Supplemental Materials

AR and VF genes or loci for Table 3.

The following genes were used for Table 1. Target *E. faecium* VF genes included *acm, bee1, bee2, bee3, cad, cbh3, ebpA, ebpB, ebpC = pilB, ebpR, ecbA, efmA, esp, fms21 = pilA, gelE, hemolysin, hyl, IS16, IS256, psaA, sagA, scm, sgrA, srt1*, and *srt2*. Target *E. faecalis* genes included *ace, agg = asa, asa373, bee1, bee2, bee3, cad, cbh, ccf, cob, cpd, cpsA, cylA, cylB, cylLl, cylLs, cylM, cylR1, cylR2, ebpA, ebpB, ebpC, ebpR, ef03341, ef0591, efaS, elrA, epaA, epaB, epaC, epaD, epaE, epaF, epaG, epaH, epaI, epaJ, epaK, epaL, epaM, epaN, epaO, epaP, epaQ, epaR, etaR, etaS, fsrA, fsrB, fsrC, fsrD, gelE, gls24-like, hydrolase, hypR, IS256, Nuc-1, perR, psaA, salB, sprE, srt1*, and *srt2*. *E. faecium* AR genes included *aac6Ie, aad9Ib, aph3IIIa, catA, ermA, ermB, ermG, FosB, lnuB, mefA, str, tetL, tetM, tetU, vanA* loci, *vanB* loci, and *vatD*. *E. faecalis* AR genes included *aac6Ie, aph3IIIa, bacA, bl2a_pc, catA, emeA, ermB, isa, lnuB, str, tetL, tetM, tetS, vanA* loci, and *vanB* loci.

Table S1. PCR primers for detection of group-specific genes of *E. faecium*

Primers	Sequences (5' - 3')	Product Sizes	Targets
Efm522_F1	GCAAAGAGGAATCGGATCTT	266 bp	*E. faecium* ortholog_522
Efm522_R1	CAGTTGCCAAGTAAAGCCAA		
Efm3722_F2	TCGAAATCCGAAATCAATCA	261 bp	*E. faecium* ortholog_3722
Efm3722_R2	TCCTCAAATTGTTCCGTGTC		
Efm2421_F2	GCATAAACCTGCTTCGACAA	288 bp	*E. faecium* ortholog_2421
Efm2421_R2	TCCTCATCAAGAGCGTCATC		
Efm4275_F1	TCACAATGATCGGACATTCA	170 bp	*E. faecium* ortholog_4275
Efm4275_R1	TAATGCATCTCCCTCGACAA		
Efm6476_F1	TTTGGTGTTTGTGCTGGATT	227 bp	*E. faecium* ortholog_6476
Efm6476_R1	TCCATTTCTTCCGAGAACG		

Table S2. CDS collection comparison between *E. faecium* and *E. faecalis*.

Species	*E. faecium*	*E. faecalis*
Number of orthologous CDS	7,017	8,032
CDS with a gene symbol	1,133 (16.15%)	1,724 (21.46 %)
CDS without a gene symbol	5,884 (83.85 %)	6,308 (78.54 %)
CDS Length (bp, mean ± SD) [a]	762.85 ± 625.15	774.11 ± 673.32
Range of CDS Length (bp)	69 to 9,522	37 to 8,331
CDS G+C content (%, mean ± SD) [a]	36.19 ± 5.15	35.17 ± 5.59 [b]

[a] SD = standard deviation

[b] Significantly different between *E. faecium* and *E. faecalis* ($P < 0.001$, Student's t-test).

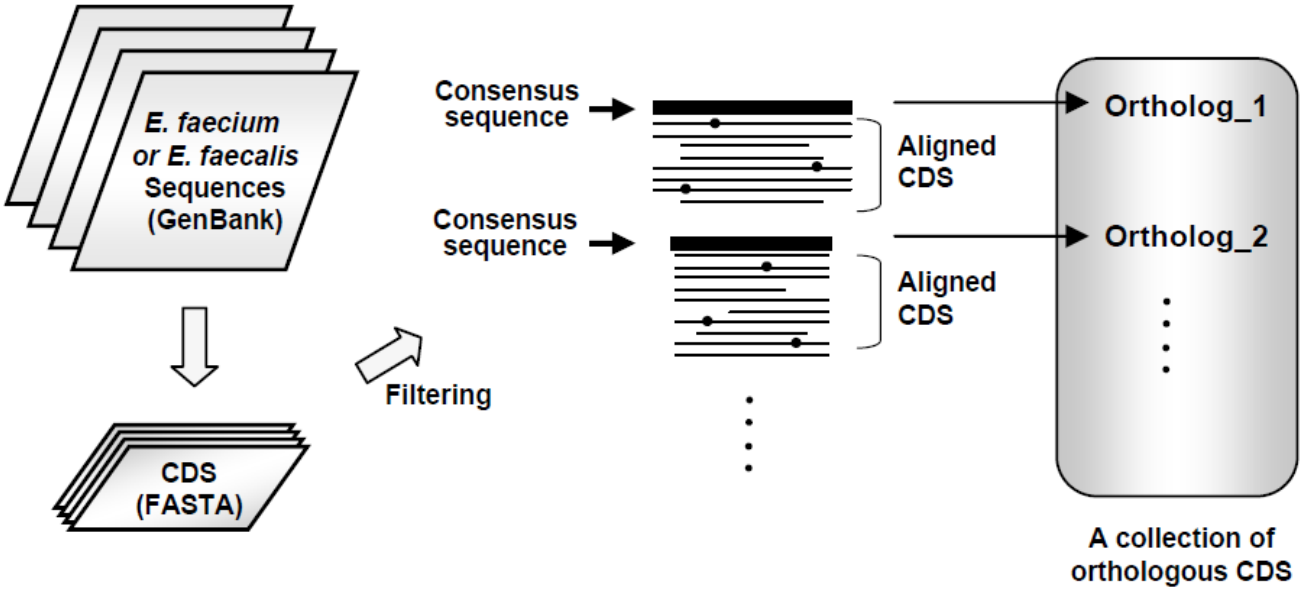

Figure S1. Construction of *E. faecium* and *E. faecalis* orthologous CDS collections. Each consensus sequence (thick line) was determined using the aligned CDS. Some single nucleotide polymorphisms (SNPs, black dots within the lines) were found among the aligned CDS.

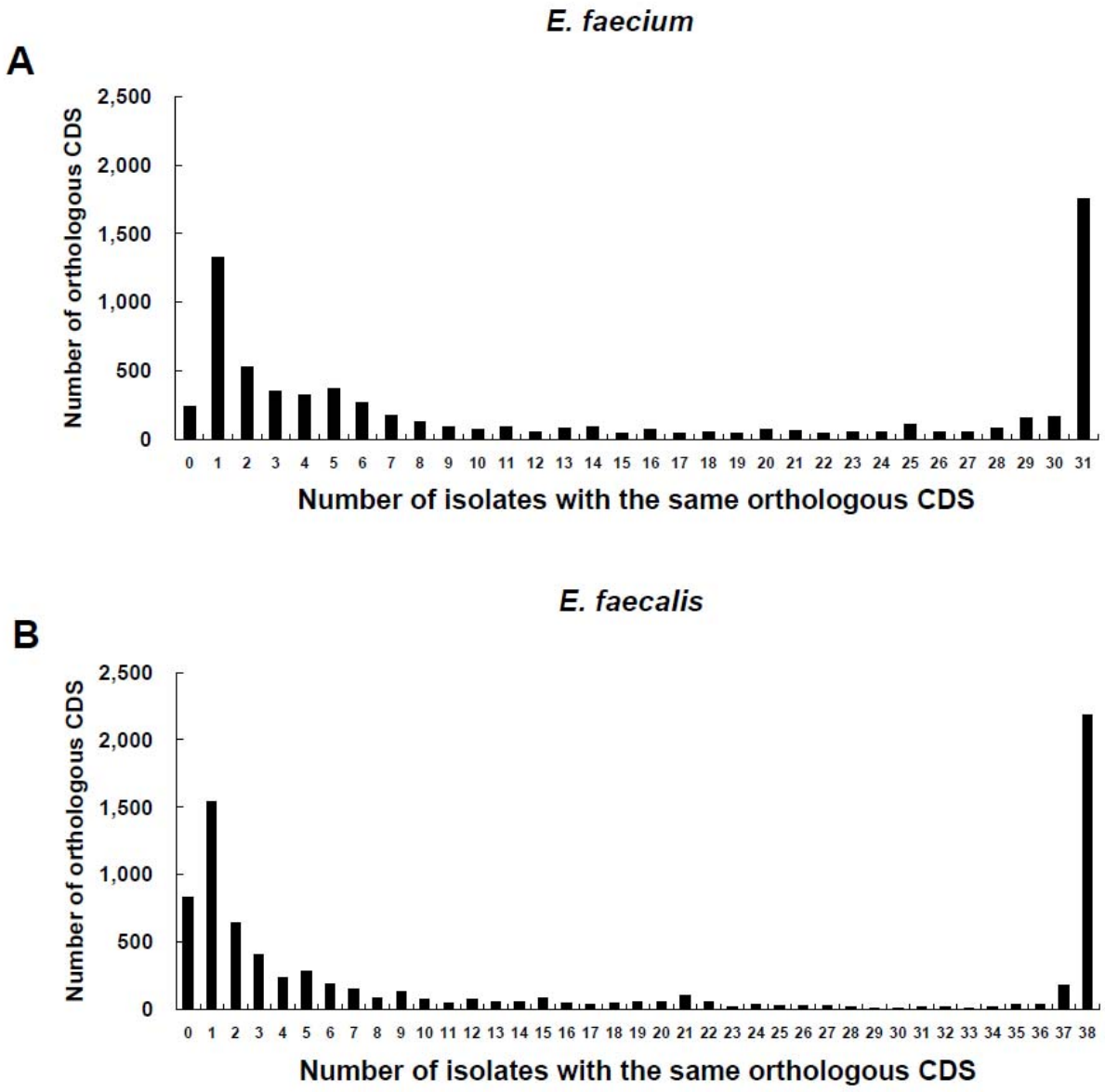

Figure S2. Distribution of orthologous CDS in 31 *E. faecium* and 38 *E. faecalis* genomes. Distribution histograms are shown for *E. faecium* (A) and *E. faecalis* (B) species. The horizontal axis indicates the number of isolates sharing the same orthologous CDS, and vertical axis represents the number of orthologous CDS shared by the indicated number of isolates. A number of accessory genes were not found in the 31 *E. faecium* and 38 *E. faecalis* genomes, which is indicated by "0" in the horizontal axis. The accessory genes were from different isolates, genomes of which are not sequenced yet.

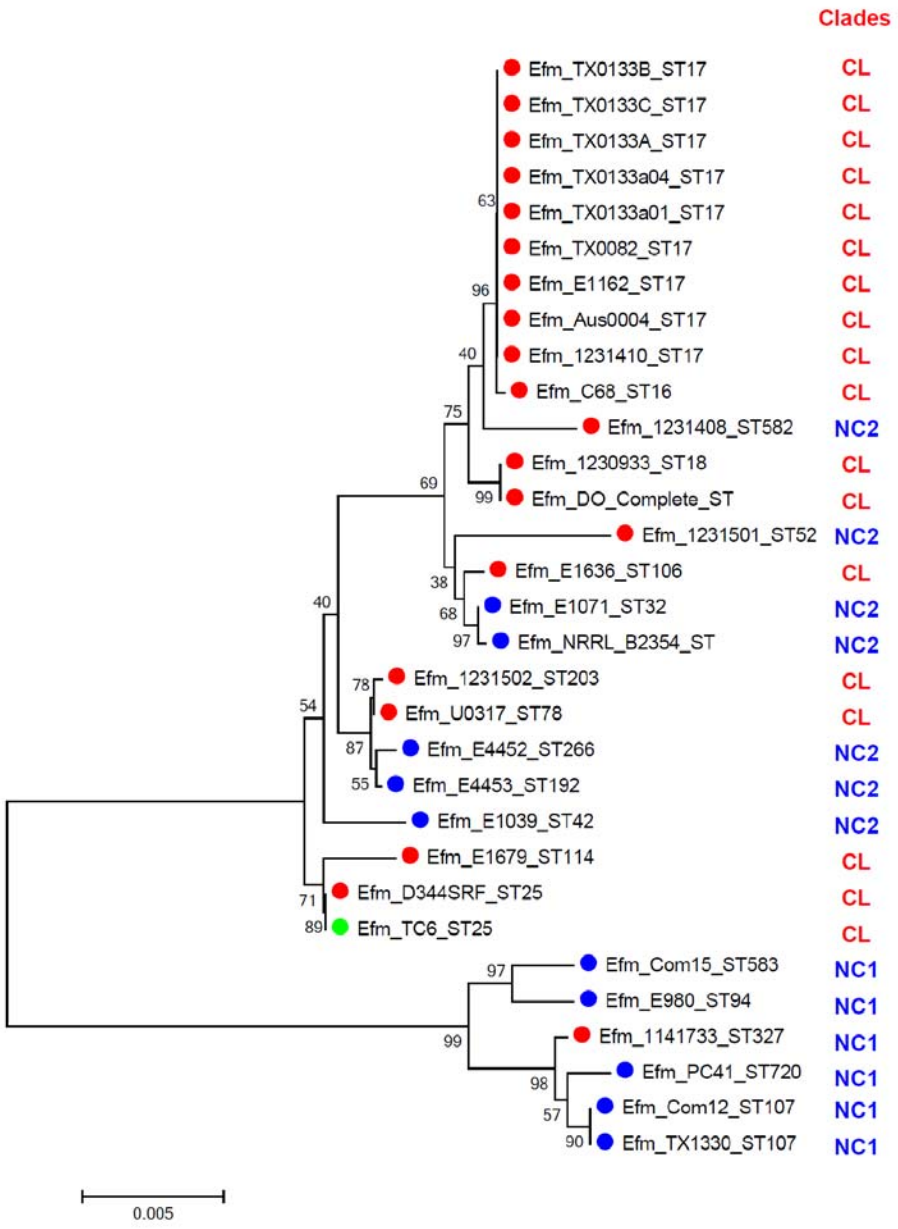

Figure S3. Phylogenetic analysis of MLST sequences found in different *E. faecium* strains.

Phylogenetic analysis was performed using MLST sequences (concatenated sequences of 7 housekeeping genes *atpA*, *ddl*, *gdh*, *purK*, *gyd*, *pstS*, and *adk*) found in 31 *E. faecium* strains. Sequence typing groups are indicated with strain names. Reproducible relationships between associated taxa clustered together were confirmed by the neighbor joining with 1,000 times of re-samplings. The scale bar indicates the number of nucleotide substitutions per site.

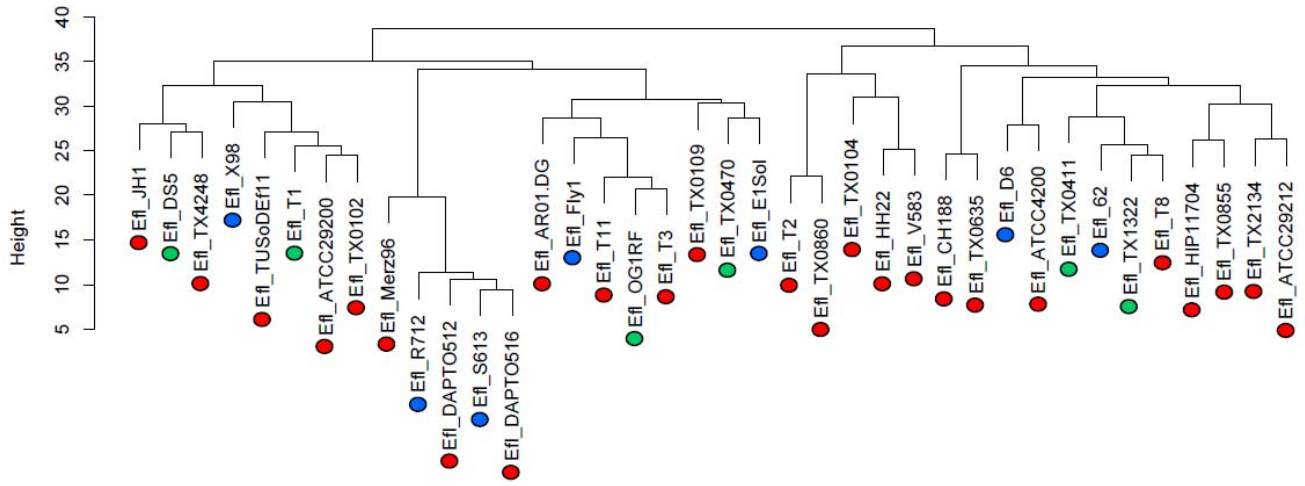

Figure S4. Hierarchical clustering of *E. faecalis* isolates based on orthologous CDS contents. NC isolates (blue ovals, N = 7), CL isolates (red ovals, N = 25), and isolates from unclear origins (green ovals, N = 6) are shown in the diagram.

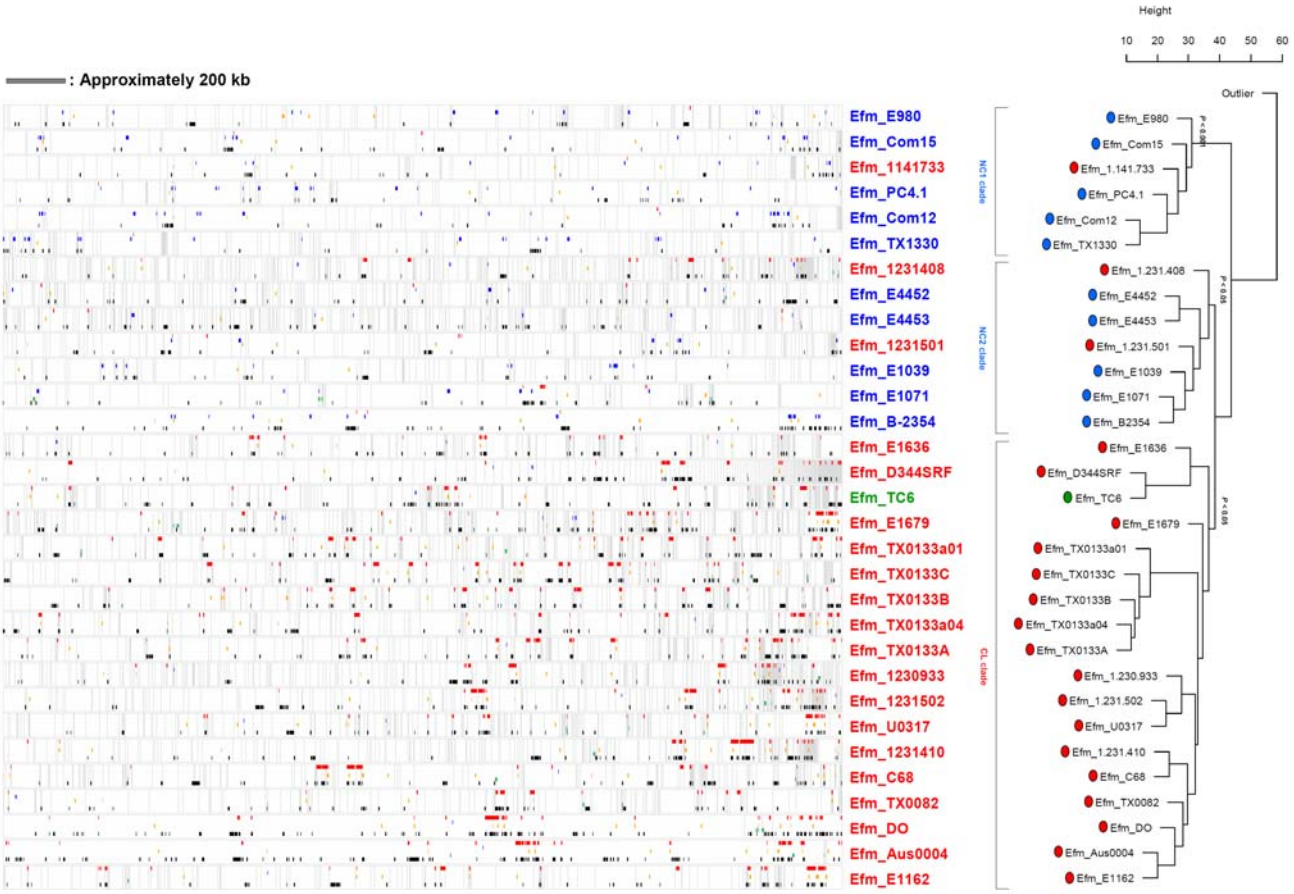

Figure S5. Gene distribution in *E. faecium* genomes for proximity analysis. Gray rectangles indicate assembled sequences such as contigs, plasmids, or chromosomes from each assembled genome. NC-enriched genes (blue), CL-enriched genes (red), VF genes (orange), AR genes (green), and ME genes (black) were aligned to the assembled sequences, and the aligned genes were visualized. The assembled sequences are shown according to their order in each GenBank file. If one gene/locus was overlapped with or next to another gene/locus within 1,000 bp in the same contig, the two genes/loci were considered co-localized. At the upper side of the gene distribution array, isolates names are listed in the same order as shown in the hierarchical clustering. A scale bar is shown at the left side of the gene distribution array.

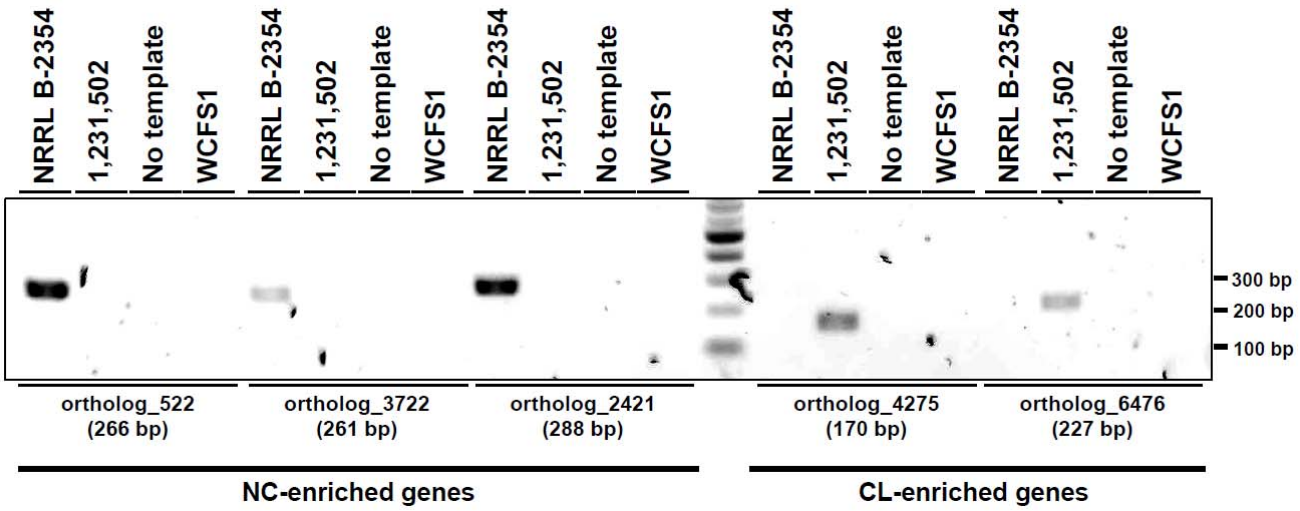

Figure S6. **PCR amplification of NC-and CL-enriched genes**. Genomic DNA from one *E. faecium* NC strain (NRRL B-2354), one *E. faecium* CL strain (1,234,502), and one negative control (*Lactobacillus plantarum* WCFS1) were used as templates for PCR detection. Each PCR product shows the same size as the expected according to Table S4.

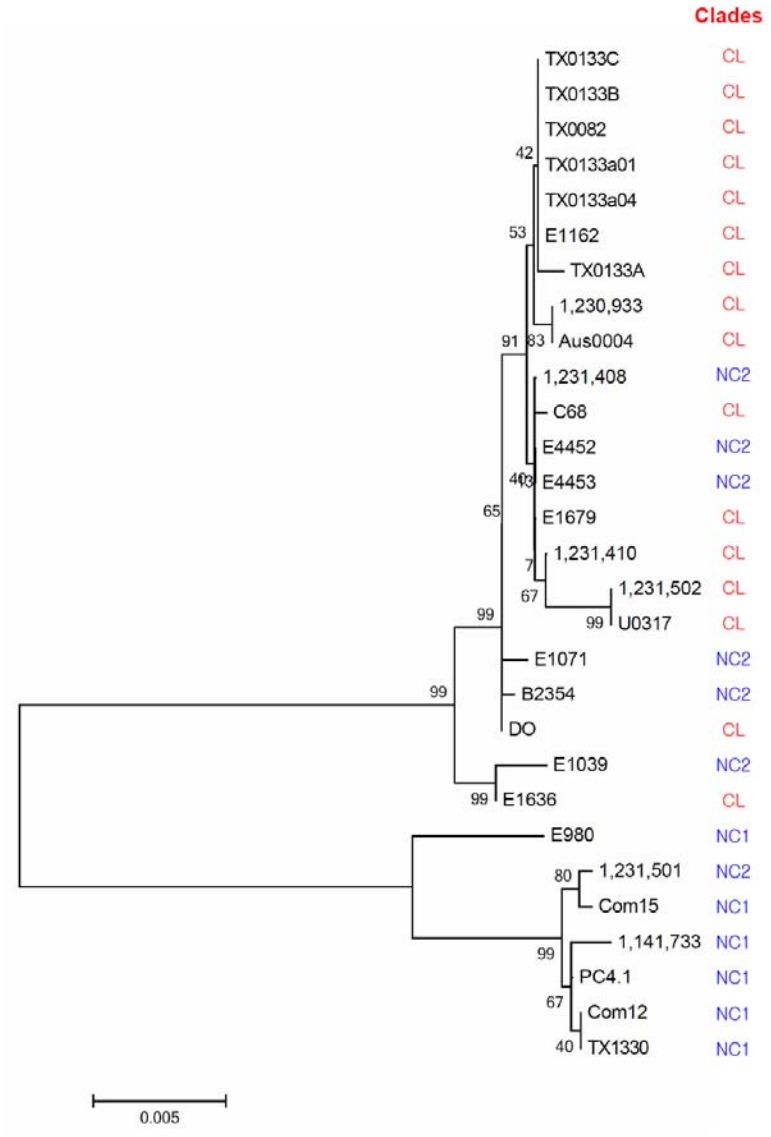

Figure S7. Phylogenetic analysis of *pbp5* sequences found in different *E. faecium* strains.

Phylogenetic analysis was performed using *pbp5* sequences found in 29 *E. faecium* strains. The *pbp5* gene was not found in two *E. faecium* genomes. Reproducible relationships between associated taxa clustered together were confirmed by the maximum likelihood methods with 500 times of re-samplings. The scale bar indicates the number of amino acid substitutions per site.