Comparative genomic analysis reveals significant enrichment of mobile genetic elements and genes encoding surface structure-proteins in hospital-associated clonal complex 2 *Enterococcus faecalis*

Margrete Solheim1*, Mari C Brekke1, Lars G Snipen2, Rob JL Willems3, Ingolf F Nes1, Dag A Brede1

**Abstract**

**Background:** Enterococci rank among the leading causes of nosocomial infections. The failure to identify pathogen-specific genes in *Enterococcus faecalis* has led to a hypothesis where the virulence of different strains may be linked to strain-specific genes, and where the combined endeavor of the different gene-sets result in the ability to cause infection. Population structure studies by multilocus sequence typing have defined distinct clonal complexes (CC) of *E. faecalis* enriched in hospitalized patients (CC2, CC9, CC28 and CC40).

**Results:** In the present study, we have used a comparative genomic approach to investigate gene content in 63 *E. faecalis* strains, with a special focus on CC2. Statistical analysis using Fisher’s exact test revealed 252 significantly enriched genes among CC2-strains. The majority of these genes were located within the previously defined mobile elements phage03 (n = 51), efaB5 (n = 34) and a vanB associated genomic island (n = 55). Moreover, a CC2-enriched genomic islet (EF3217 to -27), encoding a putative phage related element within the V583 genome, was identified. From the draft genomes of CC2-strains HH22 and TX0104, we also identified a CC2-enriched non-V583 locus associated with the *E. faecalis* pathogenicity island (PAI). Interestingly, surface related structures (including MSCRAMMs, internalin-like and WxL protein-coding genes) implicated in virulence were significantly overrepresented (9.1%; p = 0.036, Fisher’s exact test) among the CC2-enriched genes.

**Conclusion:** In conclusion, we have identified a set of genes with potential roles in adaptation or persistence in the hospital environment, and that might contribute to the ability of CC2 *E. faecalis* isolates to cause disease.

**Background**

For many years, *Enterococcus faecalis* was considered as an intestinal commensal, which only sporadically caused opportunistic infections in immunocompromised patients. During the last thirty years, however, *E. faecalis* has gained notoriety as one of the primary causative agents of nosocomial infections [1,2], including urinary tract infections, endocarditis, intra-abdominal infections and bacteremia. The ability of *E. faecalis* to cause infection has been connected to inherent enterococcal traits, enabling the bacterium to tolerate diverse and harsh growth conditions. Moreover, several putative enterococcal virulence factors have been characterized (reviewed in [3]), and the role of these virulence factors in pathogenicity have been further established in various animal infection models [4-8] and cultured cell lines [9,10]. Reportedly, several of the proposed virulence determinants are enriched among infection-derived *E. faecalis* and/or *E. faecium* isolates, including esp (enterococcal surface protein) [11], hyl (hyaluronidase) [12], genes encoding collagen binding adhesins [13,14] and other matrix-binding proteins [15], and pilin loci [16,17]. On the other hand, recent studies on enterococcal pathogenicity have shown that a number of the...
putative virulence traits are present not only in infectious isolates but also in animal and environmental isolates [18-23]. This widespread distribution of putative virulence determinants in enterococcal isolates strongly suggest that enterococcal pathogenicity is not a result of any single virulence factor, but rather a more intricate process. Indeed, the virulence potential of the newly sequenced laboratory strain \textit{E. faecalis} OG1RF was, despite its lack of several factors, comparable to that of the clinical isolate \textit{E. faecalis} V583 [24]. Bourgogne et al. [24] proposed a scenario where the virulence of V583 and OG1RF may be linked to genes that are unique to each of the two strains, but where the combined endeavor of the different gene-sets result in the ability to cause infection.

Population structure studies of \textit{E. faecalis} by multilocus sequence typing (MLST) have previously defined distinct clonal complexes (CC) of \textit{E. faecalis} enriched in hospitalized patients (CC2, CC9, CC28 and CC40), designated high-risk enterococcal clonal complexes (HiRECCs) [25,26]. In one of our previous studies, we reported an overall correlation between MLST and Bayesian phylogenetic analysis of gene content as revealed by microarray-based comparative genomic hybridization (CGH) [27]. This observation led us to speculate whether the virulence of different HiRECCs may be due to lineage-specific gene sets. In the present study we have used the comparative genomics approach to further investigate variation in gene content within \textit{E. faecalis}, with a special focus on CC2. This complex was chosen on the basis of previous Bayesian-based phylogenetic reconstruction [27]. CC2 is equivalent to the previously designated BVE complex, and comprises several clinically important \textit{E. faecalis} isolates, including the first known beta-lactamase producing isolate HH22, the first U.S. vancomycin-resistant isolate V583, and pathogenicity island (PAI)-harboring clinical bacteremia isolate MMH594 [26,28,29]. This CC represents a globally dispersed hospital-associated lineage, and identification of CC2-enriched genes may unravel novel fitness factors implicated in survival and spread of \textit{E. faecalis} clones in the hospital environment.

**Results and discussion**

**Overall genomic diversity**

To explore the genetic diversity among \textit{E. faecalis}, BLAST comparison was performed with 24 publicly available sequenced draft genomes, including the two CC2-strains TX0104 (ST2), which is an endocarditis isolate, and HH22 (ST6; mentioned above) against the genome of strain V583, which is also a ST6 isolate. The number of V583 genes predicted to be present varied between 2385 (OG1RF) and 2831 (HH22) for the 24 strains (Additional file 1). In addition, we used CGH to investigate variation in gene content within 15 \textit{E. faecalis} isolated in European hospital environments, with a special focus on a hospital-adapted subpopulation identified by MLST (CC2). Of the 3219 V583 genes represented on the array, the number of V583 orthologous genes classified as present ranged from 2359 (597/96) to 2883 (E4250). Analysis of the compiled data set (\textit{in silico} and CGH), revealed a total of 1667 genes present in all strains, thus representing the \textit{E. faecalis} core genome. None of the annotated V583 genes were found to be divergent in all the isolates analyzed.

**Putative CC2-enriched elements**

In a previous study, we identified a set of potential pathogen-specific genes, which were entirely divergent in a collection of commensal baby isolates [27]. None of these genes were found to be present in all hospital-related isolates analyzed in the present study, neither was any gene found to be unique to any HiRECC. In order to identify genes specifically enriched among strains belonging to CC2, data from the present study were supplemented with hybridization data from an additional 24 strains of various origins ([27,30] and M. Solheim, unpublished data). The additional data sets were obtained by hybridization to the same array as described above. All together, data from a total of 63 strains were analyzed, in addition to V583 (Table 1). A genome-atlas presentation of the gene content in all the strains analyzed by CGH compared to the V583 genome is shown in Figure 1.

By Fisher’s exact testing (\(q < 0.01\)), 252 genes were found to be more prevalent among CC2-strains than in non-CC2-strains (Additional file 2). The CC2-enriched genes included large parts of \textit{phage03} (\(p03; n = 51\), \textit{efaB5} (\(n = 34\)) and a phage-related region identified by McBride et al. [31](EF2240-82/EF2335-51; \(n = 55\)), supporting the notion that the \(p03\) genetic element may confer increased fitness in the hospital environment [27]. Indeed, prophage-related genes constituted a predominant proportion of the CC2-enriched genes (55.5%; \(p < 2.2e-16\), Fisher’s exact test). Interestingly, the Tn916-like \textit{efaB5} element has previously also been suggested to play a role in niche adaptation (Leavis, Williams et al. unpublished data): CGH analysis identified an \textit{efaB5}-orthologous element in \textit{E. faecium} that appeared to be common for HiRECC \textit{E. faecalis} and CC17 \textit{E. fae- cium}, a hospital-adapted subpopulation identified by MLST. To further confirm the presence of the relevant MGEs in \textit{E. faecalis}, we used PCR combining internal primers with primers targeting the genes flanking \(p03\), \textit{efaB5} and the \textit{vanB}-associated phage-related element in V583, to monitor conserved V583 junctions on either side of the elements in 44 strains (Table 1). Seven strains contained the junctions on both sides of \(p03\), of which six strains were CC2-strains. Eleven strains were positive for the junctions on both sides of \textit{efaB5},
Table 1 Enterococcus faecalis isolates used in this study. CC: clonal complex, CGH: comparative genomic hybridization, MLST: multilocus sequence typing, S: singleton, ST: sequence type

| Strain   | Year | Country   | Source          | MLST Application | Reference |
|----------|------|-----------|-----------------|-------------------|-----------|
| TX0104   |      | USA       | Clinical        | 2                 | [65]      |
| 609/96   | 1996 | Poland    | Clinical        | 2                 | [25]      |
| 372-56   | 2007 | Norway    | Blood           | 2                 | [67]      |
| 226B     | 2005 | Norway    | Feces           | 2                 | [27]      |
| 368-42   | 2007 | Norway    | Blood           | 2                 | [26]      |
| 442/05   | 2005 | Poland    | CSF             | 2                 | [25]      |
| E1828    | 2001 | Spain     | Blood           | 2                 | [66]      |
| MMH594   | 1985 | USA       | Clinical        | 2                 | [67]      |
| VS83     | 1989 | USA       | Blood           | 2                 | [27]      |
| 158B     | 2005 | Norway    | Feces           | 2                 | [27]      |
| HH22     | ≤1982| USA       | Urine           | 2                 | [27]      |
| E1834    | 2001 | Spain     | Feces           | 2                 | [26]      |
| E4250    | 2007 | Netherlands | Feces         | 2                 | [26]      |
| HIP11704 | 2002 | USA       | Clinical        | 2                 | [68]      |
| E1841    | 2001 | Spain     | Blood           | 2                 | [26]      |
| Vet179   | 1999 | Norway    | Dog_ureine      | 2                 | [69]      |
| CH188    | 1990 | USA       | Liver           | 2                 | [69]      |
| E1807    | 2002 | Spain     | Feces           | 2                 | [26]      |
| X98      | 1934 | USA       | Feces           | 2                 | [71]      |
| OG1RF    | ≤1975| USA       | Oral            | 2                 | [26]      |
| E1960    | 2001 | Spain     | Feces           | 2                 | [26]      |
| T8       | ≤1992| Japan     | Urine           | 2                 | [73]      |
| 2426/03  | 2003 | Poland    | Feces           | 2                 | [25]      |
| ATCC 29200| ≤1974| Canada   | Urogenital      | 2                 | [74]      |
| T1       | ≤1995| Japan     | Urine           | 2                 | [73]      |
| LMGT3406 | 1999 | Denmark   | Poultry_feces   | 2                 | [75]      |
| 111A     | 2005 | Norway    | Feces           | 2                 | [27]      |
| TX1322   | 2004 | USA       | Blood           | 2                 | [27]      |
| 3339/04  | 2004 | Poland    | Feces           | 2                 | [27]      |
| UC11/46  | 1999 | Finland   | Feces           | 2                 | [27]      |
| 189      | 2002-2003 | Norway | Feces           | 2                 | [27]      |
| Symbioflor 1 | 1980s| Germany   | Feces           | 2                 | [75]      |
| T2       | ≤1992| Japan     | Urine           | 2                 | [73]      |
| E1188    | 1997 | Greece    | Blood           | 2                 | [26]      |
| 383/04   | 2004 | Poland    | Blood           | 2                 | [25]      |
| E1052    | 2004 | Netherlands | Feces         | 2                 | [27]      |
| 85       | 2008 | Norway    | Feces           | 2                 | [27]      |
| 597/96   | 1996 | Poland    | Ulcer           | 2                 | [25]      |
| LMGT2333 | 1999 | Iceland   | Fish            | 2                 | [27]      |
| JH1      | ≤1974| United Kingdom | Clinical      | 2                 | [76]      |
| LMG13209 | 1999 | Greece    | Food_cheese     | 2                 | [26]      |
| 1645     | 2007 | Denmark   | Blood           | 2                 | [27]      |
| 29C      | 2004 | Norway    | Feces           | 2                 | [27]      |
| 92A      | 2005 | Norway    | Feces           | 2                 | [27]      |
| DSS      | ≤1974| Denmark   | Pig             | 2                 | [77]      |
| E2370    | 2002-2003 | Hungary | Wound           | 2                 | [27]      |
| 10S      | 2002-2003 | Norway | Feces           | 2                 | [27]      |
| D6       | 2002-2003 | Denmark | Pig             | 2                 | [31]      |
including nine CC2-strains, while thirteen strains gave positive PCR for both junctions of the phage-related element surrounding vanB, of which eleven strains belonged to CC2 (Additional file 3). These results substantiate the theory of p03, efaB5 and the vanB-associated phage as CC2-enriched elements.

A total of 178 of the 252 putative CC2-enriched genes identified here, were associated with previously defined MGEs identified in V583 [32]. In addition to p03, efaB5 and the vanB-surrounding phage element, these included p01 (n = 5), PAI (n = 7), p04 (n = 21), p06 (n = 1) and pTEF1 and pTEF2 (n = 5) (Additional file 2). In addition, a ten-gene cluster (EF3217 to -27) with significant GC skew compared to the genome-average (31.6 and 37.4%, respectively), was found to be significantly more frequent in strains belonging to CC2 than in non-CC2 strains. The deviation in GC content suggests that this genetic element may also be of foreign origin. This notion was further supported by the sequence similarities of several of the genes with known phage-related transcriptional regulators (EF3221, EF3223 and EF3227). Moreover, EF3221 to -27 showed high degree of identity (>85%) to EfmE980_2492 to -93 of the newly sequenced Enterococcus faecium E980 [33]. EfmE980_2492 holds a domain characteristic of the aspartate aminotransferase superfamily of pyridoxal phosphate-dependent enzymes. Interestingly, EF3217 encodes a putative helicase, while EF3218 encodes a putative MutT protein, both with implications in DNA repair [34,35]. A potential role of these genes in protection against oxidative DNA damage induced in the hospital environment and during infection is plausible. To further investigate the distribution of EF3217 to -27 in E. faecalis, 44 strains were screened by PCR (Additional file 3): 10 CC2-strains held all ten genes, while 19 strains including two CC2-strains were devoid of the entire element. Moreover, 2 strains contained EF3225 only, 3 strains contained EF3217 to -18, while 8 strains, including OG1RF, contained EF3226 only. The two latter patterns of presence and divergence of EF3217 to -27 were also obtained with BLASTN analysis of TX0104 and OG1RF, respectively, corroborating that these are indeed genuine polymorphisms in this locus. Notably, in the OG1RF genome five more genes (OG1RF_0214 to -18) are also located between the homologs of EF3216 and EF3230 [24], suggesting this locus may represent a hot spot for insertions. Partial sequencing across the junction between EF3216 and EF3230 suggested that several of the non-CC2 strains carry genes homologous to OG1RF_0214 to -18 in this locus (results not shown).

Mobile DNA constitutes a substantial fraction of the E. faecalis V583 genome and transfer of MGEs and transposons thus plays an important role in the evolution of E. faecalis genomes [32]. The large pool of mobile elements also represents an abundant source of pseudogenes, as indel events occurring within coding regions often render genes nonfunctional. To verify the

| Table 1 Enterococcus faecalis isolates used in this study. CC; clonal complex, CGH; comparative genomic hybridization, MLST; multilocus sequence typing, S; singleton, ST; sequence type (Continued) |
|---------------------------------|-----------------|-----------------|-----------------|
| E1Sol 1960s                     | Solomon Islands | Feces           | 93              | 93              | In silico | [78] |
| Merz96 2002                     | USA             | Blood           | 103             | 103             | In silico | [79] |
| R712                            | USA             | Clinical        | 103             | 103             | In silico | [65] |
| S613                            | USA             | Clinical        | 103             | 103             | In silico | [65] |
| LMGT3405 1999                   | Denmark         | Poultry_feces   | 116             | 116             | CGH2, PCR |
| LMGT3407 1999                   | Denmark         | Poultry_feces   | 34              | 121             | CGH2, PCR |
| Fly1 2005                       | USA             | Drosophila      | 101             | 1011^a          | In silico | [31] |
| Vet138 1998                     | Norway          | Dog_eth         | 164             | 119^a           | CGH2, PCR |
| 82                              | Norway          | Poultry_feces   | 65              | S               | CGH2, PCR |
| T11 ≤1992                      | Japan           | Urine           | 65              | S               | In silico | [73] |
| 62                              | 2002-2003 Norway | Feces           | 66              | S               | CGH2, PCR |
| ATCC 4200 1926                  | Blood           | 105             | S               | In silico       |
| AR01/DG 2001                    | New Zealand     | Dog             | 108             | S               | In silico | [80] |
| 266                             | Norway          | Feces           | 163             | S               | CGH2, PCR |
| LMGT3143 Spain                  | Animal_wood pigeon | 165           | S               | CGH2, PCR      |
| LMGT3208 2008                   | Greece          | Food_cheese     | 166             | S               | CGH2, PCR |
| 84                              | Norway          | Poultry_feces   | 249             | S               | CGH2, PCR |
| TuSoD ef11                      | USA             | Clinical        | 364             | S               | In silico | [65] |

^aClonal complexes were not predicted founder was proposed by eBURST.
^bIn Solheim et al. 2009.
^cIn Vebø et al. 2010.
^dMS, unpublished work.
Figure 1 Genome-atlas presentation of CGH data compared to the V583 genome and arranged by clonal relationship according to MLST. From inner to outer lanes: 1) percent AT, 2) GC skew, 3) global inverted repeats, 4) global direct repeats, 5) position preference, 6) stacking energy, 7) intrinsic curvature, 8) 189, 9) LMGT3208, 10) LMGT3407, 11) 92A, 12) 29C, 13) E1960, 14) 111A, 15) 105, 16) E2370, 17) 84, 18) 383/04, 19) E1188, 20) Vet179, 21) EF1841, 22) E1807, 23) LMGT3143, 24) LMGT3405, 25) OG1RF, 26) 2426/03, 27) LMGT3406, 28) 85, 29) E1052, 30) 1645, 31) LMGT3209, 32) LMGT2333, 33) 597/96, 34) 62, 35) Vet138, 36) 266, 37) UC11/96, 38) Symbioflor 1, 39) 3339/04, 40) 82, 41) E1834, 42) E4250, 43) LMGT3303, 44) 158B, 45) MMH594, 46) 372-56, 47) 609/96 and 48) annotations in V583. Elements enriched in CC2-strains are indicated with an asterisk.
expression of the CC2-enriched genes, we correlated the list of enriched genes with data from two transcriptional analyses performed in our laboratory with the same array as used in the CGH experiment described in present study ([30] and Solheim, unpublished work). Transcription was confirmed for all but fifteen of the CC2-enriched genes (results not shown), thus validating the expression of these reading frames. The fifteen genes, for which no transcripts were detected, were mainly located within efaB5 and phage04.

A constraint of the comparative genomic analyses presented here, is that the comparison of gene content is based on a single reference strain only (V583). To compensate, we conducted a CC2 pangenome analysis with the draft genomes of CC2-strains HH22 and TX0104 to identify putative CC2-enriched non-V583 genes. The pangenome analysis identified a total of 298 non-V583 ORFs in the HH22 and TX0104 (Additional file 4). Among these ORFs, one gene cluster was identified as particularly interesting (Fisher’s exact; Additional file 4 and Figure 2). Notably, HMPREF0348_0426 in TX0104 represented the best BLAST hit for all the three ORFs HMPREF0364_1864 to -66 in HH22, suggesting discrepancy in annotation between the two strains. Sequencing across the gap between contig 00034 and contig 00035 in TX0104 confirmed that HMPREF0348_0427 and HMPREF0348_0428 represent the two respective ends of a gene homologous to HMPREF0346_1863 in HH22. (Additional file 5). The presence of the putative non-V583 CC2-enriched gene cluster among E. faecalis was further elucidated by PCR in our collection of strains (Additional file 3). Strains were screened for the presence of three individual genes (HMPREF0346_1861, HMPREF0346_1864 and HMPREF0346_1868) and the entire element, with primers hmpref0346_1868-F and hmpref0346_1861-R. Fisher’s exact testing \((q < 0.01)\) on the basis of the PCR data confirmed that the gene cluster was significantly enriched among CC2. Comparative sequence analysis of the flanking regions suggests that the gene cluster is located in the HH22 and TX0104 versions of the E. faecalis pathogenicity island [36]. Recently, a microarray-based assessment of PAI-content in a set of clinical E. faecalis isolates revealed high degree of variation within the island, and an evidently modular evolution of the PAI [37], which would be consistent with acquisition by an indel event of this locus in the PAI of TX0104, HH22 and other positive CC2-strains.

CC2-enriched surface-related structures

Lepage et al. [38] have previously identified eight genes as potential markers for the V583/MMH594-lineage, of which all except one gene (EF2513) are found among the CC2-enriched genes in this study. Interestingly, several of these genes were later assigned to a recently classified family of surface proteins, with a C-terminal WxL domain, proposed to form multi-component complexes on the cell surface [39,40]. Siezen et al. [40] termed these genes cell-surface complex (csc) genes and postulated a role in carbon source acquisition. Independently, Brinster et al. [39] showed that WxL domains are involved in peptidoglycan-binding. A total of nine WxL protein-coding genes, divided into three clusters (EF2248 to -54, EF3153 to -55 and EF3248 to -53), were identified as putative CC2-enriched genes in the present study. Note that EF3153 to - 55 does not represent a complete csc gene cluster, as not all four csc gene families (cscA - cscD) are present in the cluster [40]. Interestingly, the OG1RF genome sequence revealed homologues loci encoding WxL-proteins corresponding to the gene clusters EF3153 to -55 and EF3248 to -53 in V583 (50-75% sequence identity) [24]. Such homologs may possibly explain the divergence observed between CC2 and non-CC2-strains in the present study. Indeed, BLAST analysis with the OG1RF sequences against the E. faecalis draft genomes suggested that the OG1RF_0209-10 and OG1RF_0224-25 are widely distributed among non-CC2 E. faecalis. Given the putative function in carbon metabolism, the observed sequence variation may be related to substrate specificity.

In addition to the WxL domain, EF2250 also encodes a domain characteristic for the internalin family [39]. Internalins are characterized by the presence of...
N-terminal leucine-rich repeats (LRRs). The best characterized bacterial LRR proteins are InlA and InlB from Listeria monocytogenes, known to trigger internalization by normally non-phagocytic cells [41]. Two internalin-like proteins were identified in E. faecalis V583 (EF2250 and elrA [EF2686]) [41,42]. Recently, Brinster et al. [42] presented evidence of that ElrA play a role in E. faecalis virulence, both in early intracellular survival in macrophages and by stimulating the host inflammatory response through IL-6 induction. Moreover, by quantitative real-time PCR Shepard and Gilmore [43] found that elrA was induced in E. faecalis MMH594 during exponential growth in serum and during both exponential and stationary growth in urine. Contradictory data have, however, been published for this and other strains using different methods [42,44]. Although it is tempting to speculate that EF2250 contributes to the interaction with the mammalian host, the role of internals in E. faecalis pathogenesis is still not understood, and it may therefore be premature to extrapolate function solely on the basis of shared structural domains.

Glycosyl transferase family proteins are involved in the formation of a number of cell surface structures such as glycolipids, glycoproteins and polysaccharides [45]. E. faecalis is in possession of several capsular polysaccharides [46-48], with Cps and Epa being the best characterized. The epa (entercoccal polysaccharide antigen) cluster represents a rhamnose-containing polysaccharide which was originally identified in E. faecalis OG1RF [46]. The version of the epa cluster found in the V583 genome contains an insertion of four genes (EF2185 to -88) compared to OG1RF. This insertion appeared to be enriched among CC2. While EF 2185 and EF2187 encodes transposases of the IS256 family, the two remaining genes showed 100% identity to the two respective ends of a racemase domain protein in E. faecalis TX0104. Neighboring the epa cluster, two glycosyl transferases (EF2170 and EF2167) proposed as potential virulence factors [32], are part of a three operon locus (EF2172 to -66), possibly associated with lipopolysaccharide production. Five of the genes within this locus were also found to be enriched among CC2 in the present study.

Paulsen et al. [32] also listed other putative surface-exposed virulence genes, including a choline-binding protein (CBP; EF2662) and a putative MSCRAMM (microbial surface components recognizing adhesive matrix molecules; EF2347) that based on our analysis were found to be enriched in CC2. A role of CBPs in pneumococcal colonization and virulence has been established [49,50]. A number of putative MSCRAMMs have been identified in E. faecalis [51], however, only Ace (adhesion of collagen from E. faecalis; EF1099) has been characterized in detail: Ace was shown to mediate binding to collagen (type I and IV), dentin and laminin [52-54]. Lebreton et al. [55] recently presented evidence of an in vivo function of Ace in enterococcal infections other than involvement in the interaction with extracellular matrix. It was demonstrated that an ace deletion mutant was significantly impaired in virulence, both in an insect model and in an in vivo–in vitro murine macrophage models. The authors suggested that Ace may promote E. faecalis phagocytosis and that it may also be possible that Ace is involved in survival of enterococci inside phagocytic cells. Also the structurally related MSCRAMM, Acn, found in E. faecium was recently reported to contribute to the pathogenesis of this bacterium [56].

Mucins are high molecular weight glycoproteins expressed by a wide variety of epithelial cells, including those of the gastrointestinal tract, and located at the interface between the cell and the surrounding environment [57]. The binding of bacteria to mucins through mucin-binding domain proteins is thought to promote colonization [58]. Diversity in the carbohydrate side chains creates a significant heterogeneity among mucins of different origin (e.g. different organisms or body sites), facilitating bacterial attachment to epithelial cells [58]. The non-V583 CC2-enriched gene cluster identified through in silico analysis in the present study harboured an ORF (HMPREF0346_1863 and HMPREF0348_0427/HMPREF0348_0428 in HH22 and TX0104, respectively) with homology to known mucin-binding domain proteins.

Conclusions

In conclusion, we have identified a set of genes that appear to be enriched among strains belonging to CC2. Since a significant proportion (9.1%; \( p = 0.036 \), Fisher’s exact test) of these genes code for proteins associated with cell surface structures, absence of divergence in these loci may lead to antigenic variation. Indeed, both MSCRAMMs and internals have been identified as potential antigens of E. faecalis or other Gram-positive bacteria [59-61]. It is noteworthy that the genes encoding any of the established enterococcal virulence factors were not among the CC2-enriched genes. Surface structures that promote adhesion of pathogenic bacteria to human tissue are also promising targets for creation of effective vaccines. However, functional studies of the individual CC2-enriched genes are required in order to distinguish their implications in enterococcal virulence.

Methods

Bacterial strain and growth conditions

Bacterial strains used in this study are listed in Table 1. E. faecalis strains were grown overnight (ON) in brain heart infusion broth (BHI; Oxoid) at 37° without shaking. All the strains have previously been sequence typed by the MLST scheme proposed by Ruiz-Garbajosa et al. [26].
Comparative genomic hybridization

**Microarrays**

The microarray used in this work has been described previously [27]. The microarray design has been deposited in the ArrayExpress database with the accession number A-MEXP-1069 and A-MEXP-1765.

**DNA isolation**

Genomic DNA was isolated by using the FP120 FastPrep bead-beater (BIO101/Savent) and the QiaPrep MiniPrep kit (Qiagen) as previously described [27].

**Fluorescent labeling and hybridization**

Fifteen hospital-associated *E. faecalis* strains were selected for CGH based on their representation of MLST sequence types (STs) belonging to major CCs and potential HiReCCs, with a special focus on CC2, and their variety of geographical origins within Europe. Genomic DNA was labeled and purified with the BioPrime Array CGH Genomic labeling System (Invitrogen) and Cyanine Smart Pack dUTP (PerkinElmer Life Sciences), according to the manufacturer’s protocol. Purified samples were then dried, prior to resuspension in 140 μl hybridization solution (5 × SSC, 0.1% (w/v) SDS, 1.0% (w/v) bovine serum albumin, 50% (v/v) formamide and 0.01% (w/v) single-stranded salmon sperm DNA) and hybridized for 16 h at 42°C to the *E. faecalis* oligonucleotide array in a Tecan HS 400 pro hybridization station (Tecan). Arrays were washed twice at 42°C with 2 × SSC + 0.2% SDS, and twice at 23°C with 2 × SSC, followed by washes at 23°C with 1) 0.2 × SSC and 2) H2O. Two replicate hybridizations (dye-swap) were performed for each test strain. Hybridized arrays were scanned at wavelengths of 532 nm (Cy3) and 635 nm (Cy5) with a Tecan scanner LS (Tecan). Fluorescent intensities and spot morphologies were analyzed using GenePix Pro 6.0 (Molecular Devices), and spots were excluded based on slide or morphology abnormalities. All water used for the various steps of the hybridization and for preparation of solutions was filtered (0.2 μM) MilliQ dH2O.

**Data analysis**

Standard methods in the LIMMA package [62] in R http://www.r-project.org/, available from the Bioconductor http://www.bioconductor.org were employed for preprocessing and normalization. Within-array normalization was first conducted by subtracting the median from the log-ratios for each array. A standard loess-normalization was then performed, where smoothing was based only on spots with abs(log-ratio) < 2.0 to avoid biases due to extreme skewness in the log-ratio distribution. For the determination of present and divergent genes a method that predicts sequence identity based on array signals was used, as described by Snipen et al. [63]. A threshold of 0.75 was used in order to obtain a categorical response of presence or divergence, i.e. genes with Sb-value > 0.75 were classified as present, while genes with Sb-value < 0.75 were classified as divergent. Genes with Sb-value = 0.75 remained unclassified. All genes were tested for significant enrichment among the CC2-strains by using the Fisher’s exact test.

**Microarray data accession number**

The microarray data have been deposited in the ArrayExpress database with the series accession number E-TABM-905.

**Polymerase chain reaction**

The presence of selected genes was verified by means of polymerase chain reactions (PCR). A similar approach was also applied to investigate the presence of selected mobile genetic elements (MGEs). Primers targeting the genes flanking the MGEs were combined with internal primers to monitor the presence of the junctions on either side of each MGE. PCR was carried out in 20 μl reaction volumes containing 1× buffer, 250 μM of each deoxynucleotide triphosphate and 1 U DyNAZyme II polymerase (Finnzymes). The reaction conditions included an initial denaturation step at 95°C and 35 cycles of 95°C for 30 s, 56-60°C for 30 s and 72°C for 1-5 min, followed by a final extension step at 72°C for 7 min. The primers used in this study are listed in Table 2.

**Validation of microarray data by sequencing**

Sequencing was performed using the ABI Prism Big dye Cycle Sequencing Ready Reaction kit (Applied Biosystems) in an ABI PrismTM 3100 Genetic Analyzer and primers listed in Table 2.

**In silico comparison of *E. faecalis* draft genomes**

Whole genome blast comparison against the V583 reference genome was conducted for 24 *E. faecalis* strains whose draft genomes were publicly available (GenBank accession numbers in parenthesis; Table 1): *E. faecalis* ARO1/DG (ACAK01000000); *E. faecalis* ATCC 4200 (ACAG01000000); *E. faecalis* ATCC 29200 (ACOX00000000); *E. faecalis* CH188 (ACAV01000000); *E. faecalis* D6 (ACAT01000000); *E. faecalis* DS5 (ACAI01000000); *E. faecalis* E1Sol (ACAQ01000000); *E. faecalis* Fly1 (ACAR01000000); *E. faecalis* H1Pl1704 (ACAN01000000); *E. faecalis* HH22 (ACIX00000000); *E. faecalis* JH1 (ACAP01000000); *E. faecalis* Mer296 (ACAM01000000); *E. faecalis* OG1RF (ABPI0100001); *E. faecalis* R712 (ADDQ00000000); *E. faecalis* S613 (ADDP00000000); *E. faecalis* T1 (ACAD01000000); *E. faecalis* T2 (ACAE01000000); *E. faecalis* T3 (ACAF01000000); *E. faecalis* T8 (ACOC01000000); *E. faecalis* T11 (ACAU01000000); *E. faecalis* TuSoD e11 (ACOX00000000); *E. faecalis* TX0104 (ACGL00000000);
E. faecalis TX1322 (ACGM00000000); E. faecalis X98 (ACAW01000000) [64,65], as follows: the annotated V583 genes were blasted (BLASTN) against each genome, and presence and divergence was predicted based on a score calculated as number of identical nucleotides divided by the length of the query gene. Genes obtaining a score >0.75 were predicted to be present.

Table 2 Primers used in this study

| Target gene | Primer sequences (5' → 3') | Amplicon size (bp) | Application |
|-------------|----------------------------|--------------------|-------------|
| ef1415      | F:GTGGCGGTTCGGTTCGATG      | 2818               | PCR on junction between EF1415 and EF1417 |
| ef1417      | RGCTCTGAGTATGCAATTCG       | 1252               | PCR on junction between EF2239 and EF2240 |
| ef1489      | F:GAATTCAAAATTTTGTGTTG     | 654                | PCR on junction between EF2350 and EF2352 |
| ef1895      | F:GCAACAACATTTCAATTTCCA    | 4573               | PCR on junction between EF1895 and EF1898 |
| ef1898      | RATGGAAGTGGTGCTGCTACGG     | -                  | Sequencing of junction between EF3216 and EF3230 |
| ef2239      | F:AATCTGCTGAGCGATGAGCA     | 499                | PCR         |
| ef2240      | F:ATACCTGCTGAGCGATGAGCA    | 654                | PCR on junction between EF2350 and EF2352 |
| ef2350      | F:TGTCGTGATCGTGTGTTG       | 4573               | PCR on junction between EF1895 and EF1898 |
| ef2352      | F:CAGGCGGTTCGGTTCGATG      | 2818               | PCR on junction between EF1415 and EF1417 |
| ef3216      | F:TCGCGGTTCGGTTCGATG      | 499                | PCR         |

**CC2 pangenome content analysis**

Among the newly released E. faecalis draft genomes were two CC2-strains; HH22 and TX0104. In order to extend the list of CC2-enriched genes beyond V583, we conducted a BLAST search using the annotated genes of these two strains as queries against the full genome sequences of the other draft genomes. Again, a cutoff of
75% identity to the query was used to distinguish present from divergent genes.

Additional material

Additional file 1: BLAST comparison of E. faecalis genomes. Data from BLAST comparison of 24 E. faecalis draft genomes with the annotated genes of strain V583.

Additional file 2: V583 genomes which were identified as significantly enriched among CC2-strains in the present study. A list of V583 genomes which were identified as significantly enriched among CC2-strains in the present study.

Additional file 3: PCR screening. An overview of results from PCR screening of a collection of E. faecalis isolates.

Additional file 4: Enrichment analysis of Cc6 non-V583 genes by Fisher's exact test. An overview of the presence non-V583 genes in 24 E. faecalis draft genomes Cc6 including data from enrichment analysis by Fisher's exact test.

Additional file 5: Amino acid alignment of HMPREF0346_1863 in Enterococcus faecalis HH22 and its homologue in E. faecalis TX0104. An amino acid alignment of HMPREF0346_1863 in Enterococcus faecalis HH22 and its homologue in E. faecalis TX0104.

Acknowledgements

This work was financially supported by the European Union 6th Framework Programme "Approaches to Control multi-resistant Enterococci: Studies on molecular ecology, horizontal gene transfer, fitness and prevention" (LSHE-CT-2007-037410). We gratefully acknowledge the following researchers for kindly providing strains to this study; Dr. Lars B. Jensen, Dr. Barbara E. Murray, Dr. Ewa Sadowsky, Dr. Arnfinn Sundsfjord and Dr. Atte von Wright. We also acknowledge Dr. David W. Ussery for contributing bioinformatic tools and assistance in construction of the genome-atlas and Halleger Bergum at The Norwegian Microarray Consortium for printing of the microarray slides. Finally, we acknowledge the tremendous genome sequencing efforts made by Dr. Michael S. Gilmore and coworkers at the Stephens Eye Research Institute and Harvard Medical School, the Broad Institute, and the Human Microbiome-project represented by Dr. Barbara E. Murray and co-workers at Baylor College of Medicine, Dr. George Weinstock and coworkers at Washington University, and Dr. S. Shrivastava and co-workers at the J. Craig Venter Institute.

Author details

1Laboratory of Microbial Gene Technology and Food Microbiology, Department of Chemistry, Biotechnology and Food Science, The Norwegian University of Life Sciences, N-1432 Ås, Norway. 2Section for Biostatistics, Department of Chemistry, Biotechnology and Food Science, The Norwegian University of Life Sciences, N-1432 Ås, Norway. 3Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, The Netherlands.

Authors' contributions

MS conceived and designed the study, carried out the experimental work, analyzed the data, and drafted the manuscript. MCB performed the experimental work and assisted in critical review of the manuscript. LS contributed analysis tools, performed the statistical and bioinformatic analyses and assisted in critical review of the manuscript. RALW conceived and designed the study, contributed material and assisted in critical review of the manuscript. IFN conceived the study, contributed material and assisted in critical review of the manuscript. DAB participated in the design and coordination of the study, performed bioinformatic analysis and helped to draft the manuscript. All authors read and approved the final manuscript.

Received: 1 September 2010 Accepted: 4 January 2011 Published: 4 January 2011

References

1. Richards MJ, Edwards JR, Culver DH, Gaynes RP: Nosocomial infections in combined medical-surgical intensive care units in the United States. Infect Control Hosp Epidemiol 2000, 21(8):510-515.
2. Wispelhoff H, Bischoff T, Tallent SA, Seifert H, Veresol RP, Edmond MB: Nosocomial bloodstream infections in US hospitals: analysis of 24,179 cases from a prospective nationwide surveillance study. Clin Infect Dis 2004, 39(3):309-317.
3. Hancock LE, Gilmore MS: Pathogenicity of enterococci. In Gram-positive pathogens. Edited by: Frischetti VA, Novick RP, Ferretti JJ, Portnoy DA, Rood JJ. Washington DC: ASM Press; 2006:299-311.
4. Shankar N, Lockatell CV, Baghdayan AS, Drachenberg C, Gilmore MS, Johnson DE: Role of Enterococcus faecalis surface protein Esp in the pathogenesis of ascending urinary tract infection. Infect Immun 2001, 69(7):4366-4372.
5. Chow JW, Thal LA, Perri MB, Vazquez JA, Donabedian SM, Clewell DB, Zenovs MI: Plasmid-associated hemolysin and aggregation substance production contribute to virulence in experimental enterococcal endocarditis. Antimicrob Agents Chemother 1993, 37(11):2474-2477.
6. Jett BD, Jensen HS, Nordquist RE, Gilmore MS: Contribution of the pad1-encoded cytolsin to the severity of experimental Enterococcus faecalis endophthalmitis. Infect Immun 1992, 60(6):2445-2452.
7. Schlievert PM, Gahr PJ, Assimacopoulos AR, Dinges MM, Stooehr JA, Harmala JW, Hirt H, Dunny GM: Aggregation and binding substances enhance pathogenicity in rabbit models of Enterococcus faecalis endocarditis. Infect Immun 1998, 66(1):218-223.
8. Singh KV, Nallapareddy SR, Sillanpaa J, Murray BE: Importance of the collagen adhesin ace in pathogenesis and protection against Enterococcus faecalis experimental endocarditis. PLoS Pathog 6(1):e1000716.
9. Kreft B, Marre R, Schramm U, Wirth R: Aggregation substance of Enterococcus faecalis mediates adhesion to cultured renal tubular cells. Infect Immun 1992, 60(1):25-30.
10. Olmsted SB, Dunny GM, Erlandsen SL, Wells CL: A plasmid-encoded surface protein on Enterococcus faecalis augments its internalization by cultured intestinal epithelial cells. J Infect Dis 1994, 170(6):1549-1556.
11. Shankar V, Baghdadyan AS, Huycke MM, Lindahl G, Gilmore MS: Infection- derived Enterococcus faecalis strains are enriched in esp, a gene encoding a novel surface protein. Infect Immun 1999, 67(1):193-200.
12. Rice LB, Caras L, Rudin S, Vael C, Goossens H, Konstabel C, Klare I, Hendrickx AP, Bonten MJ, van Luit-Asbroek M, Schapendonk CM, Kragten AH, Willems RJ: Distribution of genes encoding MSCRAMMs and Pili in clinical and natural populations of Enterococcus faecium. J Clin Microbiol 2009, 47(4):1203-1211.
13. Nallapareddy SR, Huang W, Murray BE: A potential virulence gene, hjfF, predominates in Enterococcus faecium of clinical origin. J Infect Dis 2003, 187(3):508-512.
14. Sillanpaa J, Nallapareddy SR, Prakash VP, Qin X, Hook M, Weinstock GM, Murray BE: Identification and phenotypic characterization of a second collagen adhesin, Scm, and genome-based identification and analysis of 13 other predicted MSCRAMMs, including four distinct pilus loci, in Enterococcus faecium. Microbiology 2008, 154(Pt 10):3199-3211.
15. Hendrickx AP, van Luit-Ashbroek M, Schapendonk CM, van Wamel WJ, Braat JC, Wijnands LM, Bonten MJ, Willems RJJ: SgrA, a nipecotic-binding LPXTG surface adhesin implicated in biofilm formation, and EcbA, a collagen binding MSCRAMM, are two novel adhesins of hospital-acquired Enterococcus faecium. Infect Immun 2009, 77(11):5097-5106.
16. Hendrickx AP, Bonten MJ, van Luit-Ashbroek M, Schapendonk CM, Kragten AH, Willems RJJ: Expression of two distinct types of pili by a hospital-acquired Enterococcus faecium isolate. Microbiology 2008, 154(Pt 10):3212-3223.
17. Sillanpaa J, Prakash VP, Nallapareddy SR, Murray BE: Distribution of genes encoding MSCRAMMs and Pili in clinical and natural populations of Enterococcus faecium. J Clin Microbiol 2009, 47(4):896-901.
18. Eaton TJ, Gasson MJ: Molecular screening of Enterococcus virulence determinants and potential for genetic exchange between food and medical isolates. Appl Environ Microbiol 2001, 67(4):1626-1635.
19. Lempainen H, Kinnunen K, Mertanen A, van Wright A: Occurrence of virulence factors among human intestinal enterococcal isolates. Lett Appl Microbiol 2005, 41(4):341-344.
20. Semedo T, Santos MA, Lopes MF, Figueiredo Marques JJ, Barreto Crespo MT, Tenereo R: Virulence factors in food, clinical and reference Enterococci: A common trait in the genus? J Appl Microbiol 2003, 94(1):13-22.

21. Cerri R, Impoli M, Bertuccini L, Fabietti F, Orephi G, Di Rosa R, Baldassarri L: Survey for virulence determinants among Enterococcus faecalis isolated from different sources. J Med Microbiol 2004, 53(Pt 1):13-20.

22. Franz CM, Muscholl-Silberhorn AB, Youil NM, Vancanteyn M, Swings J, Holzapfel WH: Incidence of virulence factors and antibiotic resistance among Enterococci isolated from food. Appl Environ Microbiol 2001, 67(9):3885-3890.

23. Mannu L, Paba A, Daga E, Cemiruan R, Zanetti S, Dupre I, Sechi LA: Comparison of the incidence of virulence determinants and antibiotic resistance between Enterococcus faecium strains of dairy, animal and clinical origin. Int J Food Microbiol 2003, 88(3):291-304.

24. Bourgogne A, Garsin DA, Qin X, Singh KV, Sillanpaa J, Yerrapragada S, Figueiredo Marques JJ, Barreto ME, et al: Large scale variation in Enterococcus faecalis illustrated by the genome analysis of strain OGI1RF. Genome Biol 2008, 9(7):R110.

25. Kawalec M, Pietras Z, Danilowicz E, Jakubczak A, Gniadkowski M, Faucett C, Furlani S, Serror P, Comparini D: Comparative expression of virulence-related genes in Enterococcus faecalis in response to biological cues in urine and urine. Infect Immun 2002, 70(8):4344-4352.

26. Vebe HC, Snipen L, Nes IF, Brede DA: The transcriptome of the nosocomial pathogen Enterococcus faecalis VFs83 reveals adaptive responses to growth in blood. PLoS One 2009, 4(11):e8660.

27. Paulussen JC, Calley KJ: Glycosyltransferases. Structure, localization, and control of cell type-specific glycosylation. J Biol Chem 1989, 264(30):17615-17618.

28. Nallapareddy SR, Wenxiang H, Weinstock GM: A cluster of genes involved in polysaccharide biosynthesis from Enterococcus faecalis OGI1RF. Infect Immun 1998, 66(9):4313-4323.

29. Frandsen LG, Gille MS: The binding of capsular polysaccharide of Enterococcus faecalis and its relationship to other polysaccharides in the cell wall. Proc Natl Acad Sci USA 2002, 99(1):1574-1579.

30. Hansson LE, Gilmore MS: Contribution of the collagen adhesin, Ace, mediates attachment to particulate matter. J Bacteriol 2009, 191(9):2832-2839.

31. Hall AE, Gorovits EL, Syribeys PJ, Domanski PJ, Ames BR, Chang CY, Vernachio JH, Patti JM, Hutchins JT: Their Transcriptional Responses to Growth in Human Urine. PLoS ONE 2010, 5(8):e12489.

32. Venkateswarlu V, Fung AL, Chan CK, Jordan WA: Enterococcal surface protein A and Evidence for Production of Ace during Human Infections. J Bacteriol 2000, 182(13):4385-4393.

33. Clasen S, Grubele M, Iriarte LA, Brunkow ME, et al: Multicopy sequence typing scheme for Enterococcus faecalis reveals hospital-adapted genetic complexes in a background of high rates of recombination. J Clin Microbiol 2006, 44(6):2220-2228.

34. Solheim M, Aakra A, Snipen LG, Brede DA, Nes IF: Comparative genomics of Enterococcus faecalis from healthy Norwegian infants. BMC Genomics 2007, 8:1244-1253.

35. Biyikli A, Bozkurt HO, Yildirim O, Dogru M, et al: Comparative genomics and genetic analysis of vancomycin-resistant Enterococcus faecalis. J Clin Microbiol 2007, 45(1):147-153.

36. Tuteja N, Tuteja R: Pyrosequencing-based comparative genome analysis of the nosocomial pathogen Enterococcus faecalis and identification of a large transferable pathogenicity island. BMC Genomics 2009, 10:194.

37. Beuneu C, Wang Y, Krueger WA, Maffod LC, Martinezos G, Boisot S, Goldmann DA, Kasper DL, Tzanabas AO, Pier GB: Isolation and Chemical Characterization of a Capsular Polysaccharide Antigen Shared by Clinical Isolates of Enterococcus faecalis and Vancomycin-Resistant Enterococcus faecium. Infect Immun 1999, 67(2):1213-1219.

38. Gosink KK, Mann ER, Guglielmo C, Tuomanen E, Masur H: Role of novel choline binding proteins in virulence of Streptococcus pneumoniae. Infect Immun 2000, 68(10):5690-5695.

39. Kugler JS, Henschen C, Wieser JN, Johnson S, Fontan P, Orntoft T, Sellebjerg F: Characterization of novel choline-binding proteins to adherence, colonization and immunogenicity of Streptococcus pneumoniae. Mol Microbiol 1997, 25(5):819-829.

40. Fronczek LF, Gilmore MS: Binding of the capsular polysaccharide of Enterococcus faecalis and its relationship to other polysaccharides in the cell wall. Proc Natl Acad Sci USA 2002, 99(1):1574-1579.

41. Vebe HC, Solheim M, Snipen L, Nes IF, Brede DA: Comparative Genomic Analysis of Pathogenic and Probiotic Enterococcus faecalis Isolates, and Their Transcriptional Responses to Growth in Human Urine. PLoS ONE 2010, 5(8):e12489.

42. McBride SM, Fischetti VA, Leblanc DJ, Moelling CR Jr, Gilmore MS: Genetic diversity among Enterococcus faecalis. PLoS ONE 2007, 2(7):e582.

43. Paulsen IT, Banerjee L, Myers GS, Nelson KE, Sehadjri R, Read TD, Fouts DE, Eisen JA, Gill SR, Hedberg-Castillo JF, et al: Role of mobile DNA in the evolution of vancomycin-resistant Enterococcus faecalis. Science 2003, 300(5621):2071-2074.

44. van Schalk W, Tan J, Caporale J, Van Den Abbeele T, van den Ende J, Schapendonck CM, Hendrickx AP, Nijman IJ, Bonten MJ, Tettelin H, et al: Pyrosequencing-based comparative genome analysis of the nosocomial pathogen Enterococcus faecalis and identification of a large transferable pathogenicity island. BMC Genomics 2011, 12:399.

45. Lenaerts SR, Snipen L, Murray BE, Hook M: A family of putative MSCRAMMs from Enterococcus faecalis. Microbiology 2004, 150(Pt 7):2053-2063.

46. Kong Y, Fang Z, Sun J, Xu Y: Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of Streptococcus pneumoniae. Mol Microbiol 1997, 25(5):819-829.

47. Vanoord WP, Evertsen RC, Boekhorst J, Muscariello L, Molenaar D, Renckens B, Sillanpaa J, Xu Y, Caspersen R, et al: Comparative genomics of Enterococcus faecalis OGI1RF and evidence for production of Ace during Human Infections. Infect Immun 2000, 68(9):4313-4323.

48. Mahapatra R, Sanyal R, Sanyal S, Sanyal A, et al: Comparative genomics of Enterococcus faecalis OGI1RF and evidence for production of Ace during Human Infections. Infect Immun 2000, 68(9):4313-4323.

49. Kondorosi A, Kondorosi E: Internalins: a complex family of versatile, widely distributed, housecleaning adhesin, Ace, mediates attachment to particulate matter. J Bacteriol 2006, 188(1):25059-25062.

50. Cui L, Guo F, Jiang D, Wang Y, et al: The MuST proteins or “Nudix” hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes. J Biol Chem 1996, 271(24):14598-14604.

51. Tuteja N, Tuteja R: Prokaryotic and eukaryotic DNA helicases. Essential motor molecular motors for cellular machinery. Eur J Biochem 2004, 271(10):1835-1846.

52. Shanks N, Baghdadjan A, Gilmore MS: Modulation of virulence within a pathogenicity island in vancomycin-resistant Enterococcus faecalis. Nature 2002, 417(6860):746-750.

53. McBride SM, Coburn PS, Baghdadjan A, Willemes RJ, Grande MJ, Shankar N, Gilmore MS: Genetic variation and evolution of the pathogenicity island of Enterococcus faecalis. J Bacteriol 2009, 191(10):3392-3402.

54. Lepage E, Brinster S, Caron C, Ducros-Crepy C, Rigaudier-Gois L, Dunny G, Henequen-Quertier C, Serror P: Comparative genomic hybridization analysis of Enterococcus faecalis: identification of genes absent from food strains. J Bacteriol 2006, 188(19):6648-6688.

55. Brinster S, Furlan S, Serror P: C-terminal WxI domain mediates cell wall binding in Enterococcus faecalis and other gram-positive bacteria. J Bacteriol 2007, 189(4):1244-1253.
61. Yu WL, Dan H, Lin M: InlA and InlC2 of Listeria monocytogenes serotype 4b are two internalin proteins eliciting humoral immune responses common to listerial infection of various host species. Curr Microbiol 2008, 56(5):506-509.
62. Smyth GA, Speed T: Normalization of cDNA microarray data. Methods 2003, 31(4):265-273.
63. Snipe L, Nyquist OL, Solheim M, Aakre A, Nes IF: Improved analysis of bacterial CGH data beyond the log-ratio paradigm. BMC Bioinformatics 2009, 10(1):91.
64. Palmer KL, Carnio K, Manson JM, Heiman D, Shea T, Young S, Zeng Q, Gevers D, Feldgarden M, Biren B, et al: High Quality Draft Genome Sequences of 28 Enterococcus sp. isolates. J Bacteriol., JB.00153-00110.
65. Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA, Deal C, et al: The NIH Human Microbiome Project. Genome Res 2009, 19(12):2317-2323.
66. Hoyeke MM, Spiegel CA, Gilmore MS: Bacteremia caused by hemolytic, high-level gentamicin-resistant Enterococcus faecalis. Antimicrob Agents Chemother 1991, 35(8):1626-1634.
67. Sahm DF, Kissinger J, Gilmore MS, Murray PR, Mulder R, Soliday J, Clarke B: In vitro susceptibility studies of vancomycin-resistant Enterococcus faecalis. Antimicrob Agents Chemother 1989, 33(9):1588-1591.
68. Moellinger RC Jr, Weinberg AN: Studies on antibiotic synergism against enterococci. II. Effect of various antibiotics on the uptake of 14C-labeled streptomycin by enterococci. J Clin Invest 1971, 50(12):2580-2584.
69. Aakre A, Nyquist OL, Snipe L, Reiersen TS, Nes IF: Survey of genomic diversity among Enterococcus faecalis strains by microarray-based comparative genomic hybridization. Appl Environ Microbiol 2007, 73(7):2207-2217.
70. Rice LB, Elipoulos GM, Wenersten C, Goldmann D, Jacoby GA, Moellinger RC Jr: Chromosomally mediated beta-lactamase production and gentamicin resistance in Enterococcus faecalis. Antimicrob Agents Chemother 1991, 35(2):272-276.
71. Wheeler SM, Foley GE: Studies on the Streptococci (Enterococci) of Lancefield Group-D.2. Recovery of Lancefield Group D Streptococci from Antemortem and Postmortem Cultures from Infants and Young Children, American Journal of Diseases of Children 1945, 70(4):207-213.
72. Murray BE, Singh KV, Ross RP, Heath JD, Dunny GM, Weinstock GM: Generation of restriction map of Enterococcus faecalis strains by microarray-based comparative genomic hybridization. J Bacteriol 1993, 175(16):5216-5223.
73. Maekawa S, Yoshida M, Kumamoto Y: Proposal of a new scheme for the serological typing of Enterococcus faecalis strains. Microbiol Immunol 1992, 36(7):671-681.
74. Ackermann HW, Caprioli T, Kasatiya SS: A large new Streptococcus bacteriophage. Can J Microbiol 1975, 21(4):571-574.
75. Domann E, Hain T, Ghsi R, Billon A, Kuenne C, Zimmermann K, Chakraborty T: Comparative genomic analysis for the presence of potential enterococcal virulence factors in the probiotic Enterococcus faecalis strain Symbioflor 1. Int Med Microbiol 2007, 297(7-8):533-539.
76. Jacob AE, Hobbs SJ: Conjugal transfer of plasmid-borne multiple antibiotic resistance in Streptococcus faecalis var. zymogenes. J Bacteriol 1974, 117(2):360-372.
77. Clewell DB, Yagi Y, Dunny GM, Schutz SK: Characterization of three plasmid deoxyribonucleic acid molecules in a strain of Streptococcus faecalis: identification of a plasmid determining erythromycin resistance. J Bacteriol 1974, 117(1):283-289.
78. Gardner P, Smith DH, Beer H, Moellinger RC Jr: Recovery of resistance (R) factors from a drug-free community. Lancet 1969, 2(7624):774-776.
79. Harrington SM, Ross TL, Gbebo KA, Metz WG: Vancomycin resistance, esp, and strain relatedness: a 1-year study of enterococcal bacteremia. J Clin Microbiol 2004, 42(12):5895-5898.
80. Manson JM, Keis S, Smith JM, Cook GM: Characterization of a vancomycin-resistant Enterococcus faecalis (VREF) isolate from a dog with mastitis: further evidence of a clonal lineage of VREF in New Zealand. J Clin Microbiol 2003, 41(7):3331-3333.

Cite this article as: Solheim et al.: Comparative genomic analysis reveals significant enrichment of mobile genetic elements and genes encoding surface structure-proteins in hospital-associated clonal complex 2 Enterococcus faecalis, BMC Microbiology 2011 11:3.

Submit your next manuscript to BioMed Central and take full advantage of:

• Convenient online submission
• Thorough peer review
• No space constraints or color figure charges
• Immediate publication on acceptance
• Inclusion in PubMed, CAS, Scopus and Google Scholar
• Research which is freely available for redistribution