hospital outbreak isolates, four clinical isolates, two hospital surveillance isolates, and five community surveillance isolates.

Similarly to *orf904-905*, *orf906* encodes a putative protein with an N-terminal signal peptide sequence (signal peptide probability, 1.000; cleavage site probability, 0.975) but no CWS, while *orf907* encodes a putative protein with no detectable signal peptide sequence but with an intact CWS. Examination of the *orf906-907* junction region suggests that *orf906* contains a premature TAG stop codon (Fig. 1C, allele 1). The region encompassing this stop codon was amplified by PCR in all isolates positive for *orf907* (60 of 131) by use of primers 906s_fw and 907_rv2 and yielded a 616-bp product. Sequencing of this fragment revealed that in 31 of the 60 isolates, a 19-bp oligonucleotide insertion at position 9175 of contig 609 caused a frameshift which merged the *orf906-907* genes, leading to a putative functional CWAP gene designated *orf906.7* encoding a protein with an N-terminal signal sequence and a CWS (Fig. 1C). In total, four different alleles were found. Allele 1 is present in TX0016, while alleles 2 to 4 resulted from three different types of 19-bp oligonucleotide insertions. Twenty-six isolates contained the TTTATAACCCGAATTCATC insertion (type A), three isolates contained the TTTACACCCCT GATTCATC insertion (type B), and two contained the TTT ACAACCCTGATTCATC insertion (type C). All three (A, B, and C) insertions putatively encode different amino acid sequences, i.e., FYNPNSS, FYTPDSS, and FYNPDSS, respectively. The predominant insertion A was found in nine clinical, seven outbreak-associated, six hospital surveillance, and four community surveillance isolates. The majority of the strains (22/26) with this insertion are hospital-derived isolates. Insertion B was found in one outbreak-associated isolate, one hospital surveillance isolate, and one community surveillance isolate. Insertion C was found in one clinical and in one community surveillance isolate. Only 12 of the 131 isolates (9.2%) carry both the T-to-C point mutation in the *orf904-905* junction and the insertion in the *orf906-907* junction. Seven of the 12 of these isolates are hospital-associated isolates.

Newly identified *orf904.5* and *orf906.7* encode 904.5 and 906.7 LPXTG proteins. To confirm that *orf904.5* and *orf906.7* encode 904.5 and 906.7 LPXTG proteins, Western blotting was performed on *E. faecium* strain E380 (containing *orf904.5* allele 2 and *orf906.7* insertion A) and strain E135 (negative control). Mouse antipeptide immune sera raised against the 904 and 906 proteins reacted with 48-kDa and ~70-kDa protein bands, respectively, in protein extracts of E380 but not in negative control E135 (Fig. 2), while mouse preimmune sera did not react with these proteins (data not shown). The observed molecular mass of the 904.5 protein correlated well with the predicted molecular mass of 48.05 kDa (after posttranslational modification). In contrast, the observed molecular mass of the 906.7 protein is higher than the predicted molecular mass (44.62 kDa with a pI of 4.4 after posttranslational modification). Interestingly, a similar effect was also observed for the Acm LPXTG surface protein of *E. faecium* (24). This aberrant migration in SDS-polyacrylamide gel electrophoresis is possibly due to an association with peptidoglycan or the acidic nature of the 906.7 LPXTG protein.

mRNA expression of the five CC17-enriched genes after growth in vitro on TSA. Total RNA was isolated of one com-

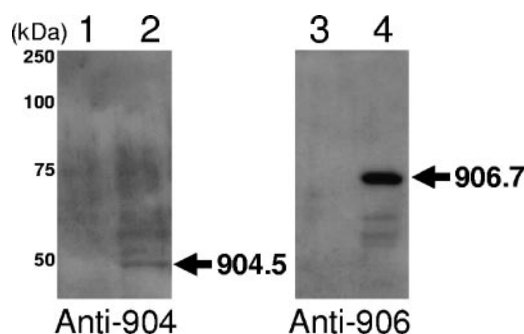

FIG. 2. Western blots of protein extracts of *E. faecium* E135 and E380 isolates probed with mouse anti-904 (left) and mouse anti-906 (right) immune sera. The Western blots show expression of the 904.5 and 906.7 LPXTG proteins in *E. faecium* E380 (lanes 2 and 4), as depicted by arrows, and not in the negative control, E135 (lanes 1 and 3). Numbers depicted on the left indicate molecular masses in kDa.

Downloaded from http://jb.asm.org/ on September 30, 2020 by guest

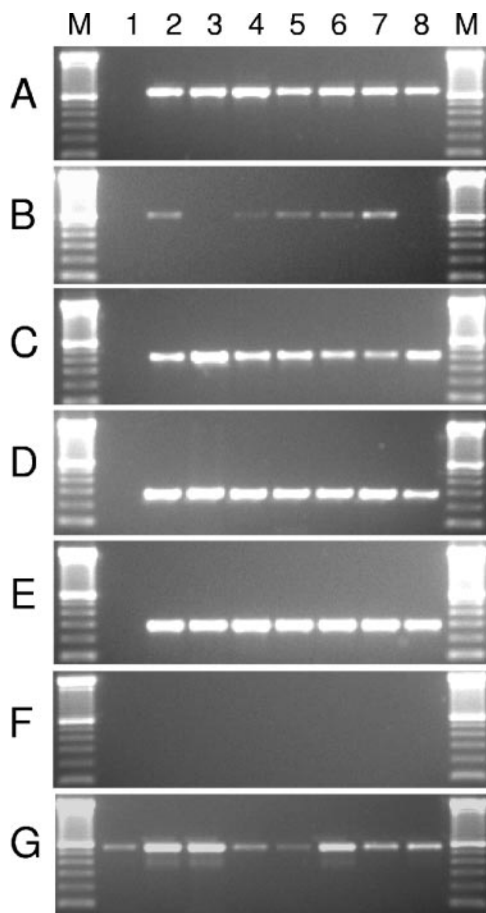

FIG. 3. mRNA expression of *orf903*, *orf905*, *orf907*, *orf2351*, and *orf2430*. Depicted is the mRNA of three outbreak-associated isolates, E155, E470, and E745 (lanes 2, 4, and 5), and of four clinical isolates, E380, E1165, E1172, and E1176 (lanes 3 and 6 to 8). Lane 1 contains mRNA of the E135 negative control, a community *E. faecium* isolate lacking all five CC17-enriched putative CWAP genes. Strains E380 and E1176 (lanes 3 and 8) are deficient for *orf2430*. (A to E) mRNA expression of the putative CWAP genes *orf2351*, *orf2430*, *orf903*, *orf905*, and *orf907*, respectively. (F) Control *ddl* PCRs on total mRNA preparations in which the reverse transcription reaction was omitted were all negative, demonstrating the absence of DNA contamination. (G) Control *ddl* reverse transcription-PCRs (internal housekeeping control) with *E. faecium*-specific *ddl* primers were all positive. The results are presented as amplified PCR products electrophoresed on the same ethidium bromide-stained 1.5% agarose gel. Lanes M show molecular mass markers.

munity surveillance isolate (E135; negative control), three outbreak-associated isolates (E155, E470, and E745), and four clinical *E. faecium* isolates (E380, E1165, E1172, and E1176) after growth on TSA plates overnight at 37°C. The mRNA expression of the five CC17-enriched putative CWAP genes (*orf903*, *orf905*, *orf907*, *orf2351*, and *orf2430*) and of *ddl* (internal housekeeping control) was analyzed by PCR on cDNA generated from total mRNA. mRNA transcripts of *orf903*, *orf905*, *orf907*, and *orf2351* were detected in seven CC17 isolates which harbor these ORFs, indicating that these ORFs are expressed and do not represent silent genes (Fig. 3A to E). mRNA transcripts of *orf2430* were detected in E155, E470, E745, E1165, and E1172, although at levels lower than those of *orf903*, *orf905*, *orf907*, and *orf2351*, suggesting that *orf2430* transcription is low during growth on TSA at 37°C (Fig. 3B). As expected, no *orf2430* mRNA transcripts were detected for E380 and E1176 (lanes 3 and 8), two strains deficient for *orf2430*. No DNA contamination was detected in control *ddl* reactions of the DNase-treated total mRNA preparations in which the reverse transcriptase reaction was omitted (Fig. 3F). The control *ddl* housekeeping gene was expressed in all isolates analyzed (Fig. 3G).

Surface-exposed expression of the 903, 904.5, 906.7, 2351, and 2430 LPXTG proteins. Surface expression of all five CC17-enriched putative surface proteins was demonstrated by transmission electron microscopy of negatively stained immunogold-labeled bacteria. By use of peptide antisera directed against the five surface proteins, gold particles were clearly associated with the cell walls of strains positive for the respective genes (Fig. 4A to E). In contrast, no gold particles were associated with the cell wall when cells were incubated with the individual mouse preimmune sera (Fig. 4F). No gold particles were detected at the cell wall when immune sera were omitted (conjugate control; rabbit anti-mouse IgG plus protein A-gold); this finding was comparable to that for strain E135 (negative control; deficient for *orf903-906.7*, *orf2351*, and *orf2430*), when E135 cells were incubated with each antiserum separately (Fig. 4G and H).

Enrichment of 22 putative CWAP genes among clinical and outbreak-associated isolates. Although the 22 putative CWAP genes studied are found in *E. faecium* isolates from different ecological niches or genetic backgrounds, they are not equally distributed (Fig. 5A). On average, clinical and outbreak-associated isolates contain 19 and 21 of the 22 identified putative CWAP genes, respectively. Human hospital surveillance isolates harbor on average 17 of the 22 CWAP genes and human community surveillance isolates harbor 14, while animal and environmental isolates contain only 13 and 15 of the 22 putative CWAP genes, respectively. These results suggest that variant numbers or various combinations of putative CWAP genes contribute to survival and growth in different ecological niches. It is interesting in this respect that isolates not belonging to the hospital-adapted genetic complex CC17 contain significantly fewer ($P = 0.02$) of the 22 putative CWAP genes than isolates that do belong to CC17 (Fig. 5B). On average, CC17 isolates contain 21 of the 22 studied CWAP genes, whereas non-CC17 isolates contain on average only 15 of these CWAP genes.

Indication of horizontal gene transfer of putative CWAP genes. An MST based upon clustering on the presence and absence of the 22 putative CWAP genes revealed a grouping of 58 isolates in a distinct CWAP-enriched cluster (Fig. 6A). Isolates belonging to this cluster belong predominantly to CC17 (n = 40/58; 70%) (Fig. 6A) and contain the highest proportion of CC17-enriched CWAP genes (Fig. 6B). However, the CWAP-enriched cluster also contains 18 non-CC17 isolates which are not evolutionary linked to CC17. Fourteen of these 18 are hospital derived. This strongly indicates horizontal transfer of putative CWAP genes among *E. faecium* isolates of distinct genetic backgrounds. The facts that isolates within the CWAP-enriched cluster are predominantly hospital related (55/58; 95%) and that of all hospital isolates 87% (55/63) group in this cluster suggest that the acquisition of CWAP genes increases the survival and spread of *E. faecium* in the hospital environment.

Downloaded from http://jb.asm.org/ on September 30, 2020 by guest

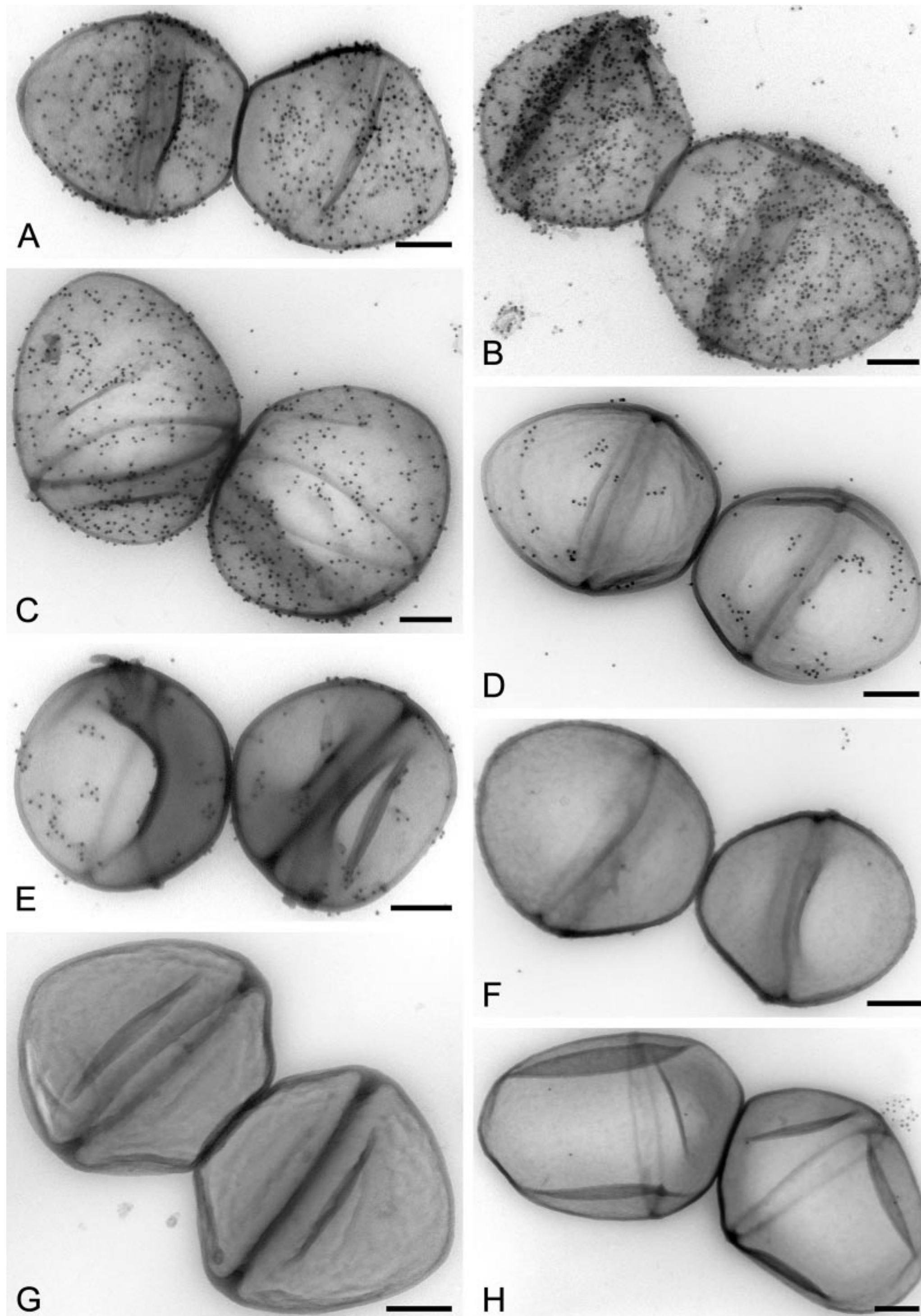

FIG. 4. Transmission electron microscope micrographs. *E. faecium* isolates E135, E380, E470, and E745 were negatively stained and labeled individually with five peptide immune sera directed against the five CC17-enriched LPXTG surface proteins followed by rabbit anti-mouse IgG and protein A-gold (15 nm). (A) E380 *E. faecium* cells (containing *orf904.5* allele 2 and *orf906.7* insertion A) incubated with anti-903 mouse immune serum. (B) E380 cells incubated with anti-904 mouse immune serum. (C) E380 cells incubated with anti-906 mouse immune serum. (D) E470 cells incubated with anti-2351 mouse immune serum. (E) E745 cells incubated with anti-2430 mouse immune serum. (F) E745 cells incubated with mouse preimmune serum. (G) E470 cells incubated with only rabbit anti-mouse IgG and protein A-gold (conjugate control). (H) E135 cells (negative control; deficient for *orf903-906.7*, *orf2351*, and *orf2430*) incubated with anti-903 immune serum (incubations with anti-904, -906, -2351, and -2430 immune sera, which also did not display gold labeling, are not shown). Bar, 200 nm; magnification, ×65,000.

Downloaded from http://jb.asm.org/ on September 30, 2020 by guest

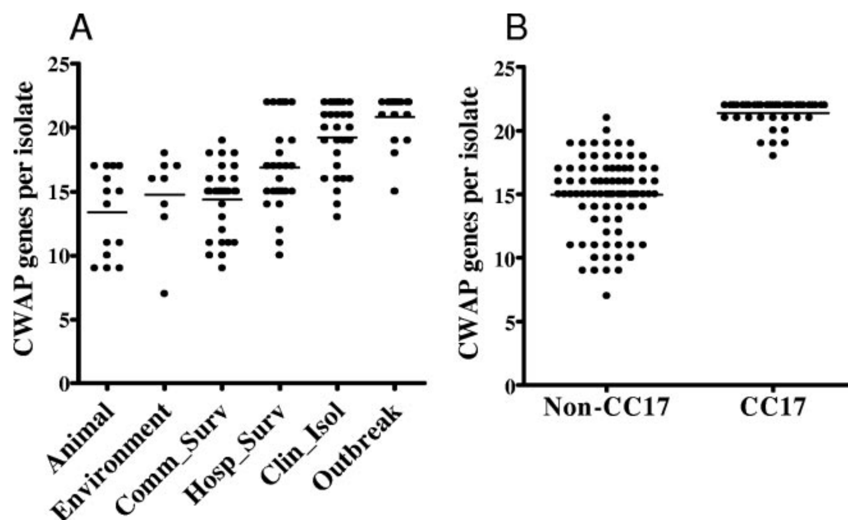

FIG. 5. Distribution of putative CWAP genes among *E. faecium* isolates. Each black circle represents an *E. faecium* isolate with a specific number of putative CWAP genes. Horizontal lines indicate the average number of putative CWAP genes in each epidemiological class of isolates. (A) Distribution among distinct epidemiological classes. The epidemiological classes are ordered with increasing clinical relevance. Annotation: Comm_surv, community surveillance isolates from human volunteers not connected to hospitals; Hosp_surv, hospital surveillance isolates from hospitalized patients not associated with enterococcal infection or outbreak; Clin_Isol, isolates from clinical sites from hospitalized patients; Outbreak, isolates from hospital outbreaks. (B) Distribution of putative CWAP genes among CC17 and non-CC17 *E. faecium* isolates. On average, CC17 isolates contain 21 of the 22 CWAP genes, while non-CC17 isolates contain 15 of the 22 CWAP genes in their genomes.

DISCUSSION

Although knowledge of the adaptive mechanisms of CC17 *E. faecium* has gradually increased, the complete nature of the process of hospital adaptation remains poorly understood. In this study, we show that clinical and outbreak *E. faecium* isolates belonging to CC17 as well as 14 hospital-derived isolates not belonging to CC17 contain on average 19 and 21 of 22 analyzed putative CWAP-encoding genes identified in the incomplete genome sequence of an endocarditis *E. faecium* isolate (TX0016) belonging to CC17. In contrast, non-CC17 isolates e.g., animal, human community, and environmental *E. faecium* strains, contain on average only 13, 14, and 15 of the 22 putative CWAP genes, respectively. The enrichment of putative CWAP genes in clinical and outbreak-associated *E. faecium* isolates was striking. The presence of variant numbers and combinations of CWAP genes by distinct epidemiological class may contribute to successful adaptation to the different ecological niches and thus may provide a selective advantage for these isolates in the hospital setting. Five LPXTG surface protein genes were specifically enriched in CC17 *E. faecium* isolates, and expression in terms of mRNA level and at the surface of the cell was confirmed. These CC17-enriched genes encoding surface-exposed LPXTG proteins do not belong to the *E. faecium* core genome, as they were absent from the majority of the non-CC17 isolates. The ORFs *orf903*, *orf905*, and *orf907* clustered together on the *E. faecium* genome on a novel genomic island covering contig 609. Two lines of evidence support this assumption. First, this region does not harbor housekeeping genes but putatively encodes accessory functions. Second, it is most likely acquired by horizontal gene transfer, as it is found in isolates that are (based upon MLST) not closely evolutionary linked. Furthermore, we hypothesize that this genomic island constitutes a putative pathogenicity island, since it is largely missing from nonclinical non-CC17 isolates. The deduced protein sequences of this region revealed an N-terminal signal sequence in the 903, 904, and 906 proteins and an LPXTG-like motif followed by a hydrophobic domain at the C termini in the 903, 905, and 907 proteins. The presence of a thymine-to-cytosine point mutation in the *orf904-905* junction and a 19-bp insertion in the *orf906-907* junction merging *orf904-905* into *orf904.5* and *orf906-907* into *906.7* was found only in a subset (9.8%) of the 131 isolates. Insertion A, the most frequently found 19-bp insertion, which removed the stop codon in *orf906*, possibly occurred through a site-specific recombinatorial event. The other two insertions, B and C, may have arisen by multiple-nucleotide mutations of insertion A. For these isolates, *orf904.5* and *orf906.7* are predicted to encode a CWAP with an N-terminal signal sequence and a CWS. For *E. faecium* strain E380, this was demonstrated using Western blotting and immunoelectron microscopy. The presence of transcriptional regulators and a sortase downstream of *orf906.7* suggests that the three putative CWAP genes are part of a functional transcriptional unit and that the 904.5 and 906.7 proteins are anchored to the cell wall by a sortase-dependent mechanism. The 903 and 904.5 proteins have similarities to uncharacterized putative CWAPs of *L. monocytogenes* and *E. faecalis*. Furthermore, the 906.7 protein is homologous to the recently discovered Bee3 protein, which plays a role in biofilm formation in *E. faecalis* (35), and therefore we hypothesize that the 903, 904.5, and 906.7 proteins may also play a role in biofilm formation of *E. faecium*. The ability to form biofilm is clinically highly relevant, as the pathogeneses of many bacterial infections have been attributed to this process (27).

The serine-rich repeat region of the 2351 LPXTG surface protein displays high similarity with the serine-rich repeat domain of the cell wall-anchored clumping factor B protein

Downloaded from http://jb.asm.org/ on September 30, 2020 by guest

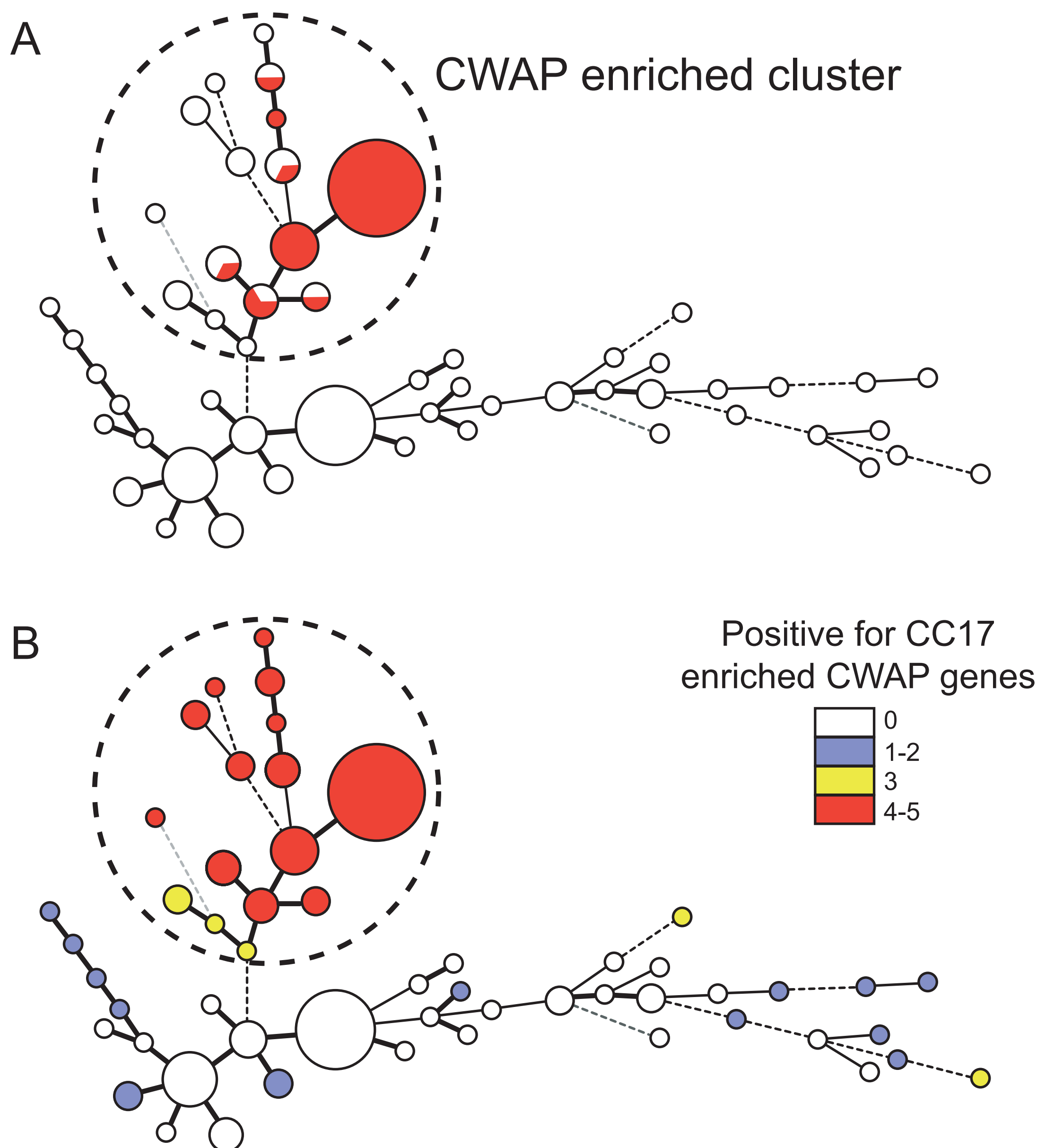

FIG. 6. MST based on clustering on the presence and absence of the 22 putative CWAP genes. A categorical coefficient and the eBURST priority rule of the highest number of single-locus changes were used for the clustering. Circles represent putative CWAP gene profiles, and the sizes of the circles indicate the numbers of isolates. Thick short lines connecting two CWAP gene types denote types differing by a single CWAP gene, thin lines connect double CWAP gene variants, and dotted lines indicate the most likely connection between two types differing by more than two CWAP genes (black) or more than four CWAP genes (grey). (A) All 40 CC17 isolates are marked red and cluster together with 18 non-CC17 clinical isolates (marked white) in a CWAP-enriched cluster. (B) The presence of 0, 1 or 2, 3, or 4 or 5 CC17-enriched CWAP genes is indicated with colors depicted in the legend.

(ClfB) of *S. aureus*. The ClfB protein is an MSCRAMM involved in fibrinogen binding (25). Fibrin and fibrinogen are major components of blood clots and are the major plasma proteins deposited on implanted foreign devices. The ability of *S. aureus* to adhere to fibrinogen/fibrin is of importance in the initiation of intravascular catheter-related infections (8, 20, 41, 42). Thus, the presence of the putative 2351 LPXTG fibrinogen binding protein could be a beneficial factor for CC17 *E. faecium* in establishing foreign body-associated infections.

The 2430 LPXTG surface protein has 94% homology to the EF1896 protein of *E. faecalis* V583, which may suggest that the gene encoding this protein has been acquired via horizontal gene transfer and that this may have occurred recently. This may explain its low prevalence (7%) in non-CC17 isolates and in only a subset (70%) of the CC17 isolates. In both *E. faecium* and *E. faecalis*, the 2430 protein has not been characterized. However, its ortholog in *E. faecalis*, EF1896, was enriched in clinical isolates (17). The exact role of the five CC17-enriched LPXTG surface proteins remains to be elucidated but will be determined by making insertional knockouts in the genes by

Downloaded from http://jb.asm.org/ on September 30, 2020 by guest

use of the improved temperature-sensitive vectors described by Nallapareddy et al. (22).

The sequence homology and surface-exposed expression of the five CC17-enriched gene products suggest that these genes encode factors involved in bacterial adhesion and biofilm formation. Attachment of a bacterium to host tissues or indwelling abiotic medical devices like stents or catheters, whether mediated through CWAPs or not, is the initial event in the pathogenesis of microbial infections (28). Therefore, the five identified CC17-enriched LPXTG surface proteins may play a role in the pathogenesis of CC17 *E. faecium* in hospital-related infections and as such may have contributed to the rapid emergence of CC17 *E. faecium* as a nosocomial pathogen over recent decades. The clear distinction of the CWAP gene profile of *E. faecium* CC17 make the 903, 904.5, 906.7, 2351, and 2430 LPXTG surface proteins possible targets for novel active or passive immunization therapies.

ACKNOWLEDGMENTS

This work was supported by ZonMW research grant 6100.008 from The Netherlands Organization for Health, Research and Development and by a grant from the European Union Sixth Framework Programme under contract LSHE-CT-2007-037410.

We thank Janice Griffith for kindly providing reagents, fruitful discussions, and technical assistance.

REFERENCES

1. **Ausubel, F. M. (ed.)** 1987. Current protocols in molecular biology. John Wiley & Sons, Inc., New York, NY.
2. **Bendtsen, J. D., H. Nielsen, G. von Heijne, and S. Brunak.** 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. **340:**783–795.
3. **Boekhorst, J., M. W. de Been, M. Kleerebezem, and R. J. Siezen.** 2005. Genome-wide detection and analysis of cell wall-bound proteins with LPxTG-like sorting motifs. J. Bacteriol. **187:**4928–4934.
4. **Bonten, M. J., M. K. Hayden, C. Nathan, T. W. Rice, and R. A. Weinstein.** 1998. Stability of vancomycin-resistant enterococcal genotypes isolated from long-term-colonized patients. J. Infect. Dis. **177:**378–382.
5. **Cheung, A. L., K. J. Eberhardt, and V. A. Fischetti.** 1994. A method to isolate RNA from gram-positive bacteria and mycobacteria. Anal. Biochem. **222:**511–514.
6. **Coque, T. M., R. J. Willems, J. Fortun, J. Top, S. Diz, E. Loza, R. Canton, and F. Baquero.** 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.
7. **Gaspar, A. H., and H. Ton-That.** 2006. Assembly of distinct pilus structures on the surface of *Corynebacterium diphtheriae*. J. Bacteriol. **188:**1526–1533.
8. **Herrmann, M., Q. J. Lai, R. M. Albrecht, D. F. Mosher, and R. A. Proctor.** 1993. Adhesion of *Staphylococcus aureus* to surface-bound platelets: role of fibrinogen/fibrin and platelet integrins. J. Infect. Dis. **167:**312–322.
9. **Iwen, P. C., D. M. Kelly, J. Linder, S. H. Hinrichs, E. A. Dominguez, M. E. Rupp, and K. D. Patil.** 1997. Change in prevalence and antibiotic resistance of *Enterococcus* species isolated from blood cultures over an 8-year period. Antimicrob. Agents Chemother. **41:**494–495.
10. **Janulczyk, R., and M. Rasmussen.** 2001. Improved pattern for genome-based screening identifies novel cell wall-attached proteins in gram-positive bacteria. Infect. Immun. **69:**4019–4026.
11. **Jett, B. D., M. M. Huycke, and M. S. Gilmore.** 1994. Virulence of enterococci. Clin. Microbiol. Rev. **7:**462–478.
12. **Joh, D., E. R. Wann, B. Kreikemeyer, P. Speziale, and M. Hook.** 1999. Role of fibronectin-binding MSCRAMMs in bacterial adherence and entry into mammalian cells. Matrix Biol. **18:**211–223.
13. **Johnson, A. P.** 1994. The pathogenicity of enterococci. J. Antimicrob. Chemother. **33:**1083–1089.
14. **Leavis, H. L., M. J. Bonten, and R. J. Willems.** 2006. Identification of high-risk enterococcal clonal complexes: global dispersion and antibiotic resistance. Curr. Opin. Microbiol. **9:**454–460.
15. **Leavis, H. L., R. J. Willems, J. Top, and M. J. Bonten.** 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.
16. **Leavis, H. L., R. J. Willems, W. J. Van Wamel, F. H. Schuren, M. P. Caspers, and M. J. Bonten.** 2007. Insertion sequence-driven diversification creates a globally dispersed emerging multiresistant subspecies of *E. faecium*. PLoS Pathogens **3:**e7.
17. **Lepage, E., S. Brinster, C. Caron, C. Ducroix-Crepy, L. Rigottier-Gois, G. Dunny, C. Hennequet-Antier, and P. Serror.** 2006. Comparative genomic hybridization analysis of *Enterococcus faecalis*: identification of genes absent from food strains. J. Bacteriol. **188:**6858–6868.
18. **Mazmanian, S. K., G. Liu, H. Ton-That, and O. Schneewind.** 1999. *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. Science **285:**760–763.
19. **Mazmanian, S. K., H. Ton-That, and O. Schneewind.** 2001. Sortase-catalysed anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. Mol. Microbiol. **40:**1049–1057.
20. **Moreillon, P., J. M. Entenza, P. Francioli, D. McDevitt, T. J. Foster, P. Francois, and P. Vaudaux.** 1995. Role of *Staphylococcus aureus* coagulase and clumping factor in pathogenesis of experimental endocarditis. Infect. Immun. **63:**4738–4743.
21. **Nallapareddy, S. R., and B. E. Murray.** 2006. Ligand-signaled upregulation of *Enterococcus faecalis ace* transcription, a mechanism for modulating host-*E. faecalis* interaction. Infect. Immun. **74:**4982–4989.
22. **Nallapareddy, S. R., K. V. Singh, and B. E. Murray.** 2006. Construction of improved temperature-sensitive and mobilizable vectors and their use for constructing mutations in the adhesin-encoding *acm* gene of poorly transformable clinical *Enterococcus faecium* strains. Appl. Environ. Microbiol. **72:**334–345.
23. **Nallapareddy, S. R., K. V. Singh, J. Sillanpaa, D. A. Garsin, M. Hook, S. L. Erlandsen, and B. E. Murray.** 2006. Endocarditis and biofilm-associated pili of *Enterococcus faecalis*. J. Clin. Investig. **116:**2799–2807.
24. **Nallapareddy, S. R., G. M. Weinstock, and B. E. Murray.** 2003. Clinical isolates of *Enterococcus faecium* exhibit strain-specific collagen binding mediated by Acm, a new member of the MSCRAMM family. Mol. Microbiol. **47:**1733–1747.
25. **Ní Eidhin, D., S. Perkins, P. Francois, P. Vaudaux, M. Hook, and T. J. Foster.** 1998. Clumping factor B (ClfB), a new surface-located fibrinogen-binding adhesin of *Staphylococcus aureus*. Mol. Microbiol. **30:**245–257.
26. **Ozawa, Y., P. Courvalin, and M. Gaiimand.** 2000. Identification of enterococci at the species level by sequencing of the genes for D-alanine:D-alanine ligases. Syst. Appl. Microbiol. **23:**230–237.
27. **Parsek, M. R., and P. K. Singh.** 2003. Bacterial biofilms: an emerging link to disease pathogenesis. Annu. Rev. Microbiol. **57:**677–701.
28. **Patti, J. M., B. L. Allen, M. J. McGavin, and M. Hook.** 1994. MSCRAMM-mediated adherence of microorganisms to host tissues. Annu. Rev. Microbiol. **48:**585–617.
29. **Rice, L. B., L. Carias, S. Rudin, C. Vael, H. Goossens, C. Konstabel, I. Klare, S. R. Nallapareddy, W. Huang, and B. E. Murray.** 2003. A potential virulence gene, hyl*Efm*, predominates in *Enterococcus faecium* of clinical origin. J. Infect. Dis. **187:**508–512.
30. **Sahm, D. F., J. Kissinger, M. S. Gilmore, P. R. Murray, R. Mulder, J. Solliday, and B. Clarke.** 1989. In vitro susceptibility studies of vancomycin-resistant *Enterococcus faecalis*. Antimicrob. Agents Chemother. **33:**1588–1591.
31. **Schneewind, O., D. Mihaylova-Petkov, and P. Model.** 1993. Cell wall sorting signals in surface proteins of gram-positive bacteria. EMBO J. **12:**4803–4811.
32. **Schouls, L. M., H. G. van der Heide, L. Vauterin, P. Vauterin, and F. R. Mooi.** 2004. Multiple-locus variable-number tandem repeat analysis of Dutch *Bordetella pertussis* strains reveals rapid genetic changes with clonal expansion during the late 1990s. J. Bacteriol. **186:**5496–5505.
33. **Smeltzer, M. S., and A. F. Gillaspy.** 2000. Molecular pathogenesis of staphylcoccal [sic] osteomyelitis. Poult. Sci. **79:**1042–1049.
34. **Snodgrass, J. L., N. Mohamed, J. M. Ross, S. Sau, C. Y. Lee, and M. S. Smeltzer.** 1999. Functional analysis of the *Staphylococcus aureus* collagen adhesin B domain. Infect. Immun. **67:**3952–3959.
35. **Tendolkar, P. M., A. S. Baghdayan, and N. Shankar.** 2006. Putative surface proteins encoded within a novel transferable locus confer a high-biofilm phenotype to *Enterococcus faecalis*. J. Bacteriol. **188:**2063–2072.
36. **Timmers, G. J., W. C. van der Zwet, I. M. Simoons-Smit, P. H. Savelkoul, H. H. Meester, C. M. Vandenbroucke-Grauls, and P. C. Huijgens.** 2002. Outbreak of vancomycin-resistant *Enterococcus faecium* in a haematology unit: risk factor assessment and successful control of the epidemic. Br. J. Haematol. **116:**826–833.
37. **Toledo-Arana, A., J. Valle, C. Solano, M. J. Arrizubieta, C. Cucarella, M. Lamata, B. Amorena, J. Leiva, J. R. Penades, and I. Lasa.** 2001. The enterococcal surface protein, Esp, is involved in *Enterococcus faecalis* biofilm formation. Appl. Environ. Microbiol. **67:**4538–4545.
38. **Top, J., L. M. Schouls, M. J. Bonten, and R. J. Willems.** 2004. Multiple-locus variable-number tandem repeat analysis, a novel typing scheme to study the genetic relatedness and epidemiology of *Enterococcus faecium* isolates. J. Clin. Microbiol. **42:**4503–4511.
39. **van den Bogaard, A. E., P. Mertens, N. H. London, and E. E. Stobberingh.** 1997. High prevalence of colonization with vancomycin- and pristinamycin-resistant enterococci in healthy humans and pigs in The Netherlands: is the addition of antibiotics to animal feeds to blame? J. Antimicrob. Chemother. **40:**454–456.
40. **Van Wamel, W. J., A. P. Hendrickx, M. J. Bonten, J. Top, G. Posthuma, and R. J. Willems.** 2007. Growth condition-dependent Esp expression of *Entero-*

Downloaded from http://jb.asm.org/ on September 30, 2020 by guest

coccus faecium affects initial adherence and biofilm formation. Infect. Immun. **75:**924–931.
41. **Vaudaux, P., D. Pittet, A. Haeberli, E. Huggler, U. E. Nydegger, D. P. Lew, and F. A. Waldvogel.** 1989. Host factors selectively increase staphylococcal adherence on inserted catheters: a role for fibronectin and fibrinogen or fibrin. J. Infect. Dis. **160:**865–875.
42. **Vaudaux, P. E., P. Francois, R. A. Proctor, D. McDevitt, T. J. Foster, R. M. Albrecht, D. P. Lew, H. Wabers, and S. L. Cooper.** 1995. Use of adhesion-defective mutants of *Staphylococcus aureus* to define the role of specific plasma proteins in promoting bacterial adhesion to canine arteriovenous shunts. Infect. Immun. **63:**585–590.
43. **Walsh, E. J., L. M. O'Brien, X. Liang, M. Hook, and T. J. Foster.** 2004. Clumping factor B, a fibrinogen-binding MSCRAMM (microbial surface components recognizing adhesive matrix molecules) adhesin of *Staphylococcus aureus*, also binds to the tail region of type I cytokeratin 10. J. Biol. Chem. **279:**50691–50699.
44. **Wann, E. R., S. Gurusiddappa, and M. Hook.** 2000. The fibronectin-binding MSCRAMM FnbpA of *Staphylococcus aureus* is a bifunctional protein that also binds to fibrinogen. J. Biol. Chem. **275:**13863–13871.
45. **Willems, R. J., W. Homan, J. Top, M. van Santen-Verheuvel, D. Tribe, X. Manzioros, C. Gaillard, C. M. Vandenbroucke-Grauls, E. M. Mascini, E. van Kregten, J. D. van Embden, and M. J. Bonten.** 2001. Variant *esp* gene as a marker of a distinct genetic lineage of vancomycin-resistant *Enterococcus faecium* spreading in hospitals. Lancet **357:**853–855.
46. **Willems, R. J., J. Top, M. van Santen, D. A. Robinson, T. M. Coque, F. Baquero, H. Grundmann, and M. J. Bonten.** 2005. Global spread of vancomycin-resistant *Enterococcus faecium* from distinct nosocomial genetic complex. Emerg. Infect. Dis. **11:**821–828.

Downloaded from http://jb.asm.org/ on September 30, 2020 by guest