ccrAB$_{Ent}$ serine recombinase genes are widely distributed in the *Enterococcus faecium* and *Enterococcus casseliflavus* species-groups and expressed in *E. faecium*

Eva Katrin Bjørkeng,[1] Girum Tadesse Tessema,[1†] Eirik Wasmuth Lundblad,[2] Patrick Butaye,[3,4] Rob Willems,[5] Johanna Ericsson Sollid,[1] Arnfinn Sundsfjord[1,2] and Kristin Hegstad[1,2*]

[1]Research group for Host-Microbe Interactions, Department of Medical Biology, Faculty of Health Sciences, University of Tromsø, N-9037 Tromsø, Norway, [2]Reference Centre for Detection of Antimicrobial Resistance (K-res), Department of Microbiology and Infection Control, University Hospital of North-Norway, N-9038 Tromsø, Norway, [3]CODA-CERVA-VAR, Brussels, Belgium, [4]Department of Pathology, Bacteriology and Poultry Diseases, Faculty of Veterinary Medicine, Ghent University, Salisburylaan 133, 9820 Merelbeke, Belgium, and [5]Department of Medical Microbiology, University Medical Centre Utrecht, Heidelberglaan 100, Rm G04.614, 3584 CX, Utrecht, The Netherlands.

[†]Present addresses: Nofima Mat AS, Osloveien 1, N-1430 Ås, and Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway.

[*]Corresponding author: Kristin Hegstad

Phone: +47 77 64 63 51

Fax: +47 77 64 53 50

E-mail: Kristin.Hegstad@uit.no

Contents category: Genes and Genomes

Running title: *ccrAB*$_{Ent}$ genes in enterococci

Abbreviations: American Type Culture Collection (ATCC), coding sequences (CDSs), Joint Genome Institute (JGI), Mueller Hinton (MH), Multi Locus Sequence Type (MLST), sequence type (ST), staphylococcal cassette chromosome (SCC)

ABSTRACT

The presence, distribution and expression of cassette chromosome recombinase (*ccr*) genes, homologous to the staphylococcal *ccrAB* genes designated $ccrAB_{Ent}$ genes, were examined in enterococcal isolates (n=421) representing 13 different species. A total of 118 (28 %) isolates were positive for $ccrAB_{Ent}$ genes by PCR, and a number of these were confirmed by Southern hybridization with $ccrA_{Ent}$ (n=76) and DNA sequencing of $ccrA_{Ent}$ and $ccrB_{Ent}$ (n=38). $ccrAB_{Ent}$ genes were present in *Enterococcus faecium* (n=58/216, 27 %), *Enterococcus durans* (n=31/38, 82%), *Enterococcus hirae* (n=27/52, 50 %), *Enterococcus casseliflavus* (n=1/4, 25 %), and *Enterococcus gallinarum* (n=1/2, 50 %). In the eight other species tested including *Enterococcus faecalis* (n=94) $ccrAB_{Ent}$ genes were not found. Thirty-eight sequenced $ccrAB_{Ent}$ genes from five different enterococcal species showed 94-100 % nucleotide sequence identity and linkage PCRs showed heterogeneity in the $ccrAB_{Ent}$ flanking chromosomal genes. Expression analysis of $ccrAB_{Ent}$ genes from the *E. faecium* DO strain showed constitutive expression as a bicistronic mRNA. The $ccrAB_{Ent}$ mRNA levels were lower during log- than stationary-phase in relation to total mRNA. MLST was performed on 39 isolates. $ccrAB_{Ent}$ genes were detected in both hospital related (n=10/29, 34 %) and non-hospital (n=4/10, 40 %) strains of *E. faecium*. Various sequence types were represented by both $ccrAB_{Ent}$ positive and negative isolates suggesting acquisition or loss of $ccrAB_{Ent}$ in *E. faecium*. In summary, $ccrAB_{Ent}$ genes, potentially involved in genome plasticity, are expressed in *E. faecium* and widely distributed in the *E. faecium* and *E. casseliflavus* species groups.

INTRODUCTION

The emergence of multidrug resistant hospital acquired *Enterococcus faecium* as one of the most important pathogens in the developed world has been a remarkable development in the last two decades (Leavis *et al.*, 2006; Werner *et al.*, 2003). Molecular epidemiological studies and comparative genomic hybridization analyses of *E. faecium* (Leavis *et al.*, 2007; Werner *et al.*, 2003) have revealed genotypic differences between hospital and community isolates (Leavis *et al.*, 2006). Mixed whole genome arrays demonstrated a distinct genetic make-up of hospital-associated *E. faecium* with more than 100 extra genes, possibly acquired by horizontal gene transfer (Leavis *et al.*, 2007). The *esp* virulence gene, located on a putative pathogenicity island, is one of the determinants acquired by hospital-associated *E. faecium*. These observations as well as current Multi Locus Sequence Type (MLST) data strongly indicate that gene flux and recombination contribute significantly to diversification and adaptation of *E. faecium* (Leavis *et al.*, 2006; van Schaik *et al.*, 2010).

Recombinases facilitate the exchange of DNA fragments within and between bacteria and are thus pivotal in genome plasticity. Staphylococcal cassette chromosome (SCC) elements are vehicles for exchange of genetic information in staphylococci. These elements are characterized by the presence of terminal inverted repeats, unique recombinase genes, and are flanked by direct repeats (Ito *et al.*, 2004; Ito *et al.*, 2001; Katayama *et al.*, 2003). So far, the major group of elements described are SCC*mec* I-VIII (International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements (IWG-SCC), 2009) responsible for the spread of methicillin resistance between staphylococci. The movement of SCC elements is dependent on the gene products of the cassette chromosome recombinase genes (*ccr*), either the *ccrA-ccrB* complex or the single product of *ccrC* (Katayama *et al.*, 2000; Noto & Archer, 2006). These proteins are serine recombinases of the

resolvase/invertase family which integrate the SCC element in a site specific manner (Ito *et al*., 1999). To our knowledge, *ccr* genes have only been reported in staphylococcal species.

Here, we report for the first time the presence of *ccrAB* genes in enterococci herby designated $ccrAB_{Ent}$ and show that they are expressed under standard *in vitro* growth conditions. Our analyses show that the $ccrAB_{Ent}$ genes are widely distributed in *Enterococcus* species belonging to the *E. faecium* and *E. casseliflavus* species groups.

MATERIALS AND METHODS

Bacterial isolates. A total of 421 *Enterococcus* isolates of 13 species from three continents (Europe, USA and Australia) were included in the study; *E. faecium* (n=216), *E. faecalis* (n=94), *E. durans* (n=38), *E. hirae* (n=52), *E. casseliflavus* (n=4), *E. avium* (n=4), *E. raffinosus* (n=3), *E. canintesti* (n=2), *E. canis* (n=2), *E. gallinarum* (n=2), *E. cecorum* (n=2), *E. asini* (n=1), and *E. dispar* (n=1). Among the 216 *E. faecium* isolates, 72 were of human origin of which 58 were clinical isolates. Among the 94 *E. faecalis* isolates, 13 were of human origin of which eight were clinical isolates. Other enterococcal species included were exclusively of animal origin (poultry, dog, bovine and pig). Six American Type Culture Collection (ATCC) strains were also included. Isolates used for phylogenetic analyses, MLST and/or PCRs to link $ccrAB_{Ent}$ with surrounding genes are displayed in Table 1.

The *E. faecium* ATCC 19434, *E. faecalis* ATCC 29212, *E. gallinarum* ATCC 49608, *E. faecalis* ATCC 19433, and *E. faecalis* ATCC 51575 were used as controls in species identification. All species were identified by *ddl* PCR (Dutka-Malen *et al.*, 1995) or tRNA intergenic spacer PCR (Baele *et al.*, 2000).

Detection of *ccrAB*$_{Ent}$ genes in the *E. faecium* DO genome by *in silico* analyses. Preliminary sequence data of the *E. faecium* DO strain were obtained from The Joint Genome Institute (JGI) website at http://genome.jgi-psf.org/mic_home.html (Version 4/06/04). Search for homologous proteins were performed using BLAST 2.0 (http://www.ncbi.nlm.nih.gov/ 2010.02.08), and FASTA 33 (http://www.ebi.ac.uk/fasta33/ 2010.02.08). Translation of coding sequences (CDSs), into amino acid sequences was done using ExPASy proteomic tools (http://au.expasy.org/tools/ 2010.02.08).

For prediction of CDSs we used ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html 2010.02.08), Gene Mark (v2.4) (Besemer & Borodovsky, 1999), FGENESB (www.softberry.com 2010.02.08), and ARTEMIS (Wellcome Trust Genome Campus, Hinxton, Cambridge, UK). Pairwise comparison and multiple sequence alignments were performed between the *E. faecium* CcrAB$_{Ent}$ proteins and the previously identified four pairs of *Staphylococcus aureus* CcrABs (CcrAB1, CcrAB2, CcrAB3, CcrAB4) (GenBank accession no. AB033763, D86934, AB037671 and AF411935) and CcrC (GenBank accession no. AB121219). Since *ccrB1* and *ccrB4* were truncated due to frame shift mutation, 1626 bp (*ccrB1*) and 1629 bp (*ccrB4*) ORFs were reconstituted by adding adenine to deleted positions in order to make the alignment better with the *Staphylococcus hominis ccr* sequence (GenBank accession no. AB063171) which has been fully sequenced (Ito *et al.*, 2001). The comparison of DNA sequences were performed in BioEdit V.7.0.5.3 (http://www.mbio.ncsu.edu/BioEdit/bioedit.html), while multiple alignments were done using ClustalW (http://www.ebi.ac.uk/Tools/clustalw2/index.html) or T-Coffee (http://www.ebi.ac.uk/Tools/t-coffee/index.html).

The evolutionary relationships of $CcrAB_{Ent}$, Ccr of staphylococci (deduced from *ccrA1, ccrA2*, *ccrA3, ccrA4, ccrB1*, *ccrB3*, *ccrB4*, and *ccrC*), and three other site specific recombinases (site-specific integrase of bacteriophage phi-FC1 found in *E. faecalis* and two site-specific recombinases from *Clostridium acetobutylicum* ATCC824) were further investigated. These were included because they have been part of previous similar analyses (Ito *et al.*, 2004) and because the *ccrA* and *ccrB* as well as one of the recombinases from *C. acetobutylicum* (AE007725) have been annotated as if they were DNA invertase Pin homologue proteins. The full-length *ccrB1* of NCTC10442 and *ccrB4* of HDE288 were reconstituted as described earlier (Ito *et al.*, 2004). A neighbour-joining tree was constructed using MEGA3 (Kumar *et al.*, 2004) by creating 2000 bootstrap replicates. Sites with gaps/missing data were excluded during analyses. Recombination within the sequenced regions of $ccrA_{Ent}$ and $ccrB_{Ent}$ was determined by phi test (Bruen *et al.*, 2006).

Protein structures were predicted using PSTPRED v2.4 (http://bioinf.cs.ucl.ac.uk/psipred/ 2008.12.17) and the determinations of protein superfamilies were done using HMM library, Genome assignment v1.65 (http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/ 2010.02.08), InterProscan (http://www.ebi.ac.uk/InterProScan/ 2010.02.08), and Pfam (http://pfam.sanger.ac.uk/ 2010.02.08). The programs EditSeq and SeqMan (DNASTAR Inc, Madison, Wisconsin, USA) were used for sequence analysis. To detect repeat sequences Nucleic Acid Dot Plot (http://arbl.cvmbs.colostate.edu/molkit/dnadot/index.html 2010.06.17) and Dotlet database (http://myhits.isb-sib.ch/cgi-bin/dotlet 2010.02.08) were used.

DNA extraction, PCR amplification, and DNA sequencing. Bacterial DNA extraction for PCR analyses was performed manually using the InstaGene matrix kit (Biorad, Irvine, California, USA) or the GenoM™-48 robotic workstation using GenoPrep™ DNA from

blood, standard kit (Genovision, West Chester, Pennsylvania, USA). DNA for hybridisation purposes was isolated using guanidium isothiocyanate (Dahl & Sundsfjord, 2003).

For long range PCR 2U DNA polymerase enzyme r*Tth* XL (Perkin Elmer, New Jersey, USA) were used per reaction and 1.4 mM $Mg(OAc)_2$ in a standard XL PCR reaction mix, or a 0.7x *Pfu* Ultra mix (Stratagene, La Jolla, CA, USA) with 2.5U *Pfu* Ultra polymerase per reaction. DNA sequencing was performed using BigDye 3.1 technology (Applied Biosystems, Foster City, CA, USA). Real-time PCR was performed using ABI Prism 7300 real-time PCR system (PE Biosystems, Warrington, England) and TaqMan universal mastermix (Applied Biosystems).

Detection of *ccrAB*$_{\mathbf{Ent}}$ genes and PCR linkage to surrounding genes. *ccrAB*$_{\mathrm{Ent}}$ genes were detected by PCR, using the primer pairs FA-RA, and FB-RB, respectively (Table 2), and for selected isolates by Southern hybridisation and DNA sequencing. PCRs were also performed on 13 of 14 *ccrAB*$_{\mathrm{Ent}}$ positive *E. faecium* isolates selected for MLST as well as two *ccrAB*$_{\mathrm{Ent}}$ positive *E. faecium* animal isolates from Norway to search for presence and conservation of gene synteny in the surrounding genes (Table 2 and Fig. 1a). Primers and probes were designed using *E. faecium* DO sequences as template.

Expression analysis of *ccrAB*$_{\mathbf{Ent}}$ genes by real-time quantitative PCR. To analyse if *ccrAB*$_{\mathrm{Ent}}$ genes are expressed, *E. faecium* DO was grown aerobically in Mueller Hinton (MH) broth at 37°C for 18-24 hours. Subsequently the culture was diluted 1:10 in MH broth and grown with agitation to A_{600} of 0.7 or to stationary phase (grown over night). The cell suspension was centrifuged and the cells were immediately frozen on dry ice or liquid N_2 before adding an RNA stabilizing solution, RNA later (Ambion, Austin, Texas, USA).

177 Alternatively RNA later or RNA protect (Qiagen, Hilden, Germany) were added directly to
178 the inoculum, according to the manufacturer's instructions. RNA extraction was performed
179 using RNeasy mini kit (Qiagen) using a prolonged lysis step of 45 min with 10 mg lysozyme
180 and 10 U mutanolysin in total volume of 100 µl. On-column DNase treatment was performed
181 according to the manufacturer's instructions. RNA integrity was determined by agarose gel
182 electrophoresis. Reverse transcription of the total RNA was performed using the ABRTR1
183 primer and the High Capacity cDNA Reverse Transcription kit (Applied Biosystems). Real
184 time PCR was performed on the cDNA using primers ccrAFre, ccrARre, ccrBFre, ccrBRre,
185 recAFre, recARre, pbp5Fre, pbp5Rre, adkFre and adkRre and probes $ccrA_{Ent}$, $ccrB_{Ent}$, recA,
186 pbp5 and adk (Table 2). Expression of *ccrAB*$_{Ent}$ genes was compared to the expression of
187 *recA, pbp5*, and *adk*. Ten-fold dilutions of *E. faecium* DO genomic DNA were used to make
188 standard curves for the relative expression of the genes. Experiments were performed in
189 triplicate.
190
191 **Analysis of *ccrAB*$_{Ent}$ mRNA linkage by RT-PCR.** RNA isolation was performed as
192 described above. RNA was treated with the DNA-free kit (Ambion). Reverse transcription of
193 total RNA was performed with SuperScript III reverse transcriptase (Invitrogen) using
194 primers CcrBRTR1 or CcrBxR. RT-PCR without reverse transcriptase was performed on total
195 RNA to check for DNA contamination. Linkage of *ccrA*$_{Ent}$ and *ccrB*$_{Ent}$ mRNAs as a
196 bicistronic mRNA was analysed by PCRs on cDNAs using primers located in *ccrA*$_{Ent}$
197 (CcrARTR1 and CcrAxF) and *ccrB*$_{Ent}$ (CcrBRTR1 and CcrBxR) (Fig. 1b and Table 2).
198
199 **Southern blot hybridisation analyses.** RFLP with *Xba*I (Promega) was performed on total
200 genomic DNA for selected *E. faecium* isolates (DO, TUH7-55, E0470, E0745, E1304, and
201 E1293). PFGE of *Sma*I digested DNA from 76 *E. faecium* isolates were performed according

to Dahl *et al.* (Dahl *et al.*, 1999). DNA fragments separated by gel electrophoresis were transferred to a positively charged nylon membrane (Boehringer, Mannheim, Germany) by vacuum blotting using a Vacugene™ XL system (Amersham Biosciences, Uppsala, Sweden). Southern blot hybridization was performed with DIG labelled (Boehringer Mannheim) *ccrA* probe based on *E. faecium* DO.

MLST. MLST was performed using the following primers: adk1n, adk2n, atpA1n, atpA2n, ddl1, ddl2, gdh1, gdh2, gyd-1, gyd2, pstS1n, pstS2, purK1n, and purK2n (Homan *et al.*, 2002) on a subset of isolates.

RESULTS AND DISCUSSION

ccrAB$_{Ent}$ sequences in the *E. faecium* DO genome

Genes similar to the *ccrA* and *ccrB* genes of *S. aureus* (GenBank accession no. D86934) were identified in the draft sequence of *E. faecium* DO genome. BLAST searches indicated two CDSs in *E. faecium* DO contig 655 (Version 4/06/04) similar to and in identical order as the staphylococcal *ccrA* and *ccrB*. They were named *ccrA*$_{Ent}$ and *ccrB*$_{Ent}$. No available reports have previously shown *ccrA*$_{Ent}$ and *ccrB*$_{Ent}$ genes in enterococci. The *ccrA*$_{Ent}$ and *ccrB*$_{Ent}$ CDSs are 1374 bp and 1638 bp in size, respectively. The two *ccrAB*$_{Ent}$ genes in *E. faecium* DO showed quite similar length as the staphylococcal *ccrAB2* (Katayama et al., 2000).

The *ccrAB*$_{Ent}$ gene synteny was confirmed to be the same as in staphylococci (Katayama *et al.*, 2000) for 14 of 15 *E. faecium* isolates by linkage PCR (Table 3). No available results have previously shown whether staphylococcal *ccrA* and *ccrB* genes are transcribed as separate units or as a bicistronic mRNA. RT-PCR analysis of total RNA from *E. faecium* DO revealed that the *ccrAB*$_{Ent}$ genes were transcribed as a bicistronic mRNA, confirming the

bioinformatics results. Knowing the function of *ccrAB* in staphylococci we hypothesize that *ccrAB*$_{Ent}$ genes in enterococci might be part of a larger integrative genetic element in *E. faecium*. The GC content of *E. faecium* DO contig 655 (35 %), the *ccrAB*$_{Ent}$ CDSs (35 %), and the whole genome (38 %) is not substantially different. No putative termini (inverted repeats) were identified in contig 655 by Nucleic Acid Dot Plot or DotLet analyses. Thus it was not possible to identify a putative integrative element. The genome sequence of contig 655 is limited to the *tnp* transposase determinant (Fig. 1a) at the left side and it has not been possible to identify the continuation of this sequence in another DO contig. The sequence at the other side of *ccrAB*$_{Ent}$ also contain a lot of putative transposases (belonging to several IS families) in addition to hypothetical proteins (http://maple.lsd.ornl.gov/cgi-bin/JGI_microbial/contig_viewer.cgi?org=efae&chr=08jun04&contig=Contig655&sort=left_bp 2010. 21.06) which may well be part of an integrative element.

Pairwise comparison and multiple sequence alignments were performed between the *E. faecium* CcrAB$_{Ent}$ proteins and the Ccrs of *S. aureus*. The similarities of CcrA and CcrB between *E. faecium* and *S. aureus* N315 were 55% and 69%, respectively. The N-terminal resolvase and recombinase domains, as well as the predicted catalytic serine residue of the recombinase active site were highly conserved between the *Staphylococcus* and *Enterococcus* CcrAB proteins. Moreover, the *Enterococcus* CcrB$_{Ent}$ was predicted to include an Ogr/delta like domain (a phage transcription activator). Two algorithms, Pfam and ProScan predicted both the resolvase and recombinase domains in the examined Ccr protein sequences (Table S1).

The evolutionary relationships of CcrAB$_{Ent}$, Ccr of staphylococci, and three other site specific recombinases were further investigated. The phylogenetic analyses revealed an evolutionary

252 relationship between $CcrA_{Ent}$ and $CcrB_{Ent}$ from enterococci and the staphylococcal CcrAB-
253 cluster (Fig. 2). However, the low identity score between the enterococcal and staphylococcal
254 proteins does not support a recent horizontal transfer of the *ccr* genes between these species.
255
256 **$ccrAB_{Ent}$ genes are expressed in *E. faecium***
257 Analyses of $ccrAB_{Ent}$ gene expression were performed during both the exponential and
258 stationary phase of *E. faecium* DO grown in rich medium. Both genes were expressed in
259 approximately the same amounts in exponential phase. $ccrAB_{Ent}$ genes were expressed
260 approximately 10^3 and 10-10^2 times lower than *recA* and *adk*, respectively, and approximately
261 2-10 folds higher than the *pbp5* gene (Fig. S1). The mRNA abundance of $ccrAB_{Ent}$ was lower
262 in stationary phase than in exponential phase.
263
264 **$ccrAB_{Ent}$ genes are dispersed among *Enterococcus* species belonging to *E. faecium* and *E.***
265 ***casseliflavus* groups**
266 Of a total of 421 enterococcal isolates 118 (28 %) were positive for $ccrAB_{Ent}$ genes in five
267 species by PCR; *E. faecium* (n = 58/216, 27 %), *E. durans* (n=31/38, 82 %), *E. hirae*
268 (n=27/52, 50 %), *E. casseliflavus* (n=1/4, 25 %), and *E. gallinarum* (n=1/2, 50 %) (Table 1).
269 One *E. hirae* isolate was PCR positive for $ccrB_{Ent}$ only. Eight other species including *E.*
270 *faecalis* were negative for $ccrAB_{Ent}$ (data not shown).
271
272 A BLAST search (2010.04.19) for the $ccrAB_{Ent}$ genes and the surrounding regions against
273 *Enterococcus* strains revealed the presence of $ccrAB_{Ent}$ in *E. faecium* E1071, 1,231,408 and
274 C68 (http://www.ncbi.nlm.nih.gov/genomes/geblast.cgi?gi=6512#SearchSet 2010.21.06) and
275 no $ccrAB_{Ent}$ sequence or protein matches with high identity scores in other available
276 *Enterococcus* genomes (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi 2010.21.06).

277 The *E. faecium* E1071 and *E. faecium* 1,231,408 showed sequence similarity with DO
278 sequence in parts of the hypothetical protein, the *ccrB*$_{Ent}$ and parts of the *ccrA*$_{Ent}$. The *E.*
279 *faecium* C68 showed similarity with DO in parts of the hypothetical protein, both *ccrA*$_{Ent}$ and
280 *ccrB*$_{Ent}$ and parts or the replication initiation factor (REP).
281
282 *ccrAB*$_{Ent}$ gene sequences (GenBank accession no. FJ572967 - FJ573039) from *E. faecium*
283 (n=14), *E. hirae* (n=10 for *ccrA*$_{Ent}$ and 11 for *ccrB*$_{Ent}$), *E. durans* (n=10), *E. gallinarum* (n=1),
284 and *E. casseliflavus* (n=1) isolates were aligned and a neighbour-joining phylogenetic tree
285 was made with bootstrap of 2000 replicates using the P-distance model (Fig. 3). The *ccrAB*$_{Ent}$
286 genes both clustered into two major clades represented by the majority of *E. faecium* (clade I)
287 and *E. hirae* (clade II) isolates, respectively. With 7 of 10 isolates clustering in clade II *E.*
288 *hirae* appears to be slightly more dispersed between the two *ccrA*$_{Ent}$ clades. *ccrAB*$_{Ent}$ from the
289 *E. gallinarum* and *E. casseliflavus* isolates clustered in clade II with the majority of *ccrAB*$_{Ent}$
290 from the *E. hirae* isolates. In *E. durans* 6 of 10 *ccrA*$_{Ent}$ genes clustered in clade I while 7 of 10
291 *ccrB*$_{Ent}$ clusters in clade II. Except for *ccrA*$_{Ent}$ from *E. faecium* E1304 the *ccrAB*$_{Ent}$ genes of
292 the human isolates clustered in clade I whereas the animal isolates were found in both. The
293 two phylogenetic trees are incongruent meaning that *ccrA*$_{Ent}$ and *ccrB*$_{Ent}$ belong to different
294 clades for 11 of the isolates, all of animal origin (Fig. 3 isolates marked with asterisk). Phi
295 tests revealed no statistically significant evidence for recombination within the sequenced
296 regions of the *ccrA*$_{Ent}$ and *ccrB*$_{Ent}$ genes. However, the incongruence suggests recombination
297 of the *ccr*$_{Ent}$ genes outside the sequenced regions of the two genes. Incongruence between
298 these genes has also been seen for *S. aureus* (Ito *et al.*, 2004).
299
300 *ccrAB*$_{Ent}$ genes were only found in isolates belonging to the *E. faecium* and *E. casseliflavus*
301 species-groups which belong to the same tree-branch in phylogenetic trees based on

enterococcal 16S and *sodA* gene diversity (Devriese *et al.*, 1993; Poyart *et al.*, 2000). The absence of *ccrAB*$_{Ent}$ in the other species could be explained by the low number of isolates tested except for *E. faecalis* or by a lack of integration sites recognised by *ccrAB*$_{Ent}$ in the strains not belonging to the *E. faecium* or *E. casseliflavus* groups. Alternatively, their *ccrAB*$_{Ent}$ genes may exhibit such a low sequence identity to the *ccrAB*$_{Ent}$ genes identified in this study that they are missed using the PCR and hybridisation conditions used in the present study.

Variations of the *ccrAB*$_{Ent}$ genes and the surrounding region between selected *E. faecium* isolates

PFGE analysis and Southern hybridisation of 76 *E. faecium* isolates with *ccrA*$_{Ent}$ probe confirmed the PCR results. One *ccrA*$_{Ent}$ PCR negative strain (399/F98/A1) was *ccrA*$_{Ent}$ positive by Southern blot hybridisation (data not shown) indicating sequence diversity affecting PCR-amplification. Also *Xba*I analyses of *ccrA*$_{Ent}$ and *ccrB*$_{Ent}$ genomic regions revealed heterogeneity and only one copy of *ccrA*$_{Ent}$. The *ccrA*$_{Ent}$ probe hybridized to an approximately 10 kb fragment in DO, TUH7-55, E1304, and E1293 isolates, while the *ccrA*$_{Ent}$-positive fragment of E0470 and E0745 was approximately 24 kb (data not shown). To investigate this in more detail, the presence of *ccrAB*$_{Ent}$ flanking genomic genes identified in the DO genome was determined by multiple PCRs in 15 *ccrAB*$_{Ent}$ positive and 16 *ccrAB*$_{Ent}$ negative isolates (Fig. 1a). Examinations of the *ccrAB*$_{Ent}$ surrounding region in several isolates showed a variable pattern of the *ccrAB*$_{Ent}$ flanking sequences with hospital-associated isolates showing most sequence similarity with the DO sequence (Table 3). All 31 isolates were positive for the *tnp* gene (belonging to the IS*30* family) specific PCR and 3 *ccrAB*$_{Ent}$ positive isolates of different sequence types (STs) were also positive for the REP gene PCR. This REP gene harbours a REP_trans domain belonging to superfamily pfam02486. This family represents probable topoisomerases that makes a sequence-specific single stranded

nick in the origin of replication. To this family belong plasmid REPs, phage REPs (RstAs) and transposon REPs (Cro/CI transcriptional regulators). Long-range PCRs confirmed linkage of these genes with *ccrAB*$_{Ent}$ and conservation of gene synteny surrounding *ccrAB*$_{Ent}$ with the exception of isolates 64/F99/H6, 399/F99/A10, 399/F99/H8, and S399/F99/A14 for which linkage of *tnp*-orf1 and orf1-*ccrB*$_{Ent}$ was not confirmed. Furthermore, a *ccrB*$_{Ent}$-*ccrA*$_{Ent}$ linkage was not shown in 64/F99/H6 (Table 3 and Fig. 1a). The inability to link genes that were positive on gene specific PCRs may indicate that the region between these genes are larger than expected or that the specific genes are located at other regions in the genome. The transposase of the IS*30* family is for instance located at more than one site in *E. faecium* DO. Annotation of contig 655 (http://maple.lsd.ornl.gov/cgi-bin/JGI_microbial/contig_viewer.cgi?org=efae&chr=08jun04&contig=Contig655&sort=left_bp 2010.21.06) also indicates that the *ccrAB*$_{Ent}$ genes are located in a region containing several transposases. The regions surrounding *ccrAB* in *Staphylococcus* contain highly variable genes encoding ORFs of unknown functions. These variable regions are called J1 and J2 and variations in these regions are used to define the SCC*mec* subtypes (International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements (IWG-SCC), 2009), thus our results from enterococci are in line with these observations of highly variable regions surrounding *ccrAB* in staphylococci. CcrA and CcrB have roles in the excision and integration of SCC*mec* in staphylococci (Wang & Archer, 2010) and we have showed that the *ccrAB*$_{Ent}$ genes are expressed in *E. faecium* DO. It has been postulated that SCC may carry the genes conferring methicillin resistance but may also confer genetic exchange of other genes among staphylococcal species (Katayama et al., 2003). However to our knowledge no studies have provided direct experimental evidence for intercellular transfer of SCC between staphylococci.

DNA sequencing of the *ccrAB*$_{Ent}$, *tnp*, and *orf1* of the 15 *ccrAB*$_{Ent}$ positive isolates showed 94 to 100 % and 96 to 100 % sequence identity in *ccrA*$_{Ent}$ and *ccrB*$_{Ent}$ genes (GenBank accession no. FJ572978-FJ572981, FJ572997-FJ573001, FJ573014-FJ573018, FJ573032-FJ573036), respectively, while sequences of *orf1* and *tnp* were 100 % identical in all isolates (data not shown). According to Hanssen and co-workers (Hanssen et al., 2004) up to 4 % variation within the *ccrAB* genes has been observed for a given staphylococcal species. The *ccrAB* genes found in SCC*mec* types II and IV can vary up to 5 % at the nucleotide level (Noto & Archer, 2006). Since both *ccrAB*$_{Ent}$ genes and the staphylococcal *ccrAB* genes show sequence variations within the recombinase genes which have the same gene synteny and variable surrounding regions we hypothesise that they may have similar functions contributing to excision and integration of surrounding genes within the genome and possibly also mobilisation of surrounding genes between cells.

Investigation of possible association between *ccrAB*$_{Ent}$ and sequence type within *E. faecium* of human origin

MLST analyses of *E. faecium* isolates (n=39) revealed that the *ccrAB*$_{Ent}$ genes are dispersed among different STs (Table 1). Ten of 29 (34 %) hospital related *E. faecium* isolates were *ccrAB*$_{Ent}$ positive, while 4 of 10 (40 %) non-hospital related isolates were positive. Furthermore, specific STs within hospital related strains were represented by both *ccrAB*$_{Ent}$ positive and negative isolates (Table 1), suggesting that *ccrAB*$_{Ent}$ are acquired and not a part of the core genome.

Concluding remarks

Cassette chromosome recombinases may be important in recombination and genome plasticity in enterococci. Expression analyses indicate that the recombinase genes are active in

377 *E. faecium* DO, and thus, may play a role in the recombination or movement of genetic
378 elements. Further investigation of the $ccrA_{Ent}$ and $ccrB_{Ent}$ will be essential to reveal the
379 contribution of these genes for recombination and mobilisation events in enterococci.
380
381 **ACKNOWLEDGEMENT**
382 This work was supported by research grants from the Norwegian Research Council (projects
383 no. 165997 and 183653/S19), Northern Norway regional Health Authority Medical Research
384 Program and the European Commission (LSHE-CT-2007-03410 "ACE").
385
386 The DNA sequence of *E. faecium* DO was supported at the JGI by the US Department of
387 Energy's Office of Science, Biological and Environmental Research Program, and by the
388 University of California, Lawrence Berkeley National Laboratory under contract No. DE-
389 AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-
390 07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-
391 06NA25396.
392
393 We would like to thank Professor Barbara E. Murray, University of Texas Medical School at
394 Houston, USA for kindly providing bacterial isolates. We also thank Dr. Stein-Christian
395 Mohn at the University of Bergen for his help and advices, and Trine Tessem at the
396 University of Tromsø for her excellent technical assistance.
397

REFERENCES

Arduino, R. C., Murray, B. E. & Rakita, R. M. (1994). Roles of antibodies and complement in phagocytic killing of enterococci. *Infect Immun* **62**, 987-993.

Baele, M., Baele, P., Vaneechoutte, M., Storms, V., Butaye, P., Devriese, L. A., Verschraegen, G., Gillis, M. & Haesebrouck, F. (2000). Application of tRNA intergenic spacer PCR for identification of *Enterococcus* species. *J Clin Microbiol* **38**, 4201-4207.

Besemer, J. & Borodovsky, M. (1999). Heuristic approach to deriving models for gene finding. *Nucleic Acids Res* **27**, 3911-3920.

Bruen, T. C., Philippe, H. & Bryant, D. (2006). A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**, 2665-2681.

Carias, L. L., Rudin, S. D., Donskey, C. J. & Rice, L. B. (1998). Genetic linkage and cotransfer of a novel, *vanB*-containing transposon (Tn*5382*) and a low-affinity penicillin-binding brotein 5 gene in a clinical vancomycin-resistant *Enterococcus faecium* isolate. *J Bacteriol* **180**, 4426-4434.

Dahl, K. H., Simonsen, G. S., Olsvik, Ø. & Sundsfjord, A. (1999). Heterogeneity in the *vanB* gene cluster of genomically diverse clinical strains of vancomycin-resistant enterococci. *Antimicrob Agents Chemother* **43**, 1105-1110.

Dahl, K. H. & Sundsfjord, A. (2003). Transferable *vanB2* Tn*5382*-containing elements in fecal streptococcal strains from veal calves. *Antimicrob Agents Chemother* **47**, 2579-2583.

Devriese, L. A., Pot, B. & Collins, M. D. (1993). Phenotypic identification of the genus *Enterococcus* and differentiation of phylogenetically distinct enterococcal species and species groups. *J Appl Bacteriol* **75**, 399-408.

Dutka-Malen, S., Evers, S. & Courvalin, P. (1995). Detection of glycopeptide resistance genotypes and identification to the species level of clinically relevant enterococci by PCR [published erratum appears in J Clin Microbiol 1995 May;33(5):1434]. *J Clin Microbiol* **33**, 24-27.

Hanssen, A. M., Kjeldsen, G. & Sollid, J. U. E. (2004). Local variants of staphylococcal cassette chromosome *mec* in sporadic methicillin-resistant *Staphylococcus aureus* and methicillin-resistant coagulase-negative staphylococci: Evidence of horizontal gene transfer? *Antimicrob Agents Chemother* **48**, 285-296.

Homan, WL., Tribe, D., Poznanski, S., Li, M., Hogg, G., Spalburg, E., van Embden, J. & Willems, R. (2002). Multilocus sequence typing scheme for *Enterococcus faecium*. *J Clin Microbiol* **40**, 1963-1971.

International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements (IWG-SCC) (2009). Classification of staphylococcal cassette chromosome *mec* (SCC*mec*): guidelines for reporting novel SCC*mec* elements. *Antimicrob Agents Chemother* **53**, 4961-4967.

Ito, T., Ma, X. X., Takeuchi, F., Okuma, K., Yuzawa, H. & Hiramatsu, K. (2004). Novel type V staphylococcal cassette chromosome *mec* driven by a novel cassette chromosome recombinase, *ccrC*. *Antimicrob Agents Chemother* **48**, 2637-2651.

Ito, T., Katayama, Y. & Hiramatsu, K. (1999). Cloning and nucleotide sequence determination of the entire *mec* DNA of pre-methicillin-resistant *Staphylococcus aureus* N315. *Antimicrob Agents Chemother* **43**, 1449-1458.

Ito, T., Katayama, Y., Asada, K., Mori, N., Tsutsumimoto, K., Tiensasitorn, C. & Hiramatsu, K. (2001). Structural comparison of three types of staphylococcal cassette chromosome *mec* integrated in the chromosome in methicillin-resistant *Staphylococcus aureus*. *Antimicrob Agents Chemother* **45**, 1323-1336.

Johnsen, P. J., Østerhus, J. I., Sletvold, H., Sørum, M., Kruse, H., Nielsen, K., Simonsen, G. S. & Sundsfjord, A. (2005). Persistence of animal and human glycopeptide-resistant enterococci on two Norwegian poultry farms formerly exposed to avoparcin is associated with a widespread plasmid-mediated *vanA* element within a polyclonal *Enterococcus faecium* population. *Appl Environ Microbiol* **71**, 159-168.

Katayama, Y., Takeuchi, F., Ito, T., Ma, X. X., Ui-Mizutani, Y., Kobayashi, I. & Hiramatsu, K. (2003). Identification in methicillin-susceptible *Staphylococcus hominis* of an active primordial mobile genetic element for the staphylococcal cassette chromosome *mec* of methicillin-resistant *Staphylococcus aureus*. *J Bacteriol* **185**, 2711-2722.

Katayama, Y., Ito, T. & Hiramatsu, K. (2000). A new class of genetic element, *Staphylococcus* cassette chromosome *mec*, encodes methicillin resistance in *Staphylococcus aureus*. *Antimicrob Agents Chemother* **44**, 1549-1555.

Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* **5**, 150-163.

Leavis, H. L., Bonten, M. J. & Willems, R. J. (2006). Identification of high-risk enterococcal clonal complexes: global dispersion and antibiotic resistance. *Curr Opin Microbiol* **9**, 454-460.

Leavis, H. L., Willems, R. J., van Wamel, W. J., Schuren, F. H., Caspers, M. P. & Bonten, M. J. (2007). Insertion sequence-driven diversification creates a globally dispersed emerging multiresistant subspecies of *E. faecium*. *PLoS Pathog* **3**, e7.

Noto, M. J. & Archer, G. L. (2006). A subset of *Staphylococcus aureus* strains harboring staphylococcal cassette chromosome *mec* (SCC*mec*) type IV is deficient in CcrAB-mediated SCC*mec* excision. *Antimicrob Agents Chemother* **50**, 2782-2788.

Poyart, C., Quesnes, G. & Trieu-Cuot, P. (2000). Sequencing the gene encoding manganese-dependent superoxide dismutase for rapid species identification of enterococci. *J Clin Microbiol* **38**, 415-418.

Poyart, C. & Trieu-Cuot, P. (1994). Heterogeneric conjugal transfer of the pheromone-responsive plasmid pIP964 (IncHlyI) of *Enterococcus faecalis* in the apparent absence of pheromone induction. *FEMS Microbiol Lett* **122**, 173-179.

van Schaik, W., Top, J., Riley, D., Boekhorst, J., Vrijenhoek, J., Schapendonk, C., Hendrickx, A., Nijman, I., Bonten, M., Tettelin, H. & Willems, R. (2010). Pyrosequencing-based comparative genome analysis of the nosocomial pathogen *Enterococcus faecium* and identification of a large transferable pathogenicity island. *BMC Genomics* **11**, 239.

Wang, L. & Archer, G. L. (2010). Roles of CcrA and CcrB in excision and integration of staphylococcal cassette chromosome *mec*, a *Staphylococcus aureus* genomic island. *J Bacteriol* **192**, 3204-3212.

Werner, G., Willems, R. J. L., Hildebrandt, B., Klare, I. & Witte, W. (2003). Influence of transferable genetic determinants on the outcome of typing methods commonly used for *Enterococcus faecium*. *J Clin Microbiol* **41**, 1499-1506.

Willems, R. J., Top, J., van, S. M., Robinson, D. A., Coque, T. M., Baquero, F., Grundmann, H. & Bonten, M. J. (2005). Global spread of vancomycin-resistant *Enterococcus faecium* from distinct nosocomial genetic complex. *Emerg Infect Dis* **11**, 821-828.

490 **Table 1.** *Enterococcus* isolates selected for MLST typing, phylogenetic analyses and/or PCRs to link $ccrAB_{Ent}$ with surrounding genes. Origin

491 and sources, type of vancomycin resistance, ST and hospital related STs (CC17 genotype), and presence of $ccrAB_{Ent}$ genes are given.

Isolate Designation	**Species**	**Geographical origin Country/ Region**	**Sample source**	**Epidemiology***	***van* type**	**ST**	***ccrAB*$_{Ent}$**	**Reference/ Provided by**
C68	*E. faecium*	USA /Ohio	Human Faeces	Clinical isolate	*vanB*	16(CC17)	AB	Carias *et al.*, 1998
E0470	*E. faecium*	Netherlands/Amsterdam	Human Blood	HO	*vanA*	16(CC17)	AB	Willems *et al.*, 2005
E0734	*E. faecium*	Netherlands/Amersfoort	Hospital Faeces	HO	*vanA*	16(CC17)	AB	Willems *et al.*, 2005
E0745	*E. faecium*	Netherlands/Utrecht	Human Faeces	HO	*vanA*	16(CC17)	AB	Willems *et al.*, 2005
TUH7-15	*E. faecium*	USA	Human Blood	HO	*vanB*	16(CC17)	AB	Dahl *et al.*, 1999
E0510	*E. faecium*	Australia/Melbourne	Human Blood	HO	*vanB*	17(CC17)	-‡	Willems *et al.*, 2005
TUH2-18	*E. faecium*	Norway/Bergen	Human urine	HO	*vanB*	17 (CC17)	-	Dahl *et al.*, 1999
TUH2-19	*E. faecium*	Norway/Bergen	Human Wound	HO	*vanB*	17 (CC17)	-	Dahl *et al.*, 1999
TUH7-55	*E. faecium*	Germany	Human Urine	Clinical isolate	*vanB*	17 (CC17)	AB	Dahl *et al.*, 1999
DO (TX0016)	*E. faecium*	USA/Houston	Human Blood	Clinical isolate	-	18 (CC17)	AB	Arduino *et al.*, 1994
E1652	*E. faecium*	Netherlands/Amersfoort	Human Faeces	HO	*vanA*	18 (CC17)	-	Willems *et al.*, 2005
E1406	*E. faecium*	Spain/Madrid	Human Blood	HP	ND†	63(CC17)	-	T M. Coque/ R. Willems
E1392	*E. faecium*	Great Britain/Centre H	Human	HP	ND	64(CC17)	-	N. Woodford/ R. Willems
E1181	*E. faecium*	Austria/Linz	Human Blood	HP	ND	78(CC17)	-	ENARE/ R. Willems
E1186	*E. faecium*	Germany	Human Blood	HP	ND	78(CC17)	-	ENARE/R. Willems
E1321	*E. faecium*	Italy/Rome	Human Catheter	HP	ND	78(CC17)	-	L. Baldassarri/ R. Willems
E1644	*E. faecium*	Germany/Freiburg	Human Catheter urine	HP	ND	78(CC17)	-	D. Jonas/ R. Willems
E0333	*E. faecium*	Israel/Centre1	Human Blood	HP	ND	80(CC17)	-	R. Schouten/ R. Willems
E1775	*E. faecium*	Belgium	Pig Faeces		ND	121(CC17)	AB	E. de Leener/ R. Willems
E1173	*E. faecium*	Portugal/Coimbra	Human Wound	Clinical isolate	*vanA*	125 (CC17)	-	Willems *et al.*, 2005
E1304	*E. faecium*	Portugal/Coimbra	Human Blood	Clinical isolate	*vanA*	132 (CC17)	AB	Willems *et al.*, 2005
E1762	*E. faecium*	Australia/Perth	Human	Hospital survey	ND	174(CC17)	-	W. Grubb/ R. Willems
U0105/	*E. faecium*	Netherlands	Human Blood	HP	ND	267(CC17)	-	A. Troelstra/ R. Willems
3332	*E. faecium*	USA/Ohio	Human	HO	*vanB*	308(CC17)	AB	Carias *et al.*, 1998
TUH4-65	*E. faecium*	USA	Human	Clinical isolate	*vanB*	313 (CC17)	-	Dahl *et al.*, 1999
E0125	*E. faecium*	Netherlands/Rotterdam	Human Bile	Clinical isolate	*vanA*	5	-	Willems *et al.*, 2005
399/F98/H2	*E. faecium*	Norway/Østfold	Human Faeces	CS	*vanA*	8	-	Johnsen *et al.*, 2005
64/3	*E. faecium*	Germany	Human Faeces	HP	-	21	-	Werner *et al.*, 2003
E0073	*E. faecium*	Netherlands/Rotterdam	Human Faeces	Clinical isolate	*vanA*	22	-	Willems *et al.*, 2005

S399/S99/H5	*E. faecium*	Norway/Østfold	Human Faeces	CS	-	48	-	Johnsen *et al.*, 2005
64/F98/H1	*E. faecium*	Norway/Østfold	Human Faeces	CS	*vanA*	48	-	Johnsen *et al.*, 2005
E1293	*E. faecium*	Italy/Genova	Human Blood	Clinical isolate	-	50	AB	Willems *et al.*, 2005
E1626	*E. faecium*	Netherlands	Human Peritoneal fluid	Clinical isolate	-	92	-	Willems *et al.*, 2005
BM4105-RF	*E. faecium*	France	Human Faeces	Wild strain	-	172	-	Poyart & Trieu-Cuot, 1994
399/F99/H8	*E. faecium*	Norway/Østfold	Human Faeces	CS	*vanA*	195	AB	Johnsen *et al.*, 2005
64/F99/H6	*E. faecium*	Norway/Østfold	Human Faeces	CS	*vanA*	246	AB	Johnsen *et al.*, 2005
399/F99/A10	*E. faecium*	Norway/Østfold	Animal Faeces	CS	*vanA*	310	AB	Johnsen *et al.*, 2005
399/F98/A1	*E. faecium*	Norway/Østfold	Animal Faeces	CS	*vanA*	311	-[§]	Johnsen *et al.*, 2005
S399/F98/H3	*E. faecium*	Norway/Østfold	Human Faeces	CS	-	312	-	Johnsen *et al.*, 2005
K17a	*E. faecium*	Belgium	Chicken		ND	ND	AB	P. Butaye
K40b	*E. faecium*	Belgium	Chicken		ND	ND	AB	P. Butaye
S399/S99/A4	*E. faecium*	Norway/Østfold	Animal Faeces	CS	-	ND	AB	Johnsen *et al.*, 2005
S399/F99/A14	*E. faecium*	Norway/Østfold	Animal Faeces	CS	-	ND	AB	Johnsen *et al.*, 2005
V63b	*E. faecium*	Belgium	Pig		ND	ND	AB	P. Butaye
V128	*E. faecium*	Belgium	Pig		ND	ND	AB	P. Butaye
K101b	*E. durans*	Belgium	Chicken		ND	ND	AB	P. Butaye
K4a	*E. durans*	Belgium	Chicken		ND	ND	AB	P. Butaye
K21b	*E. durans*	Belgium	Chicken		ND	ND	AB	P. Butaye
K70	*E. durans*	Belgium	Chicken		ND	ND	AB	P. Butaye
K89	*E. durans*	Belgium	Chicken		ND	ND	AB	P. Butaye
K116a	*E. durans*	Belgium	Chicken		ND	ND	AB	P. Butaye
K118c	*E. durans*	Belgium	Chicken		ND	ND	AB	P. Butaye
K120a	*E. durans*	Belgium	Chicken		ND	ND	AB	P. Butaye
K121	*E. durans*	Belgium	Chicken		ND	ND	AB	P. Butaye
96 b	*E. durans*	Belgium	Dog		ND	ND	AB	P. Butaye
K51b	*E. hirae*	Belgium	Chicken		ND	ND	AB	P. Butaye
K56b	*E. hirae*	Belgium	Chicken		ND	ND	AB	P. Butaye
K66a	*E. hirae*	Belgium	Chicken		ND	ND	AB	P. Butaye
K73a	*E. hirae*	Belgium	Chicken		ND	ND	AB	P. Butaye
K74b	*E. hirae*	Belgium	Chicken		ND	ND	AB	P. Butaye
K79b	*E. hirae*	Belgium	Chicken		ND	ND	AB	P. Butaye
K115b	*E. hirae*	Belgium	Chicken		ND	ND	AB	P. Butaye
K141	*E. hirae*	Belgium	Chicken		ND	ND	AB	P. Butaye

V70b	*E. hirae*	Belgium	Pig	ND	ND	AB	P. Butaye
V106c	*E. hirae*	Belgium	Pig	ND	ND	B	P. Butaye
81 a	*E. hirae*	Belgium	Dog	ND	ND	AB	P. Butaye
86	*E. casseliflavus*	Belgium	Chicken	ND	ND	AB	P. Butaye
327	*E. gallinarum*	Belgium	Chicken	ND	ND	AB	P. Butaye

[*]HO, hospital outbreak; CS, Community survey; HP, Hospitalized patient

[†]ND, not determined

[‡]-, negative

[§], positive for $ccrA_{Ent}$ only by Southern hybridization

Table 2. Oligonucleotides used for expression analyses of *ccrAB*$_{Ent}$ genes and for detection/characterisation of the *ccrAB*$_{Ent}$ region and detection of enterococcal virulence genes.

Target gene	Primer name	Sequence 5´-3´	Product size (bp)	Reference
ccrA$_{Ent}$*	ccrAFre	AACGATTGACGCAACAAAAGCT		This study
	ccrARre	CGCCATAGTACAATGGATTTTTTAGGATAT	129	
	ccrA$_{Ent}$ probe	TCCGCGAACGTCCTTT		
ccrB$_{Ent}$*	ccrBFre	TTTTCTACCACGGCAGTCAAAGAT		This study
	ccrBRre	CAATTGATGTAGCGCGCATATTCTA	68	
	ccrB$_{Ent}$ probe	ACCCTGCATAAATTTT		
*recA**	recAFre	GATTCAGTTGCTGCTTTAGTTCCA		This study
	recARre	CTTGTAACCCGACATGTGAGTCA	72	
	recA probe	TTCGCCGTCGATTTC		
*pbp5**	pbp5Fre	GATCTGGTTTGGAAATGGCTTTTGA		This study
	pbp5Rre	CACCGTCTGTATCTGTGATGCTTAA	79	
	pbp5 probe	TCCCACGAAGATCCTT		
*adk**	adkFre	CCACGTACGCTAGATCAAGCAA		This study
	adkRre	CATGGATATCGATGACAGCATCAATTTT	85	
	adk probe	ATTGCGTCCAGAGCTT		
ccrAB$_{Ent}$†	ccrAxF	CGAAAAGCGAAAAGATGAAAAACACAAAGT		This study
	ccrARTR1	ACCTCGATCCGACAAACATGGTCACATAAC	222	
	ccrBxR	ACATAGCCTAAACGTCGTCCACCTG	625	
	ccrBRTR1	TAACCCCACATCATATCGCAACAGTTCCTC	801	
ccrA$_{Ent}$‡	ccrAF	GAAATATGAACAAATTCCCCAACG	1242	This study
	ccrAR	CGGAAGTAAATCCCACAGACT		
ccrB$_{Ent}$‡	ccrBF	GGAACCATCGTTTTGATCTACTAG	1321	This study
	ccrBR	GCAGGCGTGAATTTCATTGTA		
ccrA$_{Ent}$§	ccrAF	GAAATATGAACAAATTCCCCAACG	451	This study
		TTGAAAAATATAGCGAACAATCC		

	J03/252ccrARB			
ccrB$_{Ent}$ [§]	J03/252ccrBF	TCGGAATAAAGGAGCAAGTGTG	525	This study
	ccrBR	GCAGGCGTGAATTTCATTGTA		
ccrA$_{Ent}$ [¶]	FA	CCATATGGGTATCGTTTAGTGA	453	This study
	RA	AGCTTCGGTCGGTACAATGAT		
ccrB$_{Ent}$ [¶]	FB	ATTTGTCGCCGACCGATTAAAG	390	This study
	RB	ACGATACAAGGCTTTGAYTTGCT		
ORF1	1259F1	ATTTGTTACTGAATCCAGTGCTTACTC	873	This study
	1259R1	CAATGTTATTCTGCTTGAACTTGACC		
Replication initiation	1259F2	GCTAGGAGTACAAAATATCCAACGC	721	This study
factor gene	1259R2	CTGAATAATTCTCCGTATGAGAGCG		
tnp	1259F6	CGAAGCAGCTTAAACGTGGAC	759	This study
	1259R6	GGATATGGTTTCTTTTGGACGC		

[*]primers and probes used for the expression study

[†]primers used for *ccrAB*$_{Ent}$ linkage of RT-PCR product and RT-PCR control

[§]PCRs used to sequence a part of the genes/CDSs

[‡]primers used for detection in the early phase of the study. Following sequence analysis these were later exchanged with the new primers[¶]

Table 3. Long range linkage PCR results of $ccrA_{Ent}$ and $ccrB_{Ent}$ chromosomal region among 15 $ccrAB_{Ent}$ positive *E. faecium* isolates.

Isolate designation	**ST**[*]	**Long range linkage PCRs**			
		***tnp*-orf1**	**orf1-$ccrB_{Ent}$**	**$ccrB_{Ent}$ - $ccrA_{Ent}$**	**$ccrA_{Ent}$ -*REP factor***
DO	**18**	+[†]	+	+	+
E1304	**132**	+	+	+	+
TUH7-55	**17**	+	+	+	+
3332	**308**	+	+	+	NA[§]
C68	**16**	+	+	+	NA
E0470	**16**	+	+	+	NA
E0734	**16**	+	+	+	NA
E0745	**16**	+	+	+	NA
TUH7-15	**16**	+	+	+	NA
64/F99/H6	48	-[‡]	-	-	NA
399/F99/A10	310	-	-	+	NA
399/F99/H8	195	-	-	+	NA
E1293	50	NA	NA	+	NA
S399/F99/A14	ND[¶]	-	-	+	NA
S399/S99/A4	ND	NA	NA	+	NA

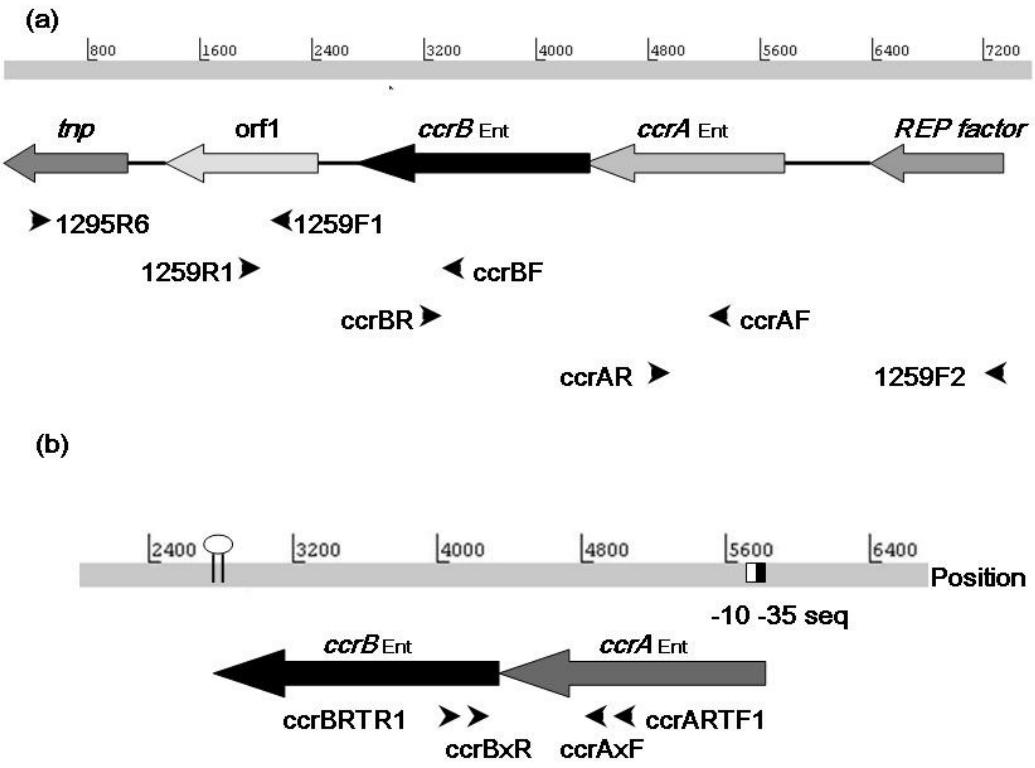
[*]STs shown in bold belong to the CC17 genotype

[†]positive

[‡]negative

[§]not applicable (one of the genes/CDSs not present)

[¶]not determined



509

510 **Fig. 1.** (a) Schematic presentation of the *ccrAB*$_{Ent}$ region of *E. faecium* DO and the long-range
511 PCRs used to link genes surrounding the *ccrA*$_{Ent}$ and *ccrB*$_{Ent}$ genes in *E. faecium*. (b)
512 Schematic presentation of the *ccrAB*$_{Ent}$ indicating the positions of the PCR primers used for
513 mRNA linkage. Linkage of *ccrA*$_{Ent}$ and *ccrB*$_{Ent}$ mRNAs was performed using combinations of
514 primers ccrAxF/ccrBRTR1, ccrAxF/CcrBxR, CcrARTF1/ccrBRTR1 and CcrARTF1/
515 CcrBxR.

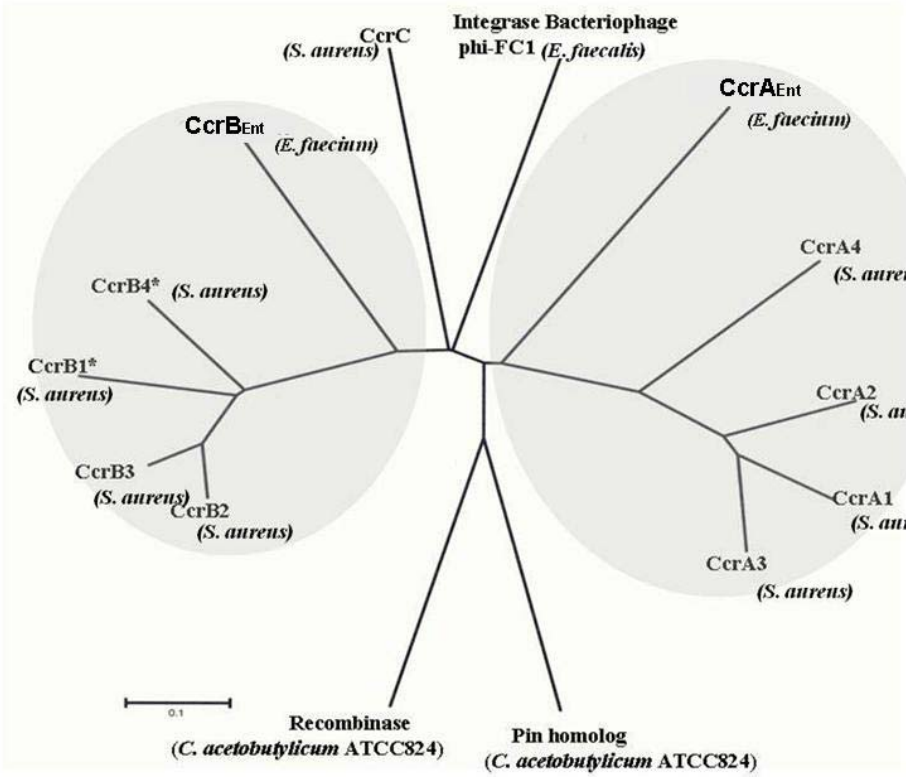
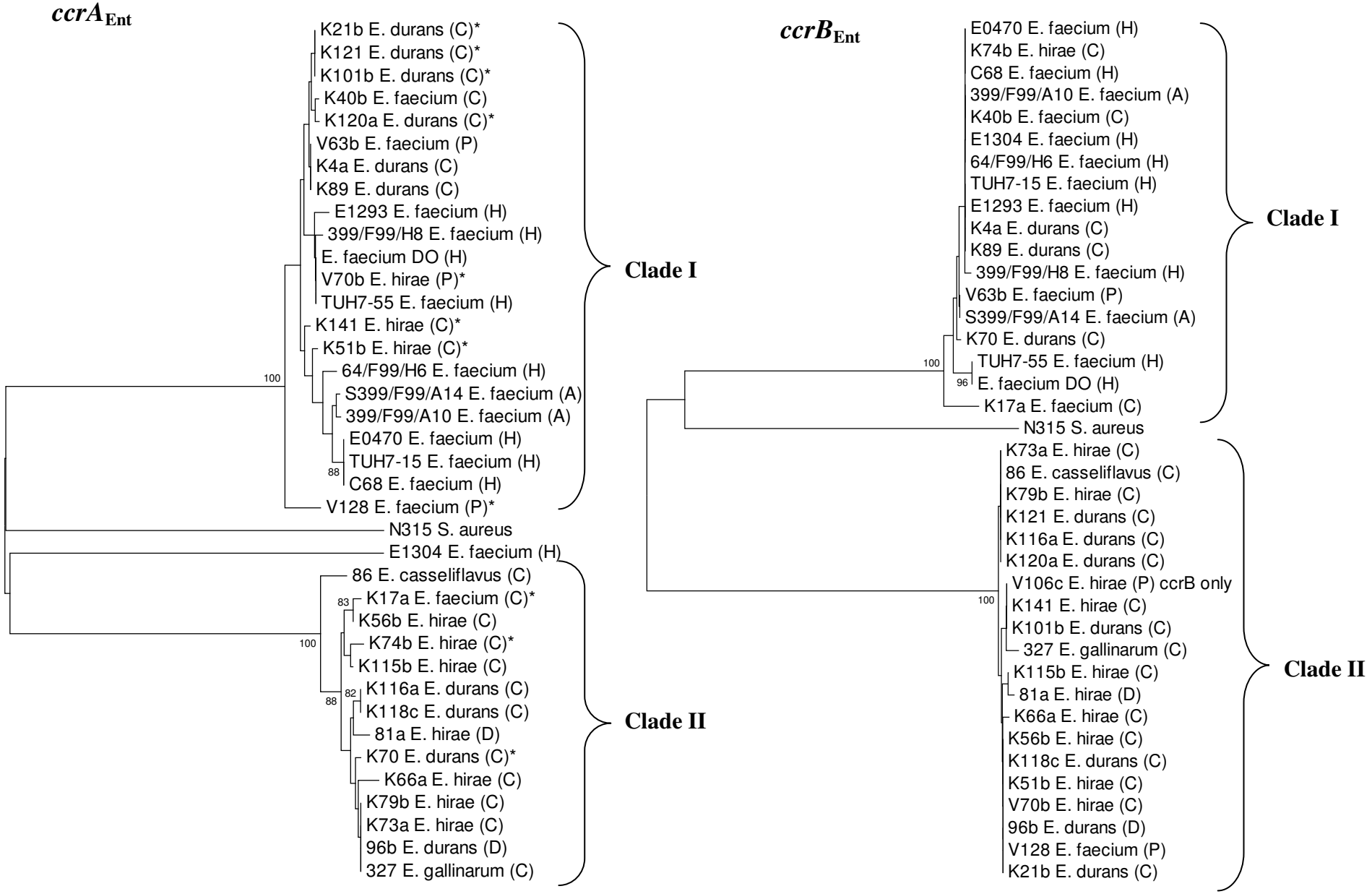


Fig. 2. Phylogram for $CcrA_{Ent}$, $CcrB_{Ent}$, other Ccrs, and three site-specific recombinase proteins. The deduced amino acid sequences of the following genes were used *ccrA1* and *ccrB1** (from NCTC10442 GenBank accession no. AB033763); *ccrA2* and *ccrB2* (from N315 GenBank accession no. D86934); *ccrA3* and *ccrB3* (from 85/2082 GenBank accession no. AB037671); *ccrA4* and *ccrB4** (from HDE288 GenBank accession no. AF411935); *ccrC* (from JSCC 3624 (WIS) GenBank accession no AB121219); site-specific integrase (from phi-FC1 GenBank accession no. AF124258), and two site-specific recombinases (from *C. acetobutylicum* ATCC824 GenBank accession no. AE007636 and AE007725). The scale indicates genetic distance in substitutions per site. The Ccr clusters are shown in circles. The amino acid sequences were aligned using T-Coffee. The neighbour-joining phylogenetic tree was constructed with MEGA3 from 2000 bootstrap replicates using the p-distance model. The scale indicates genetic distance in substitutions per site. The dataset consisted of 447 amino

acids with 37 parsimony-informative sites for CcrA and 547 amino acids with 50 parsimony-informative sites for CcrB.

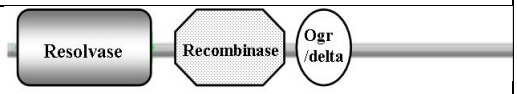

ccrA Ent
K21b E. durans (C)*
K121 E. durans (C)*
K101b E. durans (C)*
K40b E. faecium (C)
K120a E. durans (C)*
V63b E. faecium (P)
K4a E. durans (C)
K89 E. durans (C)
E1293 E. faecium (H)
399/F99/H8 E. faecium (H)
E. faecium DO (H)
V70b E. hirae (P)*
TUH7-55 E. faecium (H)
K141 E. hirae (C)*
K51b E. hirae (C)*
64/F99/H6 E. faecium (H)
S399/F99/A14 E. faecium (A)
399/F99/A10 E. faecium (A)
E0470 E. faecium (H)
TUH7-15 E. faecium (H)
C68 E. faecium (H)
V128 E. faecium (P)*
N315 S. aureus
E1304 E. faecium (H)
86 E. casseliflavus (C)
K17a E. faecium (C)*
K56b E. hirae (C)
K74b E. hirae (C)*
K115b E. hirae (C)
K116a E. durans (C)
K118c E. durans (C)
81a E. hirae (D)
K70 E. durans (C)*
K66a E. hirae (C)
K79b E. hirae (C)
K73a E. hirae (C)
96b E. durans (D)
327 E. gallinarum (C)
Clade I
Clade II
100
88
83
82
0.05
ccrB Ent
E0470 E. faecium (H)
K74b E. hirae (C)
C68 E. faecium (H)
399/F99/A10 E. faecium (A)
K40b E. faecium (C)
E1304 E. faecium (H)
64/F99/H6 E. faecium (H)
TUH7-15 E. faecium (H)
E1293 E. faecium (H)
K4a E. durans (C)
K89 E. durans (C)
399/F99/H8 E. faecium (H)
V63b E. faecium (P)
S399/F99/A14 E. faecium (A)
K70 E. durans (C)
TUH7-55 E. faecium (H)
E. faecium DO (H)
K17a E. faecium (C)
N315 S. aureus
K73a E. hirae (C)
86 E. casseliflavus (C)
K79b E. hirae (C)
K121 E. durans (C)
K116a E. durans (C)
K120a E. durans (C)
V106c E. hirae (P) ccrB only
K141 E. hirae (C)
K101b E. durans (C)
327 E. gallinarum (C)
K115b E. hirae (C)
81a E. hirae (D)
K66a E. hirae (C)
K56b E. hirae (C)
K118c E. durans (C)
K51b E. hirae (C)
V70b E. hirae (C)
96b E. durans (D)
V128 E. faecium (P)
K21b E. durans (C)
Clade I
Clade II
100
96
0.05

531

Fig. 3. Phylogram for *ccrA*$_{Ent}$ and *ccrB*$_{Ent}$ genes. The nucleotide sequences for *ccrA*$_{Ent}$ and *ccrB*$_{Ent}$ genes from *E. faecium* (n=15), *E. hirae* (n=11), *E. durans* (n=10), *E. casseliflavus* (n=1), and *E. gallinarum* (n=1) were used (GenBank accession no. FJ572967 - FJ573039). The capital letters represent the origin of the isolate: (C) chicken, (H) human, (D) dog, (P) pig, and (A) unknown animal origin. Outgroups are represented by *S. aureus* N315 *ccrA* and *ccrB* (GenBank accession no. D86934). The asterisks indicate isolates in which *ccrA*$_{Ent}$ and *ccrB*$_{Ent}$ belong to different clades. All sequences were aligned using Clustal W. The neighbour-joining phylogenetic tree was made with MEGA4.0 using bootstrap of 2000 replicates and the p-distance model. Bootstrap values higher than 80 % are shown at the branches. The scale indicates genetic distance in substitutions per site. The dataset consisted of 547 nts with 494 parsimony-informative sites for *ccrA*$_{Ent}$ and 513 nts with 227 parsimony-informative sites for *ccrB*$_{Ent}$. The two main clades of *ccrA*$_{Ent}$ and *ccrB*$_{Ent}$ are indicated.

SUPPLEMENTAL MATERIAL:

Table S1. Domains for Ccr_{Ent} proteins predicted in Pfam database. The resolvase domain is shown in grey/white, the recombinase in light grey and the Ogr/delta-like domain in white.

Ccr proteins	Domains	Start	End	E-value	Graphical view of domains
$CcrA_{Ent}$	Resolvase	20	153	1.2e-24	
	Recombinase	173	261	4e-20	
$CcrB_{Ent}$	Resolvase	9	158	6.4e-51	
	Recombinase	183	286	2.7e-31	
	Ogr/delta-like	305	360	4.4e-06	

Only Pfam-A domains are presented.

E-value (Expect value) is describing the number of hits one can "expect" to see by chance when searching the database. Start and end refer to the amino acid sequence.

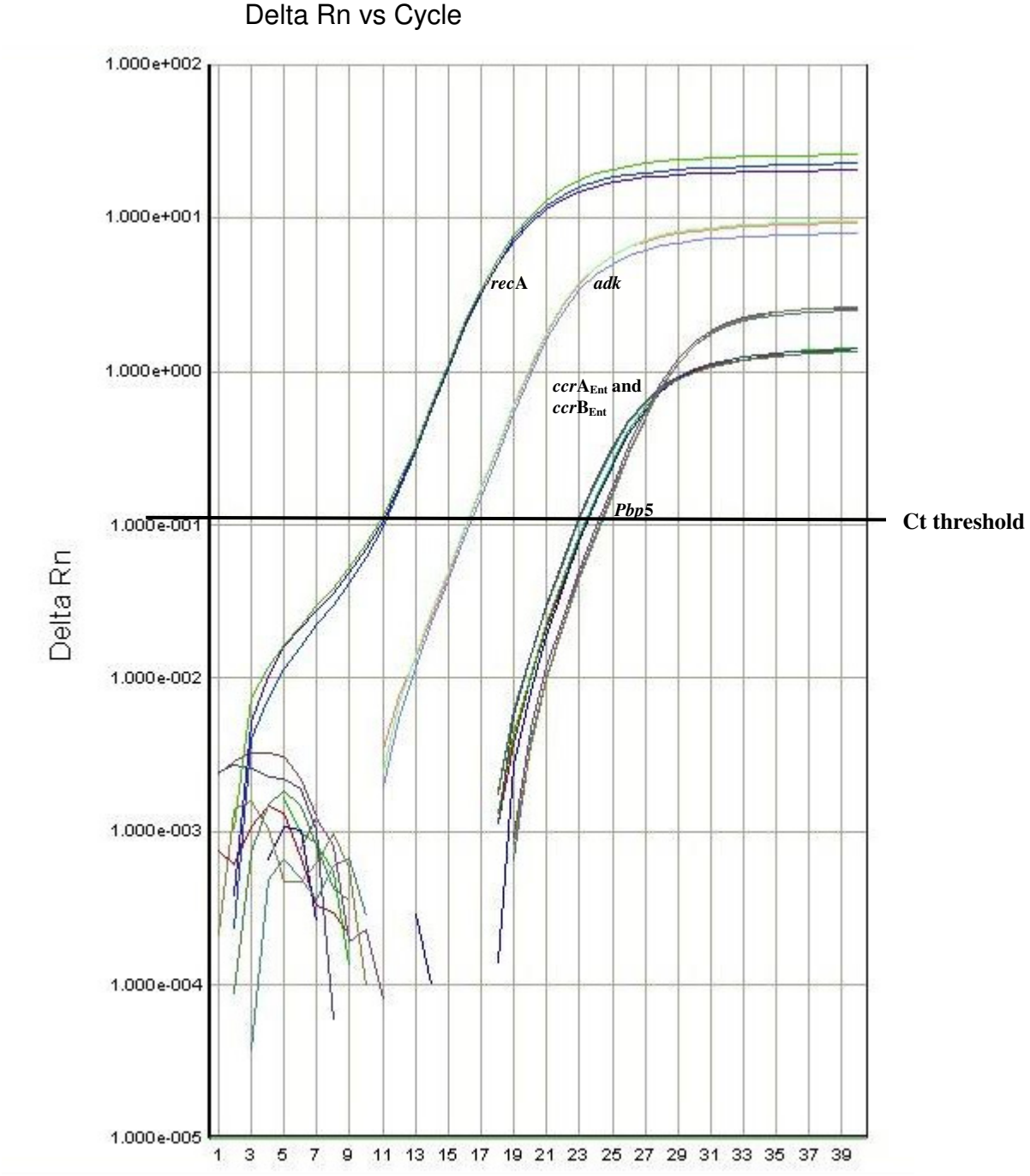


Fig. S1. Representative real-time PCR results for the expression of the genes *recA*, *adk*, $ccrA_{Ent}$, $ccrB_{Ent}$ and *pbp5*.