Supplemental Material

1. Supplemental Material and Methods
2. Supplemental Figures 1-7
3. Supplemental Table 1

Supplemental Methods

Specimen Collection

Bacterial strains isolated from blood cultures at MD Anderson Cancer Center (MDACC) in Houston TX are stored for future analysis under an MD Anderson Institutional Review Board approved protocol (PA13-0334). The MDACC routinely performs quantitative blood cultures and will perform species assignment for coagulase-negative staphylococci when >10 colony forming units are isolated from blood cultures as a means to decrease workup for contaminants. Thus, all *S. epidermidis* strains in the study met this criteria (August 2013 to December 2015) and, additionally, isolates were only analyzed if they met the United States Centers for Disease Control and Prevention criteria for bloodstream infection. A waiver of informed consent to collect clinical data and analyze the isolates was provided by the MDACC IRB (PA16-0066). Data on antimicrobial susceptibility from the MDACC clinical microbiology laboratory was abstracted from the electronic medical record with linezolid susceptibility determined initially by Vitek and then confirmed by E-test if the Vitek result showed a minimum inhibitory concentration (MIC) ≥ 4 mg/L.

Illumina Short Read Genome Sequencing

Genomic DNA from 176 *S. epidermidis* isolates was extracted from a single colony of the stocked cultures grown overnight in Muehller-Hinton (MH) broth at 37º C under shaking conditions. High-quality gDNA was isolated using the MasterPure Kit (Illumina Inc., San Diego CA). An average of 10 µg of gDNA was input for paired-end (PE) sequencing on the Illumina MiSeq instrument using TruSeq chemistry at the MD Anderson DNA Sequencing and Microarray Facility.

Short Read Processing and Genome Assembly

Short read quality was assessed using the FASTQC toolkit (Babraham Institute), and adaptors as well as low quality reads were trimmed using Trimmomatic v 0.33 [1]. Genome assembly was performed using SPAdes v 3.9.1 [2]; 28 genomes whose SPAdes assemblies either failed, or were deemed poor (> 300 contigs), were reassembled with SeqMan Ngen v 14.1 (DNASTAR® Inc, Madison WI). Custom scripts were used to remove contigs less than 300 bp and the phi X 174 sequence, respectively. This pipeline was implemented on the high performance computing cluster at MD Anderson.

Complete Genome Assembly of MB151

Genomic DNA from overnight culture was isolated using a phenol/chloroform extraction protocol and subjected to large-insert PacBio library preparation following the User Bulletin – Guideline for Preparing 20 kb SMRTbellTM Templates (version 2) and Procedure and Checklist – 20kb Template Preparation Using BluePippin Size-Selection (Version 3) (http://www.pacificbiosciences.com/support/pubmap/documentation.html). SMRTbell templates were subjected to standard SMRT sequencing using an engineered phi29 DNA polymerase on the PacBio RS system according to the manufacturer's protocol. *de novo* genome assembly using PacBio reads was performed with the HGAP pipeline [3]. Subsequently, PE Illumina short reads were mapped to the polished PacBio assembly for error correction using Bowtie2 v2.2.3[4, 5].

Characterization of the MB151 Chromosome and its Plasmids

Annotation of the closed MB151 chromosome along with the three plasmids was performed with RASTtk [6]. Genomic comparison between the chromosome of MB151 and 7 other published *S. epidermidis* strains - 14.1.R1, 1457, ATCC12228, BPH0662, PM221, RP62A, SEI – was performed with the BLAST Ring Image Generator (BRIG) [7]. Phage insertions in the MB151 genome were identified with PHAST [8]. Staphylococcal cassette chromosome

(SCCmec) elements typing was reported following the International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements guidelines (http://www.sccmec.org/Pages/SCC_TypesEN.html). The *orfX*, IS-elements and transposons, *mec*, and the *ccr* gene complex were identified via RASTtk annotation and were further annotated using BLAST searches against the NCBI non-redundant database. Each BLAST hit was then confirmed by a reciprocal protein BLAST before the *ccr* type was assigned to MB151. BlastN search with the *cfr* sequence (AM408573) was used to determine the presence or absence of *cfr* in the MB151 genome. Plasmid RASTtk annotations were screened for the presence of antibiotic resistance genes.

Phylogenomic Computations

Here we deployed three distinct core SNP-based methods to reconstruct whole genome sequencing (WGS)-based phylogenies: a) kSNP v 3.0, an alignment- and reference-independent method, which identifies SNPs using a k-mer approach and then reconstructs phylogenies using maximum parsimony [9]; b) Harvest suite v 1.0, which is designed to reconstruct phylogenies using a large number bacterial genomes while taking into account recombination events that may distort phylogenetic signal [10]. Draft genomes were aligned to the MB151 reference genome using *parsnp* and approximate maximum-likelihood (ML) trees reconstructed by FastTree 2 were visualized with *Gingr*. Finally, c) we developed an in-house GATK inspired pipeline to reconstruct WGS phylogenies [11]. Briefly, trimmed PE reads were mapped to the MB151 reference genome with Bowtie2 v 2.2.3. Read alignments were sorted and indexed uing Samtool v 1.3.1 [12], duplicate reads were removed using Picard v 2.9.0-1 (Broad Institute, Cambridge, MA), and local realignments around indels were performed with the RealignerTargetCreator and IndelRealigner modules available in the GATK 3.7 toolkit [11]. Since all strains share a uniform reference, a multi-sample pileup of aligned reads containing read depth, base quality scores was generated with the mpileup option in Samtools and then

used for variant calling with bcftools v 1.3.1 [13]. Only SNPs that had an allele depth greater than 10 and genotype call quality score greater than 10 were considered for downstream phylogenetic analysis. ML-based phylogenetic reconstruction was performed with RAxML v 8.2.10 using the GTR + Γ nucleotide substitution model [14]. 100 bootstrap replicates were evaluated to determined branch support.

Characterization of *cfr* Location in Strain MB151, Plasmid Transfer, and Plasmid Presence Characterization

S1 nuclease assays were used to detect and estimate the size of *cfr*-carrying large bacterial plasmids [15] in *S. epidermidis* isolates. Briefly, agarose gel plugs containing bacterial cells were incubated for 45 min with 0.01 U of *Aspergillus oryzae* S1 nuclease (Sigma- Aldrich, St. Louis, MO) in 200 ul of 50 mM NaCl, 30 mM sodium acetate (pH 4.5), and 5 mM ZnSO4. Digested plugs were subjected to PFGE using an initial time of 2 h, final time of 28 h and a run of 24 h (14ºC). Subsequently, plasmids were transferred to a nylon positive-charged membrane and hybridized with a digoxigenin labelled *cfr* probe, obtained by PCR amplification using total DNA from strain MB151 as target, with primers *cfrF:* 5' tgaagtataaagcaggttgggagtca *3'* and *cfrR*: *5'*accatataattgaccacaagcagc *3'* Hybridization was performed using a non-radioactive methodology following the manufacturer recommendations (Roche). Conjugative transfer of *cfr* was performed by filter mating [16] using *S. epidermidis* MB151 as donor and *S. aureus* RN4220RF as recipient. Different ratios between the donor and recipient were used (1:10, 1:5, 10:1) and the selection of transconjugants was performed using brain heart infusion (BHI) agar plates supplemented with chloramphenicol (20 ug/ml) and fusidic acid (25 ug/ml). Chloramphenicol was used since *cfr* primarily confers resistance to this antibiotic [17]. We also attempted to transfer B-lactam (using ampicillin) and gentamicin resistance determinants from *S. epidermidis* MB151 (ampicillin and gentamicin MICs >16 ug/ml and >32 ug/ml, respectively) to *S. aureus* RN4220-RF (ampicillin and gentamicin MICs 0.094 and 1 ug/ml, respectively).

Transconjugants were selected on BHI supplemented with ampicillin (16 ug/ml) and fusidic acid (25 ug/ml) and gentamicin (32 ug/ml) and fusidic acid (25 ug/ml) with ratios of donor/recipient of 1:10, 1:5 and 10:1. As positive controls of the filter mating experiments, we used *cfr*-positive *E. faecalis* 603-50427X) as donor and *E. faecalis* OG1RF as recipient. We successfully transferred the *cfr* gene to the recipient strains as we have previously reported [18].

Strains from which reads covered more than 80% of the total block length of pMB151 were considered as having pMB151a present. The mapping patterns of pMB151a across our population revealed that certain neighbor genes either mapped well together or did not map at all. For quantification purposes, these regions were defined as blocks.

Linezolid Use Data

Our in-house pharmacy informatics database was queried to identify all inpatient and ambulatory infusion-clinic based administrations of linezolid and daptomycin. Daptomycin was chosen as a comparator agent as it is used in similar patient populations and provides a similar spectrum of activity, but would not be expected to lead to cross-resistance with linezolid. Cumulative exposure (defined as the number of unique days on which a drug was administered) and any use (defined as at least one dose) was considered for each drug at three time points: 30, 60, and 90 days preceding the first isolation of *S. epidermidis*.

Linezolid resistance mechanisms

The presence or absence of *cfr* was determined by a local Blastn search. Briefly, a local blast database comprised of 176 draft genomes was generated with *makeblastdb*. The MB151 *cfr* gene sequence was used as query to interrogate the local genome database. The prevalence of *cfr*-containing plasmid pMB151a was also investigated. Briefly, cleaned-up PE reads were mapped to pMB151a using BOWTIE2, and then mapped reads were visualized with Geneious® v 9.1.8 and quantified with seqMONK (Babraham Institute). Strains with reads that mapped to

the reference plasmid with greater than 95% coverage were considered as pMB151a; and those with greater 70% (but less than 95%) were considered as pMB151a derivatives. Mapped sequence data was manually inspected for each strain to confirm the presence or absence of *cfr*. Additionally, mutations in the 23S rRNA, L3p, L4p and L22p were also catalogued. Percentage of reads carrying mutations in all 23S rRNA copies were calculated after read mapping and then reported in *E. coli* 23S rRNA sequence numbering. Sequences encoding L3p, L4p, L22p were extracted, translated and aligned using MUSCLE as implemented in MEGA v 7 to identify sequence variation.

Identification of Acquired Non-Linezolid Antibiotic Resistance Genes

Homologs to antibiotic resistance genes were identified in our cohort using local protein BLAST searches. 176 draft genome sequences were queried against the Comprehensive Antibiotic Resistance Database v1.1.8 [19]. BLAST hits which met an E-value of 1e-100, an identity > 60% and a gene coverage > 70% were considered true homologs.

Microbiome Data Analysis

The origin of stool samples for microbiome analyses have been previously described with all patients provided written informed consent [20]. In brief, stool samples were obtained from twice-weekly from patients undergoing induction therapy for acute myelogenous leukemia from time of enrollment until recovery of neutrophil count which generally took about 30 days. DNA was extracted as described and processed for 16s rRNA analysis using the V3-V4 region. Staphylococcal dominance of the gastrointestinal microbiome was defined as patients having at least two consecutive stool samples in which ≥ 30% of the 16S rRNA reads mapped to the staphylococcal genus and in which the baseline stool sample had ≤ 10% of reads mapping to *Staphylococcus* [21]. Note that for staphylococci, our 16S rRNA data was not capable of species designation. Residual DNA from the stool microbiome samples was analyzed for *S.*

epidermidis via SYBR green using previously published *S. epidermidis*-specific primers [22] and *cfr* using the primer set 5' cctgagatgtatggagaagcaaacg and 3' agcagcgtcaatatcaatcccaaat. Total 16S rRNA was used to normalize *cfr* Ct values via previously published Eubacteria universal primers UniF340 and UniR514 [23]. Input genomic DNA extracted from stool samples were normalized to 1 ng/μL and then amplified using the QuantiTect SYBR® Green PCR kit (QIAGEN, Valencia, CA) on the ABI StepOnePlus Real-Time PCR Systems (Applied Biosystem, CA). PCRs were performed with the following cycling parameters: denaturation at 95°C for 15 min, followed by 40 cycles at 94°C for 15 s, 61°C for 40 s and 72 °C for 30 s. The reaction mix consisted of 1ng gDNA and 0.1 μM of each primer pair. Standard and melting curve analyses were performed to verify the amplification efficiency and specificity.

To ensure that our *S. epidermidis* specific primers were accurately amplifying S*. epidermidis*, we plated out the 6 stool samples that had tested positive for *S. epidermidis* by our PCR based method onto Columbia CNA agar (Becton-Dickinson). CNA agar contains colistin and nalidixic acid to select for gram-positive bacteria. Genomic DNA was isolated from 10 individual colonies from each sample (total 60 isolates), and 16s rRNA was amplified using the universal primers 8F and 519R [24]**.** PCR products were subjected to Sanger sequencing followed by BLASTn anlaysis. For all 60 isolates, the 16s 16s rRNA data most closely matched with *S. epidermidis*. To ensure the specificity of our protocol, we plated three stool samples with significant levels of staphylococcus in our microbiome analysis but from which our *S. epidermidis* primers failed to amplify a product. 16s rRNA amplified from colonies from these samples matched to a variety of gram-positive including *S. aureus* (2 samples) and *S. capitis* (1 sample) along with several non-staphylococcal genera (*Streptococcus*, *Bacillus*, etc.) but no matches with *S. epidermidis* were found.

Statistical Analyses

A change in the rate of linezolid resistance over time was assessed using the Cochran-Armitage test for trend. Beginning in June 2010, an antimicrobial stewardship intervention aimed at reducing use of linezolid was introduced in our hospital. Therefore, variance in the monthly use of linezolid (reported as days of therapy per 1,000 patient days) was compared using segmented regression analysis before and after implementation of this intervention[25]. Individual use of daptomycin and linezolid were analyzed using Fisher's exact test and cumulative exposure was assessed using the Wilcoxon Rank-Sum test. A p-value ≤ 0.05 was considered significant. All analyses were performed using Stata v13.1 software (StataCorp LP, College Station, TX).

REFERENCES FOR SUPPLEMENTAL METHODS

1. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **2014** Aug 01; 30:2114-20.
2. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **2012** May; 19:455-77.
3. Liao YC, Lin SH, Lin HH. Completing bacterial genome assemblies: strategy and performance comparisons. Scientific reports **2015** Mar 04; 5:8747.
4. Koren S, Harhay GP, Smith TP, et al. Reducing assembly complexity of microbial genomes with single-molecule sequencing. Genome Biol **2013**; 14:R101.
5. Liu Y, Schmidt B. Long read alignment based on maximal exact match seeds. Bioinformatics **2012** Sep 15; 28:i318-i24.
6. Aziz RK, Bartels D, Best AA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics **2008**; 9:75.
7. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics **2011**; 12:402.
8. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res **2011** Jul; 39:W347-52.
9. Gardner SN, Hall BG. When whole-genome alignments just won't work: kSNP v2 software for alignment-free SNP discovery and phylogenetics of hundreds of microbial genomes. PLoS One **2013**; 8:e81760.
10. Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol **2014**; 15:524.

11. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Current protocols in bioinformatics **2013**; 43:11 0 1-33.

12. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics **2009** Aug 15; 25:2078-9.

13. Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics **2016** Jun 01; 32:1749-51.

14. Stamatakis A, Ludwig T, Meier H. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics **2005** Feb 15; 21:456-63.

15. Barton BM, Harding GP, Zuccarelli AJ. A general method for detecting and sizing large plasmids. Anal Biochem **1995** Apr 10; 226:235-40.

16. Tomita H, Pierson C, Lim SK, Clewell DB, Ike Y. Possible connection between a widely disseminated conjugative gentamicin resistance (pMG1-like) plasmid and the emergence of vancomycin resistance in Enterococcus faecium. J Clin Microbiol **2002** Sep; 40:3326-33.

17. Toh SM, Xiong L, Arias CA, et al. Acquisition of a natural resistance gene renders a clinical strain of methicillin-resistant Staphylococcus aureus resistant to the synthetic antibiotic linezolid. Mol Microbiol **2007** Jun; 64:1506-14.

18. Diaz L, Kiratisin P, Mendes RE, Panesso D, Singh KV, Arias CA. Transferable plasmid-mediated resistance to linezolid due to cfr in a human clinical isolate of Enterococcus faecalis. Antimicrob Agents Chemother **2012** Jul; 56:3917-22.

19. Jia B, Raphenya AR, Alcock B, et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res **2016** Oct 26.

20. Galloway-Pena JR, Smith DP, Sahasrabhojane P, et al. The role of the gastrointestinal microbiome in infectious complications during induction chemotherapy for acute myeloid leukemia. Cancer **2016** May 3.

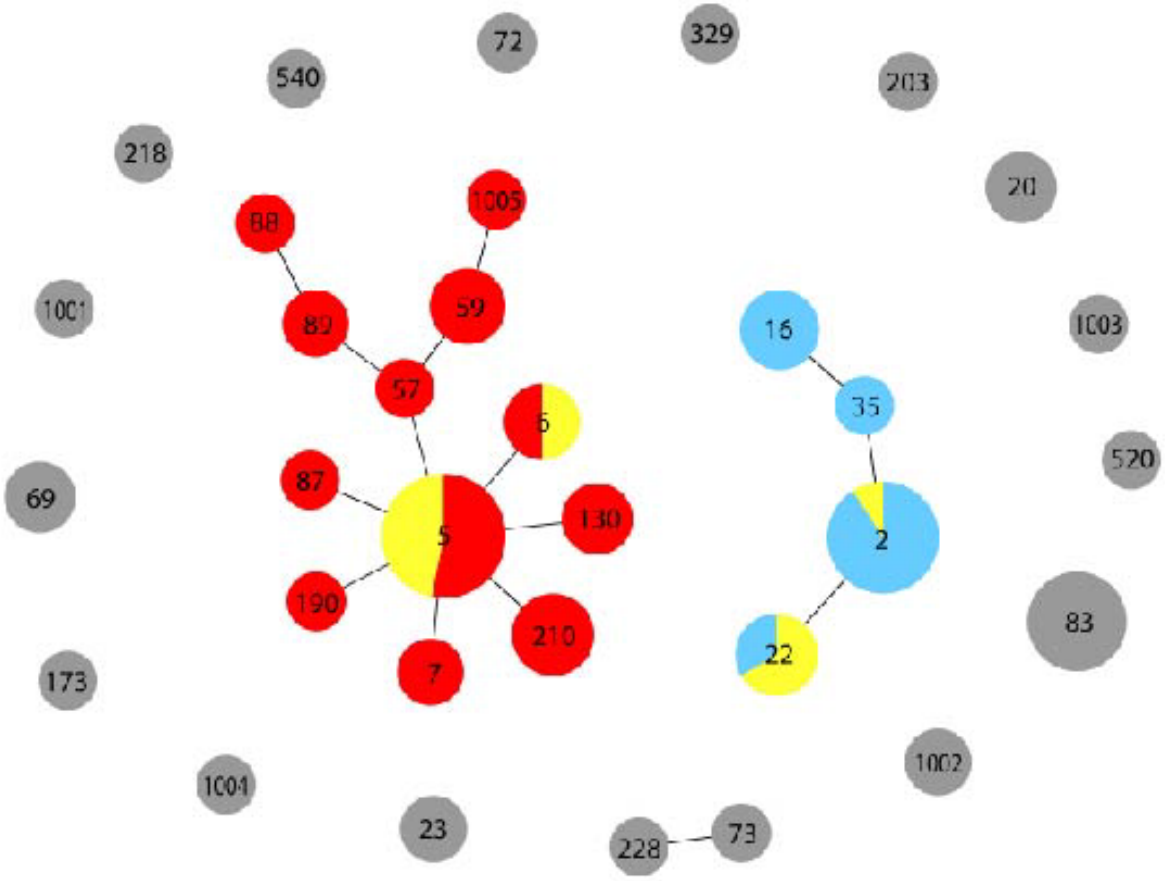
21. Taur Y, Xavier JB, Lipuma L, et al. Intestinal domination and the risk of bacteremia in patients undergoing allogeneic hematopoietic stem cell transplantation. Clin Infect Dis **2012** Oct; 55:905-14.

22. Martineau F, Picard FJ, Roy PH, Ouellette M, Bergeron MG. Species-specific and ubiquitous DNA-based assays for rapid identification of Staphylococcus epidermidis. J Clin Microbiol **1996** Dec; 34:2888-93.

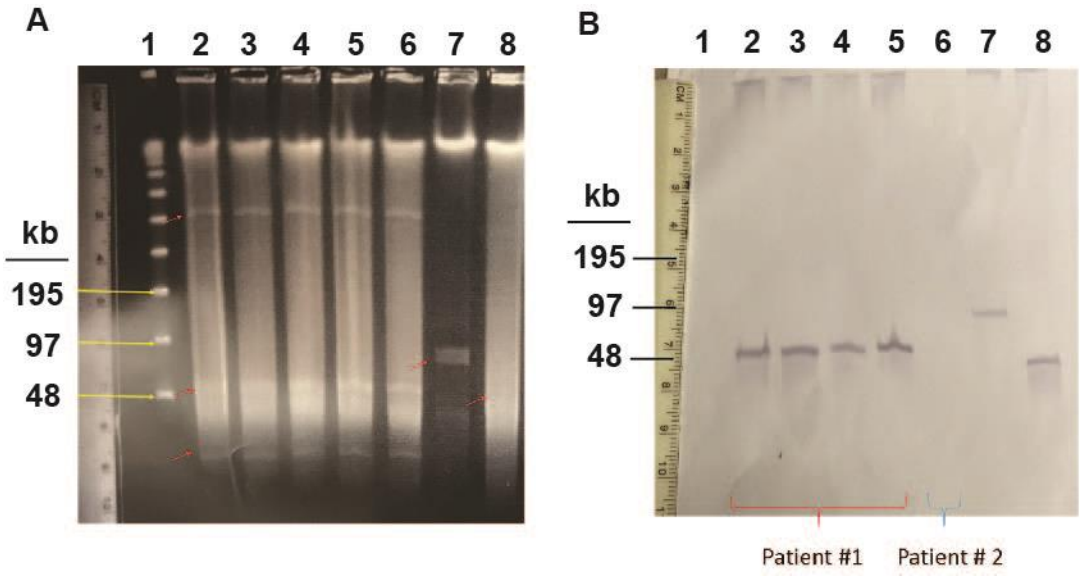
23. Barman M, Unold D, Shifley K, et al. Enteric salmonellosis disrupts the microbial ecology of the murine gastrointestinal tract. Infect Immun **2008** Mar; 76:907-15.

24. Turner S, Pryer KM, Miao VP, Palmer JD. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. J Eukaryot Microbiol **1999** Jul-Aug; 46:327-38.

25. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. J Clin Pharm Ther **2002** Aug; 27:299-309.

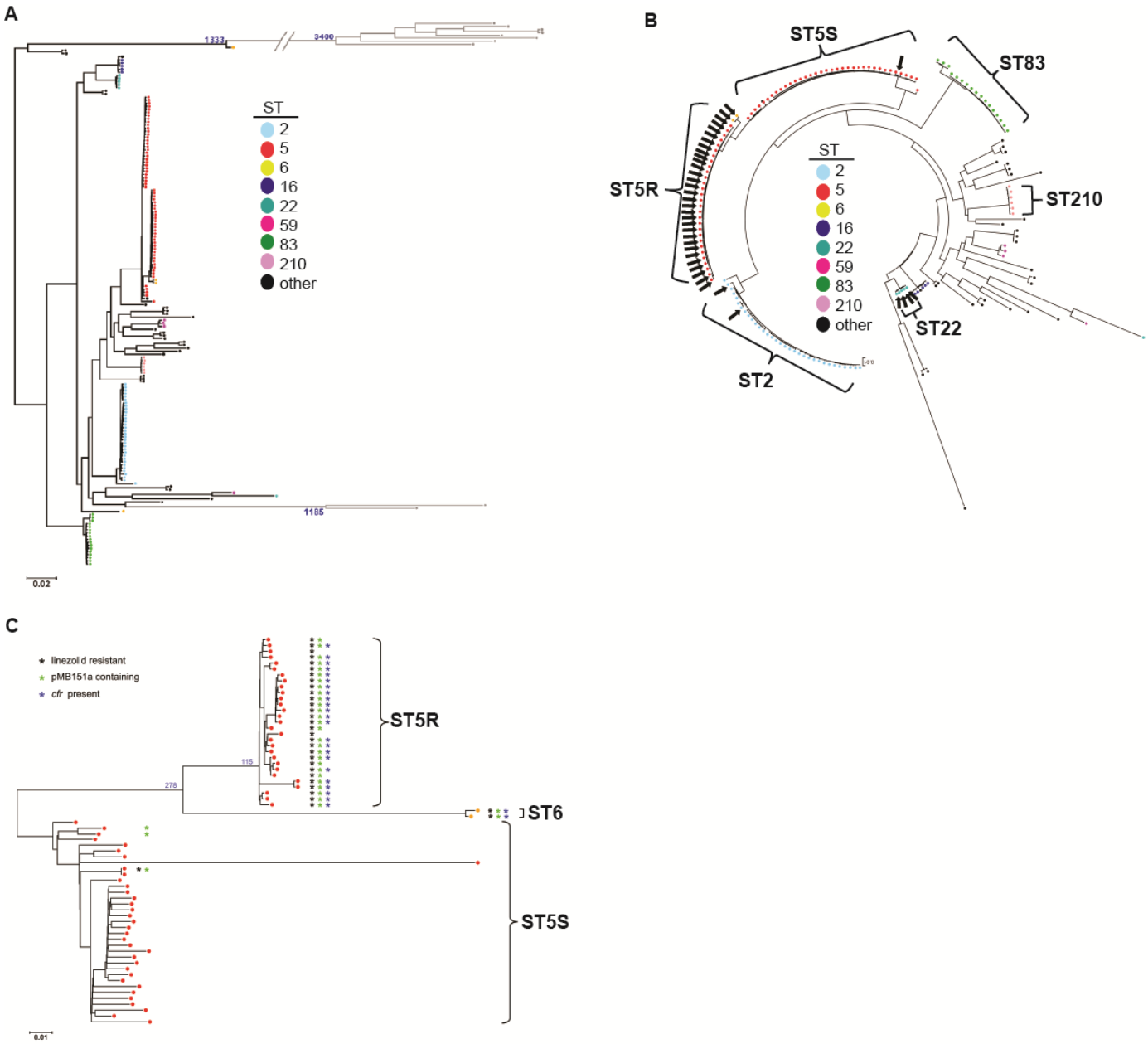


Supplemental Figure 1. Linezolid resistance in *S. epidermidis* is predominantly found in clonal complex (CC)-5 strains. Eburst analysis of MLST data was performed. Circle size is relative to number of isolates with sequence type indicated by number. CC-5 strains are colored red whereas CC-2 strains are colored blue. Yellow pie within circle indicates percent of strains that are resistant to linezolid.



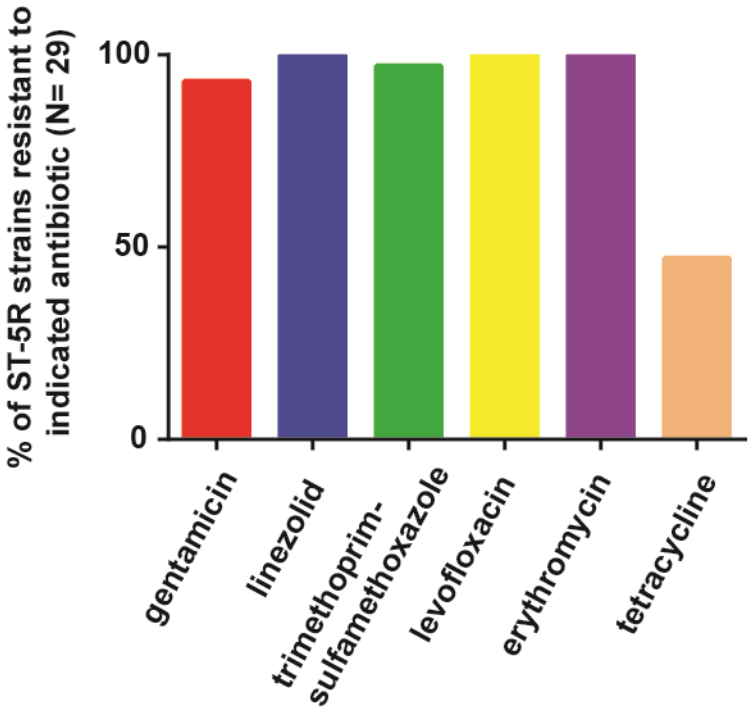
Supplemental Figure 2. Identification of *cfr* on a 49 kb plasmid. (A) S1 nuclease assay showing presence of a plasmid at size of ~49 kb as predicted by sequencing data. Lanes 2-5 are serial isolates of MB151 collected during prolonged bacteremia. Lane 6 is ST5 S. epidermidis strain that lacked cfr by sequencing data. Lanes 7 and 8 are strains harboring *cfr* in plasmids. Lane 7 is *Enterococcus faecalis* 603-50427X56 and lane 8 is *Staphylococcus aureus* 004-737X57. (B) cfr hydridization assay showing presence of *cfr* on 49 kb plasmid. Lanes are as shown for panel B.

Supplemental Figure 3. Examples of read mapping patterns used to determine presence of pMB151a and *cfr*. Shown are read mapping coverage relative to pMB151a. *Cfr* gene is shown in red. Blue on the y-axis shows depth of read mapping. The top strain contains all of pMB151a. The middle strain contains pMB151a but lacks *cfr*. The bottom strain lacks pMB151 entirely. The blue present in bottom panel indicates read cross-mapping given the presence of similar genes on the chromosome. Genes and transposases present in the pMB151a mentioned in the manuscript are labeled.

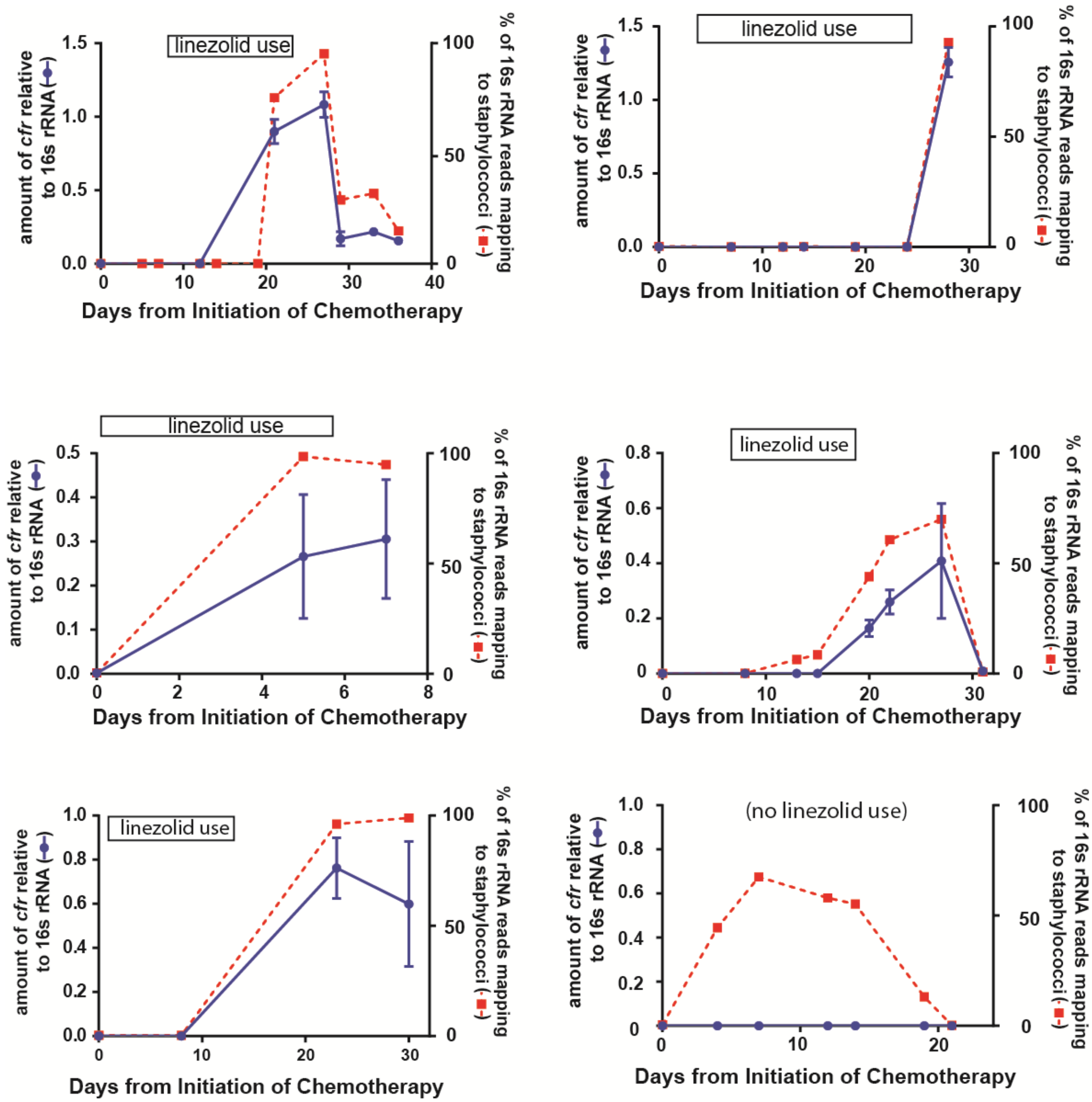

A
1333
3499
1185
0.02
ST
2
5
6
16
22
59
83
210
other
B
ST5S
ST83
ST5R
ST210
ST22
ST2
ST
2
5
6
16
22
59
83
210
other
C
linezolid resistant
pMB151a containing
cfr present
ST5R
ST6
ST5S
115
278
0.01

Supplemental Figure 4. Whole genome phylogenetic analysis using the kSNP v 3.0 identified two distinct ST5 lineages. (A) Whole genome phylogeny of 176 strains. Parsimony tree was reconstructed based on an alignment of 33,060 core SNPs. To minimize any potential bias introduced via long branch attraction 10 strains (in grey) were excluded from subsequent analysis (B) Whole genome phylogeny of 166 strains. Parsimony tree was reconstructed based on an alignment of 25,516 core SNPs. Major STs are indicated in colored circles. Arrows show linezolid resistant strains. (C) Parsimony tree was reconstructed based on an alignment of 3865 core SNPs. ST-5 strains separated into linezolid resistant (ST5R) and sensitive clusters (ST5S). Two ST-6 isolates also clustered with the ST5R isolates. The presence of linezolid resistance plasmid pMB151a (green asterisk) and *cfr* (blue asterisk) is indicated among linezolid resistant (ST5R) strains (black asterisk). For (A-C), the number of SNPs unique to each node are shown above the branches. Branch lengths were expressed in terms of changes per number of SNPs. The clustering of linezolid resistant ST-5 strains of both ST5 phylogeny and 166 strains phylogeny were consistent with read-based trees (see Figure 4).

Supplemental Figure 5. Whole genome phylogenetic analysis using the HARVEST suite ver 1.0 identified two distinct ST5 lineages. Each strain in the phylogeny (left margin) reconstructed using an approximately ML approach is paired with its corresponding row in the multi-alignment. The SNP density plot (right) shows the unique phylogenetic signatures of ST5S and ST5R. SNPs are visualized in gingr as vertical purple lines.



Supplemental Figure 6. Broad antibiotic resistance present in ST5R strains. Data graphed are % resistance to various antibiotics for all ST5R strains (N = 29 isolates).



Supplemental Figure 7. Linezolid use precedes staphylococcal emergence and cfr detection in the gastrointestinal microbiome. The six panels represent individual patients who experienced emergence of staphylococci while undergoing induction remission chemotherapy for acute myelogenous leukemia. X-axis shows day of stool sampling relative to start of chemotherapy. Right y-axis and red squares with dotted red line show % of 16s rRNA reads mapping to staphylococcal genus from the stool samples. Left y-axis and squares with solid

blue line depict amount of *cfr* present in stool samples relative to total 16s rRNA as determined by QRT-PCR with data graphed being mean ± standard deviation of samples analyzed in triplicate. The time of linezolid use is shown in the rectangle. The patient depicted in the bottom right panel had no linezolid use or detectable *cfr*.

Supplemental Table 1. Strain details.

Strain Name	*arcC*	*aroE*	*gtr*	*mutS*	*pyrR*	*tpiA*	*yqiL*	Sequence type	Assigned ST for this study (includes single nucleotide variants)	Clonal complex	ST5 subclone (if applicable)	Linezolid resistant	Cfr present	Plasmid pMB151a present	Comments
MB1016	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB1057	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB1091	7	1	2	2	4	7	1	22	22	2		Yes	No	No	
MB1096b	1	1	1	2	2	1	1	5	5	5	5R	Yes	No	No	
MB1191	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB1193	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB1226	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB1278	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB1335	1	1	2	2	2	1	1	6	6	5		Yes	Yes	Yes	
MB1427	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB1513	1	1	1	2	2	1	1	5	5	5	5R	Yes	No	Yes	
MB151C	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB1521	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB1593	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB1619	7	1	2	2	4	7	1	22	22	2		Yes	No	No	
MB1656	7	1	2	2	4	1	1	2	2	2		Yes	No	No	
MB1669	1	1	1	2	2	1	1	5	5	5	5R	Yes	No	Yes	
MB1679	1	1	1	2	2	1	1	5	5	5	5S	Yes	No	Yes	
MB1769	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB1851	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB1871a	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB1962	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB1999	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB2033	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB2126	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB2177a	1	1	1	2	2	1	1	5	5	5	5R	Yes	No	No	
MB2268	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB3116	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB3129	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB609	7	1	2	2	4	1	1	2	2	2		Yes	No	No	
MB746	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB793	1	1	1	2	2	1	1	5	5	5	5R	Yes	No	Yes	
MB806	1	1	1	2	2	1	1	5	5	5	5R	Yes	No	Yes	
MB927	7	1	2	2	4	7	1	22	22	2		Yes	No	No	
MB929	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB931	7	1	2	2	4	1	1	2	2	2		Yes	No	No	
MB939	1	1	2	2	2	1	1	6	6	5		Yes	Yes	Yes	
MB947	1	1	2	2	4	7	1	22	22	2		Yes	No	No	
MB951	1	1	1	2	2	1	1	5	5	5	5R	Yes	Yes	Yes	
MB037	1	2	1	2	1	1	10	83	83			No	No	No	
MB10	1	1	1	2	2	1	1	5	5	5	5S	No	No	No	
MB101	1	1	1	2	4	1	1	7	7	5		No	No	No	
MB1034	1	1	1	2	2	1	1	5	5	5	5S	No	No	No	
MB1037	1	1	1	2	2	1	1	5	5	5	5S	No	No	No	
MB1043	1	1	2	1	2	1	7	88	88	5		No	No	No	
MB1048	7	1	2	2	4	1	1	2	2	2		No	No	No	
MB1064	1	1	1	2	2	1	1	5	5	5	5S	No	No	No	
MB108	1	1	unknown	2	2	1	1	unknown	5	5	5S	No	No	No	single nucleotide variant from ST5
MB1082	1	2	1	2	1	1	10	83	83			No	No	No	
MB1119	7	1	2	2	4	1	1	2	2	2		No	No	No	
MB113	8	2	2	4	9	6	9	72	72			No	No	No	
MB1143	1	1	1	2	2	1	25	210	210	5		No	No	No	
MB1145	7	1	2	2	4	1	1	2	2	2		No	No	No	
MB1153	7	1	2	2	4	1	1	2	2	2		No	No	No	
MB1182	2	1	2	2	15	1	1	16	16	2		No	No	No	
MB1208	7	1	2	2	4	1	1	2	2	2		No	No	No	
MB1211	1	1	1	2	2	1	1	5	5	5	5S	No	No	No	
MB1229	1	1	unknown	2	1	1	3	unknown	20			No	No	No	single nucleotide variant from ST20
MB123	7	1	2	1	3	3	1	23	23			No	No	No	
MB1230	2	1	2	2	15	1	1	16	16	2		No	No	No	
MB1260	1	1	2	1	2	1	1	89	89	5		No	No	No	
MB1298	1	2	1	2	1	1	10	83	83			No	No	No	
MB1299	1	1	1	1	2	1	1	57	57	5		No	No	No	
MB1306	7	1	2	2	4	1	1	2	2	2		No	No	No	
MB1312	1	1	1	2	2	1	1	5	5	5	5S	No	No	Yes	
MB1347	1	1	1	2	2	unknown	1	unknown	5	5	5S	No	No	No	single nucleotide variants from ST5
MB1364	1	1	1	2	1	1	1	130	130	5		No	No	No	
MB1368	1	1	1	2	2	1	1	5	5	5	5S	No	No	No	
MB1421	2	1	1	1	2	1	1	59	59	5		No	No	No	
MB1457	1	1	1	2	5	1	unknown	unknown	190	5		No	No	No	single nucleotide variant from ST190
MB1464	1	1	2	2	2	1	1	6	6	5		No	No	No	
MB1489	1	1	1	2	2	1	1	5	5	5	5S	No	No	No	
MB149	1	1	1	2	2	1	1	5	5	5	5S	No	No	No	
MB1502	7	1	2	2	4	1	1	2	2	2		No	No	No	
MB1532	1	1	1	2	2	1	1	5	5	5	5S	No	No	No	
MB1542	1	1	1	2	2	1	1	5	5	5	5S	No	No	No	
MB1546	1	2	1	2	1	1	10	83	83			No	No	No	
MB1547	1	2	1	2	1	1	10	83	83			No	No	No	
MB1567	7	1	2	2	4	1	1	2	2	2		No	No	No	
MB1568	1	2	1	2	1	1	10	83	83			No	No	No	
MB1569	7	1	2	2	4	1	1	2	2	2		No	No	No	
MB1586	1	1	1	2	2	1	25	210	210	5		No	No	No	
MB1595	7	1	2	2	4	1	1	2	2	2		No	No	No	
MB1609	1	1	2	1	2	1	1	89	89	5		No	No	No	
MB1617	7	1	2	2	4	1	1	2	2	2		No	No	No	
MB1651	7	1	2	2	4	1	1	2	2	2		No	No	No	
MB1675	1	2	1	2	1	1	10	83	83			No	No	No	
MB1691a	1	1	1	2	2	1	1	5	5	5	5S	No	No	No	
MB1695	1	1	1	2	2	1	1	5	5	5	5S	No	No	Yes	
MB1708	7	1	2	2	4	1	1	2	2	2		No	No	No	
MB1709	1	1	1	2	2	1	25	210	210	5		No	No	No	
MB1715	1	1	1	2	2	1	25	210	210	5		No	No	No	
MB1721	7	1	2	2	4	1	1	2	2	2		No	No	No	
MB1750	7	1	2	2	4	1	1	2	2	2		No	No	No	
MB1772	unknown		1	1	2	2	1	unknown	5	5	5S	No	No	No	single nucleotide variant from ST5
MB1812	3	16	9	5	3	4	5	203	203			No	No	No	

MB182	12	48	5	5	10	16	21	unknown	1001					No	No	No				
MB1825	1	2	1	2	1	1	10	83	83					No	No	No				
MB1830	12	29	9	8	6	5	8	329	329					No	No	No				
MB1849	2	1	2	2	15	1	1	16	16	2				No	No	No				
MB1872	1	1	2	6	2	42	33	540	540					No	No	No				
MB1907	1	1	1	2	2	1	25	210	210	5				No	No	No				
MB1908	28	3	8	5	7	5	4	unknown	1002					No	No	No				
MB1910	2	1	1	1	2	1	1	59	59	5				No	No	No				
MB1915	7	1	2	1	3	3	1	23	23					No	No	No				
MB192	1	5	2	6	2	1	unknown	unknown	73					No	No	No	single nucleotide variant from ST73			
MB1935	3	29	24	5	7	4	11	unknown	1003					No	No	No				
MB1956	1	2	1	2	1	1	10	83	83					No	No	No				
MB1961	1	1	1	2	2	1	1	5	5	5	5S			No	No	No				
MB1985	1	1	2	2	4	1	1	2	2	2				No	No	No				
MB198B	1	1	1	2	2	1	1	5	5	5	5S			No	No	No				
MB2013	7	1	1	unknown	2	1	1	unknown	87	5				No	No	No	single nucleotide variant from ST87			
MB2050a	1	1	2	2	1	1	3	20	20					No	No	No				
MB2095	7	1	2	2	4	1	1	2	2	2				No	No	No				
MB2193	1	1	1	2	2	1	1	5	5	5	5S			No	No	No				
MB227A	1	1	1	2	2	1	1	5	5	5	5S			No	No	No				
MB247	7	1	2	2	4	1	1	2	2	2				No	No	No				
MB255	28	44	5	5	7	5	11	unknown	1004					No	No	No				
MB268	1	2	1	2	1	1	10	83	83					No	No	No				
MB283	2	1	1	1	2	1	1	59	59	5				No	No	No				
MB312	13	5	2	6	2	1	6	228	228					No	No	No				
MB316	1	1	1	2	1	1	1	130	130	5				No	No	No				
MB323	7	unknown	2	2	4	1	1	unknown	2	2				No	No	No	single nucleotide variant from ST2			
MB339	28	3	8	5	7	5	4	unknown	1002					No	No	No				
MB383	1	1	1	2	2	1	1	5	5	5	5S			No	No	No				
MB396	1	1	1	2	2	1	1	5	5	5	5S			No	No	No				
MB415	1	6	6	1	2	1	10	173	173					No	No	No				
MB422	7	1	2	2	4	1	1	2	2	2				No	No	No				
MB429	1	1	2	1	1	1	33	520	520					No	No	No				
MB434	7	1	2	2	4	1	1	2	2	2				No	No	No				
MB445	1	1	1	2	1	1	1	130	130	5	130			No	No	No				
MB446	2	1	1	1	2	1	7	unknown	1005	5				No	No	No				
MB449	1	2	1	2	1	1	10	83	83					No	No	No				
MB463	7	1	2	2	4	1	1	2	2	2				No	No	No				
MB466	7	1	2	2	4	1	1	2	2	2				No	No	No				
MB469	7	1	2	2	4	1	1	2	2	2				No	No	No				
MB492	1	1	1	2	2	1	1	5	5	5	5S			No	No	No				
MB496	1	1	2	2	2	1	1	6	6	5				No	No	No				
MB497	8	24	unknown	23	unknown	16	unknown	(unknown)	1006					No	No	No				
MB503	7	1	2	2	4	1	1	2	2	2				No	No	No				
MB524	1	1	1	2	2	1	1	5	5	5	5S			No	No	No				
MB526	1	1	1	2	2	1	25	210	210	5				No	No	No				
MB536	1	18	6	2	2	1	1	69	69					No	No	No				
MB539	1	1	1	2	4	1	1	7	7	5				No	No	No				
MB539A	1	2	1	2	1	1	10	83	83					No	No	No				
MB546	7	1	2	2	4	1	1	2	2	2				No	No	No				
MB551	7	1	2	2	4	1	1	22	22	2				No	No	No				
MB56	1	2	1	2	1	1	10	83	83					No	No	No				
MB567	1	1	2	6	2	16	1	218	218					No	No	No				
MB585B	2	1	1	1	2	1	1	59	59	5				No	No	No				
MB591	7	1	2	2	4	1	1	2	2	2				No	No	No				
MB599	1	2	1	2	1	1	10	83	83					No	No	No				
MB601D	2	1	2	2	15	1	1	16	16	2				No	No	No				
MB606	1	1	1	2	2	1	1	5	5	5	5S			No	No	No				
MB619A	7	1	2	2	4	1	1	2	2	2				No	No	No				
MB634	1	1	1	2	2	1	1	5	5	5	5S			No	No	No				
MB714	1	18	6	2	unknown	1	1	unknown	69					No	No	No	single nucleotide variant of ST69			
MB716	1	1	1	2	2	1	1	5	5	5	5S			No	No	No				
MB724	7	1	2	2	4	7	1	22	22	2				No	No	No				
MB740	1	1	1	2	2	1	1	5	5	5	5S			No	No	No				
MB775	2	1	2	2	4	1	1	35	35	2				No	No	No				
MB821	1	2	1	2	1	1	10	83	83					No	No	No				
MB860	1	1	1	2	2	1	1	5	5	5	5S			No	No	No				
MB883	1	1	1	2	2	1	1	5	5	5	5S			No	No	No				
MB888a	2	1	2	2	15	1	1	16	16	2				No	No	No				
MB895	unknown	29	5	5	7	4	4	unknown	1007					No	No	No				
MB896	1	1	1	2	2	1	1	5	5	5	5S			No	No	No				
MB899	1	2	1	2	1	1	10	83	83					No	No	No				
MB905	1	1	1	2	2	1	1	5	5	5	5S			No	No	No				
MB906	1	18	6	2	2	1	1	69	69					No	No	No				
MB918	7	1	2	2	4	1	1	2	2	2				No	No	No				
MB941	7	1	2	2	4	1	1	2	2	2				No	No	No				
MB972	1	1	1	2	2	1	1	5	5	5	5S			No	No	No				
MB988	1	1	1	2	2	1	1	5	5	5	5S			No	No	No				
MB995	1	1	2	2	1	1	3	20	20					No	No	No				
MB999	1	2	1	2	1	1	10	83	83					No	No	No				