Applied and Environmental Microbiology, Nov. 2008, p. 7094–7097 Vol. 74, No. 22
0099-2240/08/$08.00+0 doi:10.1128/AEM.01378-08

Identification of a Novel Genomic Island Specific to Hospital-Acquired Clonal Complex 17 *Enterococcus faecium* Isolates▽†

Esther Heikens,* Willem van Schaik, Helen L. Leavis, Marc J. M. Bonten, and Rob J. L. Willems

Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, The Netherlands

Received 19 June 2008/Accepted 21 September 2008

Hospital-acquired clonal complex 17 (CC17) *Enterococcus faecium* strains are genetically distinct from indigenous strains and are enriched with resistance genes and virulence genes. We identified a genomic island in CC17 *E. faecium* tentatively encoding a metabolic pathway involved in carbohydrate transport and metabolism, which may provide a competitive advantage over the indigenous *E. faecium* microbiota.

Enterococcus faecium strains acquired during hospitalization and responsible for the majority of the hospital burden are genetically distinct from indigenous *E. faecium* strains that belong to the normal intestinal flora (11, 19). The underlying mechanisms explaining the ecological dominance of these hospital-acquired *E. faecium* strains, currently labeled clonal complex 17 (CC17) based on multilocus sequence typing, are not well understood. Previous studies suggest that the acquisition of antibiotic resistance traits, such as penicillin and quinolone resistance, and cell surface proteins may have contributed to the ecological success of CC17 *E. faecium* (6, 7, 10, 12, 17–19). In addition, comparative genomic hybridizations using a mixed whole-genome microarray identified a specific *E. faecium* clade encompassing CC17 *E. faecium* strains containing more than 100 clade-specific genes, including resistance genes, putative virulence genes, and insertion sequence elements (13).

An additional method to identify CC17 *E. faecium*-acquired genes or gene clusters (GeCs) in a genome is base composition analysis. At the time of transfer, horizontally acquired genes often differ in their codon usages, GC percentages, and dinucleotide frequencies, since horizontally acquired genes share these characteristics with the DNA of the bacterium from which they originated (8, 9). A recently described Web-based tool for the detection of horizontally transferred genes and GeCs is the δρ-Web model (16). The δρ-Web model allows whole-genome composition analysis to visualize anomalous DNA in a prokaryotic genome based on differences in both GC percentages and dinucleotide frequencies. Horizontally acquired genes or GeCs, such as genomic islands (GIs), often encode accessory functions, such as additional metabolic activities and antibiotic resistance, or functions involved in microbial fitness, symbiosis, or pathogenesis (1, 4). In this study, we used the δρ-Web model as an initial, quick screen to identify anomalous GeCs in the genome of *E. faecium* DO, a CC17 *E. faecium* strain that may have contributed to increased fitness and enhanced survival of CC17 *E. faecium*. In addition, PCR and dot blotting were performed on a large set of *E. faecium* isolates to confirm whether these anomalous GeCs were CC17 specific.

Identification of a CC17 *E. faecium*-specific GeC. In order to submit the publicly available *E. faecium* DO draft genome sequence (GenBank accession no. AAAK00000000) to the δρ-Web model, an in silico concatenated genome sequence was created by linking all contigs larger than 2,000 bp ($n = 41$), encompassing 64% of the whole genome, in order from large to small. After submission, the concatenated genome sequence was divided into nonoverlapping fragments of 10,000 bp, as recommended by the user guidelines supplied at http://deltarho.amc.uva.nl. The difference in dinucleotide frequency (δ* value) between each fragment and the complete sequence and the GC percentage of each fragment were calculated. The model identified five fragments with both a high δ* value and an aberrant GC percentage compared to the average genome values for the concatenated *E. faecium* DO contigs. These fragments represented sequences located in contigs 608, 624, 638, 654, and 656. Contig 656 was previously identified as a contig harboring many CC17-specific genes (13). Therefore, we chose to focus on the fragments located in the other four contigs. For each fragment, one gene (*orf877*, *orf1155*, *orf1482*, and *orf2303* of contigs 608, 624, 638, and 654, respectively) was chosen and the presence of this gene was assessed by PCR and dot blotting on chromosomal DNA from 134 *E. faecium* isolates, 41 CC17 *E. faecium* isolates, and 93 non-CC17 *E. faecium* isolates (see Table S1 in the supplemental material). The preparation of chromosomal DNA and dot blotting were performed as described previously (7). The primers used for PCR and for the generation of DNA probes are listed in Table S2 in the supplemental material. *E. faecium* DO was used as a positive control and *E. faecalis* V583 as a negative control (15). PCR (data not shown) and dot blotting revealed that one of the four genes, *orf1482* on contig 638, was specific to CC17 *E. faecium* (Table 1). This gene was detected in 97.56% (40/41) of the CC17 *E. faecium* isolates and in only 4.30% (4/93) of the non-CC17 *E. faecium* isolates ($P < 0.0001$; Fisher's exact test). *orf1482* encodes a putative transcriptional regulator belonging to the AraC family. Transcriptional regulators of the AraC family are widely spread among bacteria and regulate genes with diverse functions, ranging from carbon metabolism and stress response to pathogenesis (3, 14). Since AraC-type tran-

* Corresponding author. Mailing address: Department of Medical Microbiology, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands. Phone: 31 88 7555006. Fax: 31 30 2541770. E-mail: E.heikens@umcutrecht.nl.

† Supplemental material for this article may be found at http://aem.asm.org/.

▽ Published ahead of print on 3 October 2008.

TABLE 1. Prevalence of genes in CC17 and non-CC17 *E. faecium* isolates as determined by dot blotting

Gene	No. (%) of CC17 isolates harboring gene	No. (%) of non-CC17 isolates harboring gene
orf1474	41 (100)	93 (100)
orf1475	40 (97.56)	4 (4.30)
orf1476	40 (97.56)	4 (4.30)
orf1477	40 (97.56)	4 (4.30)
orf1478	40 (97.56)	4 (4.30)
orf1479	40 (97.56)	4 (4.30)
orf1480	40 (97.56)	4 (4.30)
orf1481	40 (97.56)	4 (4.30)
orf1482	40 (97.56)	4 (4.30)
orf1483	41 (100)	93 (100)

scriptional regulators are often found close to or on GIs (5), the presence or absence of genes located upstream and downstream of *orf1482* was determined. This revealed that all isolates that contained *orf1482* also contained a set of seven genes just upstream of *orf1482*, while isolates lacking *orf1482* also lacked this set of genes, indicating that the *araC*-like gene is located in an 8.5-kb GeC, which is specifically enriched in CC17 *E. faecium* (Table 1). This means that the genes of the GeC may also serve as a marker to distinguish CC17 *E. faecium* strains from other *E. faecium* strains. *orf1474* and *orf1483*, flanking this 8.5-kb GeC, belong to the *E. faecium* core genome. The four non-CC17 *E. faecium* isolates that harbor this GeC represent two hospital outbreak isolates (E300 and E1679) and two clinical isolates (E1172 and E1721). The single CC17 *E. faecium* isolate that does not harbor this GeC represents a clinical isolate (E1263).

Organization and genetic features of the CC17 *E. faecium*-specific GeC. Direct and inverted repeats were found upstream and downstream of the CC17 *E. faecium*-specific GeC (Fig. 1). Furthermore, a putative integrase is located downstream of *orf1482*. The presence of a member of the AraC family of transcriptional regulators, of direct and inverted repeats upstream and downstream of this GeC, and of an integrase downstream and the finding that this cluster of genes was found in clinically relevant isolates and not in surveillance isolates suggest that this region encompasses a distinct GI that is acquired by horizontal transfer.

BLAST searches in GenBank of the predicted proteins encoded by this GI revealed that the GI genes putatively encode two glycosyl hydrolases, two binding protein-dependent transporter proteins, a sugar binding protein, two proteins with unknown functions, and an AraC transcriptional regulator (Table 2). Considering the putative functions of the predicted GI proteins, this GI may represent a novel metabolic island involved in carbohydrate transport and metabolism, possibly providing CC17 *E. faecium* a competitive advantage over indigenous commensal *E. faecium*, in particular ecological niches. The CC17 *E. faecium*-specific GI proteins share only low-level identity with orthologs in other prokaryotes, and these orthologous proteins originate from a wide range of taxonomically distinct groups. This indicates that this GI originates from an as-yet-unidentified biological source.

TABLE 2. Identities of the predicted proteins encoded by the GI specifically enriched in CC17 *E. faecium* as determined by BLAST

Predicted protein	Annotation	Organism	Amino acid identity (%)[a]
1475	Glycosyl hydrolase	*Citrobacter koseri*	52
1476	Protein with unknown function	*Paenibacillus* species	37
1477	Glycosyl hydrolase	*Paenibacillus* species	50
1478	Binding protein-dependent transporter protein	*Halothermothrix orenii*	36
1479	Binding protein-dependent transporter protein	*Petrotoga mobilis*	43
1480	No homologues found		
1481	Sugar binding protein	*Bacillus clausii*	22
1482	AraC transcriptional regulator	*Clostridium bolteae*	36

[a] Amino acid identities represent top BLAST hits. BLAST searches were performed in GenBank.

Transcriptional analysis of the GI. To assess whether this GI is actively expressed, total RNA was isolated from mid-exponential cultures of five CC17 *E. faecium* isolates (E470, E734, E1133, E1162, and DO) by using Tri reagent (Ambion, Austin, TX). Residual DNA was removed with Turbo DNase according to the protocol supplied with the Turbo DNA-free kit (Ambion). cDNA was synthesized with a SuperScript III first-strand synthesis system (Invitrogen Corp., Carlsbad, CA), using random hexamers according to the manufacturer's instructions. The expression levels of the eight GI genes were then assessed by PCR with gene-specific primers (see Table S2 in the supplemental material). Expression was detected at the mRNA level for all the GI genes (data not shown), indicating that these genes are expressed and do not represent silent

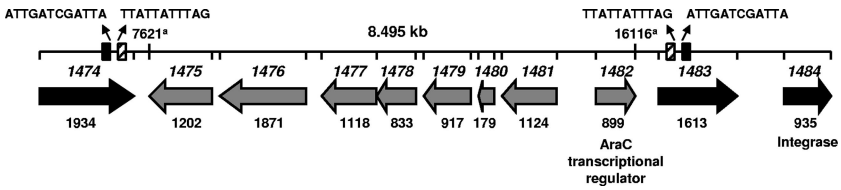

FIG. 1. Genomic organization of the 8.5-kb GI (*E. faecium* DO contig 638) specifically enriched in CC17 *E. faecium*. The direction of transcription is indicated by arrows. The gray arrows represent the genes which belong to the GI, and the black arrows represent the flanking genes. The numbers below the arrows indicate gene sizes. Direct (dashed boxes) and inverted (black boxes) repeats were found at positions 7184 and 16391 and positions 7061 and 16803, respectively. Open reading frame numbers are indicated in italics. a, nucleotide reference position relative to that of the *E. faecium* DO contig 638 sequence (GenBank accession no. AAAK03000019).

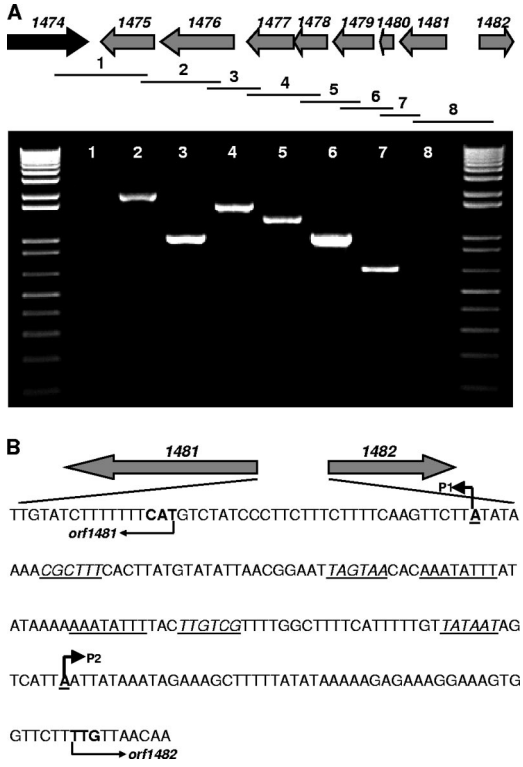

FIG. 2. Transcriptional analysis of the GI specifically enriched in CC17 *E. faecium*. (A) Cotranscription of *orf1475* to *orf1481*, demonstrated by using primer pairs designed to span the entire region, resulting in overlapping amplification products. The molecular size marker is the 1-kb ladder (Invitrogen Corp.). (B) Intergenic region of *orf1481-orf1482*, with the start codons and orientations of *orf1481* and *orf1482* in bold and indicated by arrows below the sequence. Transcriptional start sites and directions are in bold, underlined and indicated by arrows above the sequence. Putative −35 and −10 boxes are underlined and in italics. Direct repeats, representing a putative binding site of a transcriptional regulator protein, are between the two promoters (P1 and P2) and are underlined.

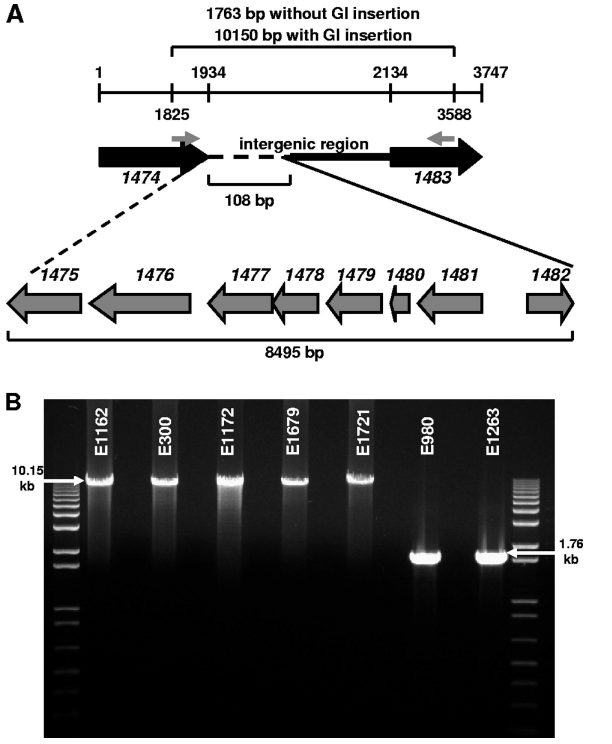

FIG. 3. Analysis of the insertion site of the GI specifically enriched in CC17 *E. faecium*. (A) Schematic representation of insertion of the GI in the intergenic region from *orf1474* to *orf1483*, resulting in deletion of a 108-bp fragment (dashed line). The primers used to analyze the insertion site are indicated by small gray arrows. Numbers indicate the start and stop positions of *orf1474* and *orf1483* and the start position of the primers. The sizes of the two amplicons, with and without GI insertion, the GI, and the deleted fragment are indicated. Open reading frame numbers are indicated in italics. (B) PCR was performed with one CC17 *E. faecium* isolate (E1162), four non-CC17 *E. faecium* isolates that harbor the gene cluster (E300, E1172, E1679, and E1721), one non-CC17 *E. faecium* isolate (E980), and the single CC17 *E. faecium* isolate that does not harbor the GI (E1263). The molecular size marker is the 1-kb ladder (Invitrogen Corp.).

genes. The secondary structures of the mRNA from the region between *orf1474* and *orf1475* and that from the region downstream of *orf1482* were predicted using Mfold (20), which revealed very stable stem-loop structures with highly negative ΔG values of −8.11 kcal/mol and −14.01 kcal/mol, respectively. This suggests the presence of transcriptional terminators at these sites and that *orf1475* to *orf1481* are part of a single operon. To confirm this, reverse transcription-PCR was performed with cDNA by using gene-specific primer pairs (see Table S2 in the supplemental material) designed to span the entire region, resulting in overlapping amplification products. Products of the expected size were observed with primer pairs covering *orf1475* to *orf1481*, showing that these genes are cotranscribed in a single operon and that *orf1474* and *orf1482* are not part of the operon (Fig. 2A). In addition, promoter mapping of *orf1481* and *orf1482* was performed using 5′ rapid amplification of cDNA ends (Invitrogen Corp.) according to the manufacturer's instructions. Total RNA from *E. faecium* DO was isolated and reverse transcribed using primers 1481R and 1482R (see Table S2 in the supplemental material). The subsequent PCR was performed using primers 1481R2 and 1482R2 and the abridged anchor primer provided with the system. Sequencing of the PCR products revealed that two transcriptional start sites were located in the *orf1481-orf1482* intergenic region (Fig. 2B). Direct repeats were found between the two promoters (P1 and P2), which may represent a putative binding site of a transcriptional regulator protein (2).

GI insertion site. To analyze the insertion sites of GIs in other CC17 *E. faecium* and non-CC17 *E. faecium* isolates, PCR and DNA sequencing were performed using primer pair 1474F2-1483R (see Table S2 in the supplemental material), designed from the flanking genes in *E. faecium* DO, *orf1474* and *orf1483*. The insertion site was analyzed for one CC17 *E. faecium* isolate (E1162), one non-CC17 *E. faecium* isolate (E980), the single CC17 *E. faecium* isolate that does not harbor the gene cluster (E1263), and the four non-CC17 *E. faecium* isolates that harbor the gene cluster (E300, E1172, E1679, and E1721). In E1162, E300, E1172, E1679, and E1721, the GI was found to be inserted at exactly the same position as in *E. faecium* DO. In E980 and E1263, *orf1474* and *orf1483*, the flanking genes of the GI, are located directly adjacent to each

other. The insertion of the GI resulted in the deletion of a 108-bp fragment located in the intergenic region from *orf1474* to *orf1483* (Fig. 3A and B). The observation of an identical insertion site in isolates that carry this GI suggests site-specific recombination.

Conclusions. By using the δρ-Web model, PCR, and dot blotting, we identified a GI tentatively encoding a novel metabolic pathway involved in carbohydrate transport and metabolism. Our finding that all CC17 *E. faecium* isolates but one harbor this island and that none of the non-CC17 *E. faecium* human surveillance, environmental, and animal isolates harbors it indicates that this GI is acquired by CC17 *E. faecium* via horizontal transfer. We hypothesize that this GI may provide CC17 *E. faecium* a competitive advantage over the indigenous commensal *E. faecium* flora by enabling CC17 *E. faecium* to effectively colonize the gastrointestinal tracts of hospitalized patients.

This work was supported by the European Union Sixth Framework Programme "Approaches to control multiresistant enterococci: studies on molecular ecology, horizontal gene transfer, fitness and prevention" (ACE) under contract LSHE-CT-2007-037410 and ZonMW "Vaccine-development to combat the emergence of vancomycin-resistant *Enterococcus faecium*" project number 0.6100.0008.

REFERENCES

1. **Dobrindt, U., B. Hochhut, U. Hentschel, and J. Hacker.** 2004. Genomic islands in pathogenic and environmental microorganisms. Nat. Rev. Microbiol. **2:**414–424.
2. **Gaballa, A., and J. D. Helmann.** 2007. Substrate induction of siderophore transport in *Bacillus subtilis* mediated by a novel one-component regulator. Mol. Microbiol. **66:**164–173.
3. **Gallegos, M. T., R. Schleif, A. Bairoch, K. Hofmann, and J. L. Ramos.** 1997. AraC/XylS family of transcriptional regulators. Microbiol. Mol. Biol. Rev. **61:**393–410.
4. **Hacker, J., and E. Carniel.** 2001. Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. EMBO Rep. **2:**376–381.
5. **Hacker, J., and J. B. Kaper.** 2000. Pathogenicity islands and the evolution of microbes. Annu. Rev. Microbiol. **54:**641–679.
6. **Heikens, E., M. J. Bonten, and R. J. Willems.** 2007. Enterococcal surface protein Esp is important for biofilm formation of *Enterococcus faecium* E1162. J. Bacteriol. **189:**8233–8240.
7. **Hendrickx, A. P., W. J. Van Wamel, G. Posthuma, M. J. Bonten, and R. J. Willems.** 2007. Five genes encoding surface-exposed LPXTG proteins are enriched in hospital-adapted *Enterococcus faecium* clonal complex 17 isolates. J. Bacteriol. **189:**8321–8332.
8. **Karlin, S.** 2001. Detecting anomalous gene clusters and pathogenicity islands in diverse bacterial genomes. Trends Microbiol. **9:**335–343.
9. **Karlin, S., and C. Burge.** 1995. Dinucleotide relative abundance extremes: a genomic signature. Trends Genet. **11:**283–290.
10. **Leavis, H. L., J. Top, N. Shankar, K. Borgen, M. J. Bonten, J. D. van Embden, and R. J. L. Willems.** 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.
11. **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.
12. **Leavis, H. L., R. J. L. 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.
13. **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 *Enterococcus faecium*. PLoS. Pathog. **3:**e7.
14. **Martin, R. G., and J. L. Rosner.** 2001. The AraC transcriptional activators. Curr. Opin. Microbiol. **4:**132–137.
15. **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.
16. **van Passel, M. W., A. C. M. Luyf, A. H. C. van Kampen, A. Bart, and A. van der Ende.** 2005. δρ-Web, an online tool to assess composition similarity of individual nucleic acid sequences. Bioinformatics **21:**3053–3055.
17. **Van Wamel, W. J. B., A. P. A. Hendrickx, M. J. M. Bonten, J. Top, G. Posthuma, and R. J. L. Willems.** 2007. Growth condition-dependent Esp expression by *Enterococcus faecium* affects initial adherence and biofilm formation. Infect. Immun. **75:**924–931.
18. **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.
19. **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.
20. **Zuker, M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31:**3406–3415.