

# Comparative Analysis of the First Complete *Enterococcus faecium* Genome

Margaret M. C. Lam,<sup>a,f</sup> Torsten Seemann,<sup>b</sup> Dieter M. Bulach,<sup>b</sup> Simon L. Gladman,<sup>c</sup> Honglei Chen,<sup>d</sup> Volker Haring,<sup>d</sup> Robert J. Moore,<sup>d,e</sup> Susan Ballard,<sup>f</sup> M. Lindsay Grayson,<sup>f,g</sup> Paul D. R. Johnson,<sup>a,f</sup> Benjamin P. Howden,<sup>a,f,h,i</sup> and Timothy P. Stinear<sup>a,i</sup>

Department of Microbiology and Immunology, University of Melbourne, Victoria, Australia<sup>a</sup>; Victorian Bioinformatics Consortium, Monash University, Clayton, Victoria, Australia<sup>b</sup>; CSIRO Food and Nutritional Sciences, Werribee, Australia<sup>c</sup>; Australian Animal Health Laboratory, CSIRO, Geelong, Victoria, Australia<sup>d</sup>; ARC Centre of Excellence in Structural and Functional Microbial Genomics, "Clayton, Victoria," Australia<sup>e</sup>; Austin Centre for Infection Research (ACIR), Infectious Diseases Department, Austin Health, Heidelberg, Victoria, Australia<sup>f</sup>; Department of Medicine, University of Melbourne, Heidelberg, Victoria, Australia<sup>g</sup>; Microbiology Department, Austin Health, Heidelberg, Victoria, Australia<sup>h</sup>; and Department of Microbiology, Monash University, Clayton, Victoria, Australia<sup>i</sup>

**Vancomycin-resistant enterococci (VRE) are one of the leading causes of nosocomial infections in health care facilities around the globe. In particular, infections caused by vancomycin-resistant *Enterococcus faecium* are becoming increasingly common. Comparative and functional genomic studies of *E. faecium* isolates have so far been limited owing to the lack of a fully assembled *E. faecium* genome sequence. Here we address this issue and report the complete 3.0-Mb genome sequence of the multilocus sequence type 17 vancomycin-resistant *Enterococcus faecium* strain Aus0004, isolated from the bloodstream of a patient in Melbourne, Australia, in 1998. The genome comprises a 2.9-Mb circular chromosome and three circular plasmids. The chromosome harbors putative *E. faecium* virulence factors such as enterococcal surface protein, hemolysin, and collagen-binding adhesin. Aus0004 has a very large accessory genome (38%) that includes three prophage and two genomic islands absent among 22 other *E. faecium* genomes. One of the prophage was present as inverted 50-kb repeats that appear to have facilitated a 683-kb chromosomal inversion across the replication terminus, resulting in a striking replicore imbalance. Other distinctive features include 76 insertion sequence elements and a single chromosomal copy of Tn1549 containing the *vanB* vancomycin resistance element. A complete *E. faecium* genome will be a useful resource to assist our understanding of this emerging nosocomial pathogen.**

*Enterococcus faecium* is a commensal bacterium of the human gastrointestinal tract and an important nosocomial pathogen, particularly through its ability to develop resistance to multiple antibiotics, including vancomycin (9, 18). Vancomycin-resistant enterococci (VRE) were first reported in hospitals in the 1980s and have since been reported in health care settings worldwide. Resistance to vancomycin is typically mediated by acquisition of the *vanA* or *vanB* gene cluster (5, 58). Rates of acquired antibiotic resistance are significantly higher in *E. faecium* than in other enterococcal species, and together with observations of increasing *E. faecium* infections, they suggest that there is a strong correlation between antimicrobial resistance acquisition and the infectious nature of the pathogen (30).

Multilocus sequence typing (MLST) is frequently employed in molecular epidemiological analyses of *E. faecium* strains, and analysis by eBURST suggests the emergence of a lineage, termed clonal complex 17 (CC17), that appears to represent a hospital-adapted subpopulation of *E. faecium* strains, as they have been associated with clinical infections and *E. faecium* outbreaks on five continents (54, 58). Strains belonging to CC17 are proposed to possess particular traits for enhancing persistence in the health care environment, including acquiring ampicillin and quinolone resistance, and a pathogenicity island that commonly harbors the *esp* gene encoding the putative virulence factor enterococcal surface protein (6, 11, 53). A recent comparative study, aligning 100 orthologs from 21 publicly available *E. faecium* genomes, identified 5,932 single nucleotide polymorphisms (SNPs) and was used to infer a phylogeny that showed a distinct separation between hospital- and community-acquired *E. faecium* (20).

VRE first emerged in the United Kingdom and Europe in the late 1980s, 15 years after glycopeptide antibiotic use was first in-

troduced to clinics and hospitals (39). Following the first reported emergence of VRE, incidence rates have continued to increase (17). In the Australian context, VRE was first isolated in 1994 at the Austin Hospital in Victoria from a liver transplant recipient (33). This isolate was an *E. faecium* strain with *vanA*-type glycopeptide resistance. In contrast, the majority of VRE isolates in Australia today, in particular, *E. faecium*, possess *vanB*-type resistance (8, 43). Following the first Australian case in 1994, VRE was soon reported in other Australian health care facilities, with outbreaks involving VRE colonization of the hospital environment. In the past decade, the rates of VRE and vancomycin-susceptible *E. faecium* (VSE) colonization and, to a lesser extent, infection have increased dramatically Australia-wide (32, 33, 43). A recent molecular epidemiological analysis of *E. faecium* bacteremia isolates at one Australian institution demonstrated that ST17 was the predominant MLST clone prior to 2005; however, this was subsequently replaced by a new clone, ST203, which was associated with a significant increase in episodes of bacteremia (32). The bacterial and genomic factors associated with the emergence of ST203 VRE, the manner in which its genome and phenotypic behavior contrast with those of the former predominant ST17 clones, and the increase in *E. faecium* bacteremia overall are not understood, and

Received 17 February 2012 Accepted 18 February 2012

Published ahead of print 24 February 2012

Address correspondence to Timothy P. Stinear, [tstinear@unimelb.edu.au](mailto:tstinear@unimelb.edu.au).

B.P.H. and T.P.S. are co-senior authors.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.00259-12

TABLE 1 *E. faecium* genome sequences used in this study

<i>E. faecium</i> strain	Accession no. <sup>a</sup>	Comments	Reference
C68	NZ_ACJQ00000000.1	Clinical isolate, ST16, CC17; closely related to ST17; represents most prevalent VRE clone in 11 Cleveland hospitals in 1996	Broad Institute, unpublished data; strain first described in reference 12
TC 6	NZ_ACOB00000000.1	Colonizing (mouse model) transconjugant strain resulting from the mating between C68 and D344SRF	Broad Institute, unpublished data; strain first described in reference 47
D344SRF	NZ_ACZZ00000000.1	Noncolonizing (mouse model) recipient strain	Broad Institute, unpublished data
E1039	NZ_ACOS00000000.1	Fecal isolate from community surveillance program (nonhospitalized person), ST42; isolated in the Netherlands, 1998; no antimicrobial resistance	55
E1162	NZ_ABQJ00000000.1	Clinical isolate, bloodstream infection, ST17, CC17; isolated in France, 1997; high resistance to ampicillin	55
E1071	NZ_ABQI00000000.1	Fecal isolate from hospital surveillance program, hospitalized patient; ST32, not CC17; isolated in the Netherlands, 2000; <i>vanA</i> VRE	55
E1636	NZ_ABRY00000000.1	Clinical isolate, bloodstream infection. ST106; isolated in the Netherlands, 1961; low-level resistance to ampicillin	55
E1679	NZ_ABS00000000.1	Clinical isolate, from a vascular catheter; ST114; isolated in Brazil, 1998; <i>vanA</i> VRE. high resistance to ampicillin; resistant to ciprofloxacin	55
U0317	NZ_ABSW00000000.1	Clinical isolate, urinary tract infection; ST78, CC17; isolated in the Netherlands, 2005; high resistance to ampicillin; resistance to ciprofloxacin	55
DO	NZ_ACYI00000000.1	Obtained from the University of Texas Medical School, ST18	19; strain first described in reference 2
1,230,933	NZ_ACAS00000000.1	Obtained from a clinical isolate repository (Eurofins Medinet), ST18	44
1,231,502	NZ_ACAX00000000.1	Obtained from a clinical isolate repository (Eurofins Medinet), ST203	44
1,231,501	NZ_ACAY00000000.1	Obtained from a clinical isolate repository (Eurofins Medinet), ST52	44
1,231,410	NZ_ACBA00000000.1	Obtained from a clinical isolate repository (Eurofins Medinet), ST17	44
1,231,408	NZ_ACBB00000000.1	Obtained from a clinical isolate repository (Eurofins Medinet), ST582	44
TX0082	NZ_AEBU00000000.1	Clinical isolate, endocarditis/bloodstream infection, ST17; isolated in Texas, 1999; resistant to vancomycin and ampicillin	Unpublished data; strain first described in reference 41
TX0133A	NZ_AECH00000000.1	Clinical isolate, bloodstream infection; isolated in Texas, 28 March 2006; resistant to vancomycin, <i>vanA</i> , ST17	Strain first described in reference 4
TX0133B	NZ_AECI00000000.1	Clinical isolate, bloodstream infection; isolated in Texas, 9 May 2006; susceptible to vancomycin	Strain first described in reference 4
TX0133a01	NZ_AEJC00000000.1	Isolate obtained from inhibition zone surrounding vancomycin Etest strip; resistant to vancomycin	Strain first described in reference 4
TX0133a04	NZ_AEBC00000000.1	Isolate obtained from outside the inhibition zone surrounding vancomycin Etest strip; susceptible to vancomycin	Strain first described in reference 4
TX0133C	NZ_AEBG00000000.1	Clinical isolate, bloodstream infection; isolated in Texas, 26 May 2006; resistant to vancomycin, <i>vanA</i>	Strain first described in reference 4
AUS0085	SRX014391	Clinical isolate, bloodstream infection; ST203, CC17	32

<sup>a</sup> All accession numbers are from GenBank, except SRX014391, which is from SRA.

this is in part hampered by the lack of any publicly available complete *E. faecium* genome sequence.

The first partially assembled, draft genome sequence for *E. faecium* strain TX0016 (also referred to as strain DO), isolated from an endocarditis patient in 1992, was published in 2000 (19). At least 23 additional *E. faecium* strains have since been sequenced and partially assembled. In these studies, the genome data have been used to analyze the evolutionary relationships and genomic diversity between strains isolated from different sources. Recent genome-based studies have revealed significant differences (up to 12%) between the gene contents of *E. faecium* strains and have highlighted the frequency and ease with which this species acquires and loses mobile genetic elements (55, 56). The genomic plasticity observed in *E. faecium* strains is presumed to be largely responsible for the different properties exhibited by commensal and clinical isolates (36). Thus, while there are several published draft genome sequences available for *E. faecium*, no one has yet produced a fully assembled reference for this species (19, 55).

Here, we have addressed this issue by completing the genome sequence of a representative *vanB*-positive, vancomycin-resistant *E. faecium* strain (Aus0004), isolated from the bloodstream of a patient at the Austin Hospital, in Melbourne, Australia, in 1998. This strain belongs to MLST group ST17, a member of clonal complex 17, which also represents the majority of *E. faecium* strains isolated from hospitals globally (28, 58).

## MATERIALS AND METHODS

**Bacterial sequences and strains.** The *E. faecium* isolates and genome sequences used in this study are listed in Table 1.

**Genome sequencing.** The complete genome sequence of *E. faecium* Aus0004 was determined using 454 GS FLX 3-kbp paired-end sequencing (Roche Diagnostics, Basel, Switzerland), yielding 48.7 Mbp from 120,556 reads and assembled with Newbler (v2.6). Finishing was facilitated using primer walking and managed with the genome assembly program Gap4 (v1.7.1b) (10). Read mapping of our previously published Illumina paired-end reads (accession no. SRX014390) was used to correct errors in the final assembly as described previously (32). An additional 25- to 40-

kb-insert-size bacterial artificial chromosome (BAC) library was also prepared to resolve chromosomal duplications (51).

**Genome analysis.** Genome annotation was managed using in-house tools Prokka and Wasabi (<http://www.bioinformatics.net.au>). Regions containing prophage or other foreign DNA were predicted using Phast and Alien Hunter (57, 59). CRISPR loci were detected by CRISPRfinder (23). For comparative genomics, a read-mapping approach was developed to define an *E. faecium* core genome and the accessory genome for a given strain. Sequence reads from each *E. faecium* strain (Table 1) were aligned separately with the completed Aus0004 reference genome using SHRiMP 2.0 (48). Those positions in Aus0004 that were covered by at least three reads from each and every genome defined an *E. faecium* core genome. Reads from Aus0004 that did not map to any *E. faecium* genome were used to define Aus0004 regions of difference. Reads from any strain that did not map to the core genome represented that strain's accessory genome. SNPs were identified using Nsoni v0.35 (<http://www.bioinformatics.net.au>), which used the sequence reads for each genome aligned with the above-defined core genome to construct a tally of putative differences at each nucleotide position, including substitutions, insertions, and deletions. This tally was then employed in a Bayesian model to decide whether a base (or deletion) could be called for the position, and if so, whether it differed from the reference sequence. A similar procedure was used to determine the presence or absence of insertions between positions in the reference. NCBI BLASTN+, Artemis, BRIG, and Easyfig were also used for intra- and inter-Aus0004 genome comparisons (1, 13, 52). Phylogenetic analyses were performed using a distance method, based on pairwise nucleotide sequence alignments for the *E. faecium* core genome among all strains. Phylogeny was inferred by both neighbor-joining and split decomposition analyses, using uncorrected *p* distances with bootstrapping as implemented in SplitsTree4 v4.11.3 (29). Tree drawing was managed with FigTree (46).

**Nucleotide sequence accession numbers.** The complete genome sequence for *E. faecium* Aus0004 has been deposited in GenBank under accession numbers CP003351 to CP003354.

## RESULTS AND DISCUSSION

**Genome summary.** The genome of *E. faecium* strain Aus0004 comprises a single circular chromosome of 2,955,294 bp (GC content, 38%) and three circular plasmids, Aus0004\_p1, Aus0004\_p2, and Aus0004\_p3, with lengths of 56,520 bp, 4,119 bp, and 3,847 bp and GC contents of 35%, 37%, and 39%, respectively (Fig. 1A). An asymmetric chromosomal GC skew pattern was observed, indicating a recent, large DNA inversion (see later discussion) (Fig. 1A). Comparison of an *in silico* AUS0004 chromosome NcoI map against an Aus0004 NcoI optical map, generated in a previous study (32), confirmed the integrity of our final assembly (Fig. 1B). A total of 2,860 protein-coding DNA sequences (CDS), 47 tRNA genes, and 6 rRNA operons were predicted in the Aus0004 chromosome.

**Comparisons of Aus0004 with other *E. faecium* strains.** The availability of a complete *E. faecium* reference sequence allowed us to define a core *E. faecium* genome and an accessory Aus0004 genome and to produce a high-resolution *E. faecium* phylogeny. The DNA sequence reads from 22 other publicly available *E. faecium* strains of predominantly clinical origin (Table 1) were mapped to the Aus0004 chromosome and revealed that 1,866,631 of 3,019,780 bp of Aus0004 (62%), comprising 456 segments at least 200 bases long, were conserved among all 23 strains. That is, 38% of Aus0004 can be considered accessory (Fig. 1C). This is a higher percentage than that observed from comparisons of conserved CDS in *E. faecalis* with accessory genomes of around 30% (38, 49). Pairwise alignment of the 1,866,631-bp core among all 23 *E. faecium* strains uncovered 54,250 variable nucleotide positions

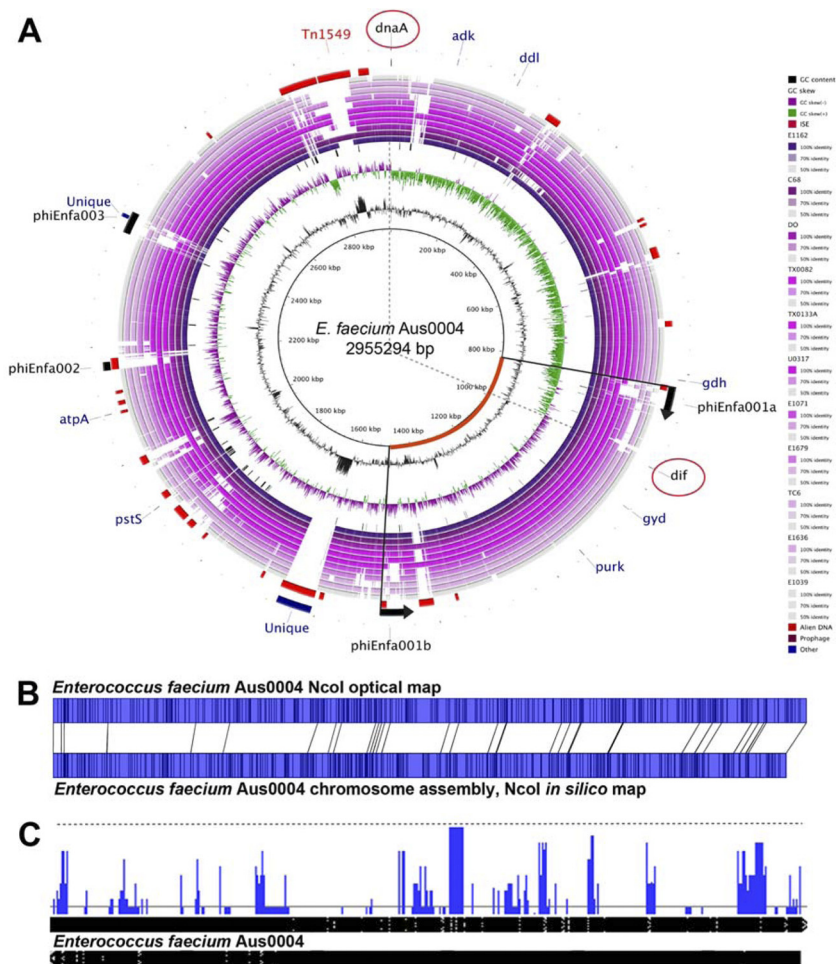
among these strains (data not shown). These comparisons were used to construct a distance matrix, and a high-resolution phylogeny was inferred by neighbor joining (Fig. 2A). Aus0004 is most closely related to 1\_231\_410, an ST17 clinical isolate (Table 1) (44). A phylogeny was also inferred with the same core genome data using split decomposition (29) (Fig. 2B). While the same relationships between strains were displayed, parallel edges with high bootstrap support linking TC6, E1636 (ST106), and E1039 (ST42) are suggestive of recombination shaping the evolution of these strains. The topologies of our trees were mostly congruent with that reported from a comparison among 100 orthologous *E. faecium* CDS using essentially the same genome sequence data (20). However, our whole-genome approach increased resolution 10-fold (54,250 versus 5,932 variable sites) and therefore offered additional discrimination among the ST17 and ST16 isolates (Fig. 2).

**Aus0004 accessory genome.** To explore accessory genome content differences among strains compared to Aus0004, we next mapped the contigs from 21 partially assembled, publicly available *E. faecium* genomes (Fig. 1A). Contig and read mapping in conjunction with DNA composition analysis (Alien Hunter) were used to identify DNA segments potentially acquired by horizontal gene transfer (HGT). In addition to detecting the Tn1549-like conjugative transposon harboring the vancomycin resistance-encoding *vanB* locus (nucleotide positions 2835604 to 2869174), we identified several pathogenicity islands and prophage (Fig. 1A). Hot spots for insertion of the Tn1549-like *vanB* transposon have previously been proposed, but the insertion site of Tn1549 in Aus0004 does not match any reported in that study (34). However, comparison of Aus0004 with strain C68, also a VRE strain, suggests that Tn1549 has been inserted into the same chromosomal site (Fig. 1A). Given the similar insertion sites, there may be several preferred target insertion sites. Adjacent to Tn1549 in *E. faecium* Aus0004 is a 13.8-kb pathogenicity island that spans nucleotide positions 2808879 to 2822670. This region carries the enterococcal surface protein-encoding *esp* gene, which is a putative virulence factor commonly present in clinical enterococcal isolates. This pathogenicity island was first described in *E. faecium* strain E300 (35), and alignment of this element with the same element from strain E300 revealed 97% overall sequence identity. An additional 60-kb genomic island, spanning nucleotides 1607500 to 1667500, was also identified. This region is not present in the other *E. faecium* strains (Fig. 1A and C), suggesting that it was recently acquired by Aus0004. The CDS within this region encode hypothetical proteins, putative transcriptional regulators, and various phage-related genes. Some of the phage elements are identical to phage components in *E. faecalis* D6 and *E. faecalis* ATCC 4200.

Three distinct prophages were present in AUS0004. Their positions are indicated in Fig. 1A, and we named them phiEnfa001a/b (duplicated), phiEnfa002, and phiEnfa003. Notably, these phages are present in all ST17 genomes, but they appear variously present in other *E. faecium* isolates (Fig. 1A). They share a high degree of similarity to phages found not only in enterococci but also in species of other genera, including *Clostridium*, *Listeria*, *Lactobacillus*, and *Staphylococcus*. The frequency and diverse origins of the mobile genetic elements seen in Aus0004 suggest that the barriers to foreign DNA acquisition are low.

Clustered, regularly interspaced short palindromic repeats (CRISPR) are observed in up to 45% of bacterial genomes and



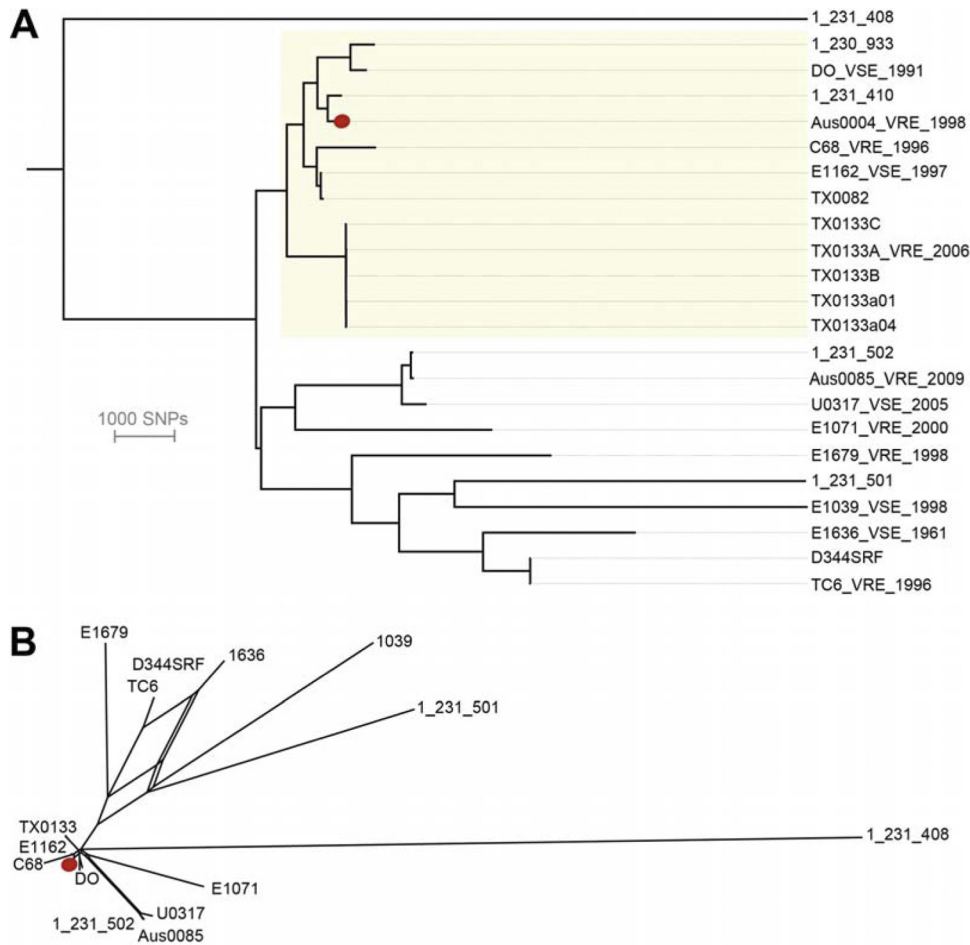


**FIG 1** Analysis of the *Enterococcus faecium* AUS0004 complete genome. (A) Circular map of AUS0004 by comparative BLASTN analysis against the contigs from the partially assembled genomes of 11 *E. faecium* strains, showing the locations of a large chromosomal inversion and Aus0004 accessory genome elements. Track identification, moving outwards, is as follows: G+C content, GC skew (G-C/G+C), IS elements, *E. faecium* isolates (next 11 tracks, as listed here and in Table 1), followed by regions (red arcs) potentially acquired by HGT revealed by AlienHunter, location of prophage, Aus004 unique regions revealed by read mapping against 23 *E. faecium* genomes (blue arcs), and finally the location of the eight MLST genes. Dotted lines indicate likely replication origin (*dnaA*) and terminus (*dif*) and highlight a replicore imbalance, caused by a phiEnfa001-mediated chromosomal inversion. (B) NcoI optical map of *E. faecium* AUS0004 compared with *in silico*-derived NcoI map demonstrating correct chromosome assembly. (C) Artemis linear view of Aus0004 chromosome and plasmids (appended), with vertical blue bars identifying the positions of accessory genome elements as determined by read mapping against 23 publicly available partially assembled genome sequences. Increasing height of vertical blue lines on this map indicates increasing specificity for Aus0004.

together with *cas* genes serve as a sequence-specific defense mechanism against the integration of exogenous foreign DNA into the bacterial genome (7, 50). Clinical *E. faecium* isolates, in particular CC17 strains are known to lack CRISPR loci, and this is thought to explain their increased propensity to acquire exogenous DNA compared with that of *E. faecium* strains from nonclinical environments (55). A CRISPR-*cas* system was not detected in *E. faecium* Aus0004. However, two paralogous CDS (EFAU004\_00823 and EFAU004\_01436) encoding a protein homologous to a CRISPR-associated protein (encoded by *cas2*) were identified within phiEnfa001 and phiEnfa002. However, without the other required *cas* genes, spacers, and repeat sequences that comprise the CRISPR-*cas* system, the two paralogous CDS most likely bear no functional significance.

Phage are major contributors to genome plasticity and have contributed to substantial chromosome remodeling in Aus0004. Prophage phiEnfa001a and phiEnfa001b are present as 50-kb inverted

repeats (Fig. 1A). Their locations align with an asymmetrical GC skew pattern that suggests they have facilitated a 683-kb chromosomal inversion that crosses the predicted replication terminus, resulting in a significant replicore imbalance (Fig. 1A). We confirmed the likely location of the replication terminus in Aus0004 by identifying a single *dif* sequence (ACTTTGTATAATATATATTATGTAAACT, positions 964426 to 964453) that aligns with the shift in GC skew (Fig. 1A). The *dif* motif is strongly associated with replication termini (27). It is not known how widespread this rearrangement is among *E. faecium* isolates, because chromosomes from other isolates have not been sufficiently well assembled, as the inversion occurs around a large duplication that is unlikely to be resolved by automated assembly processes. However, a normal GC skew pattern, with balanced replicores, is found in *E. faecalis* V583. Replicore imbalances have been proposed to cause a fitness cost (37). Simple growth curve analysis of Aus0004 in the absence of competition shows no growth defect (data not shown).



**FIG 2** (A) Neighbor-joining tree showing relationship of AUS0004 compared to 22 other publicly available *E. faecium* genome sequences. Phylogeny was inferred by alignment of the 54,250 variable nucleotide positions (gaps removed) among all 23 *E. faecium* strains. Strains belonging to ST17 are highlighted. Red dot indicates Aus0004. *E. faecium* 1\_231\_408 was used as an outgroup to root the tree. (B) Network structure revealed among the same strains by split decomposition analysis of the same data set. All nodes in both trees had greater than 90% bootstrap support (1,000 replicates).

**CDS associated with virulence.** Genes encoding a number of putative enterococcal virulence factors were identified, including hemolysin (EFAU004\_01337). Hemolysin is expressed in many *E. faecium* bacteremia isolates, and its cytolytic properties permit the lysis of a wide range of cells at the primary site of infection in the human host and subsequently enable the release of the isolate into the bloodstream (15). Also present were the genes for enterococcal surface protein (*esp*; EFAU004\_02750) and collagen-binding adhesin (*acm*; EFAU004\_02292). Enterococcal surface protein and collagen-binding adhesin reportedly promote colonization and hence the establishment of infection, although a number of recent studies have demonstrated that enterococcal surface protein is not essential for infection in murine infection models (26, 42, 45). Genes encoding other putative enterococcal virulence factors commonly associated with *E. faecalis* or *E. faecium* blood culture isolates, such as gelatinase, aggregation substance, lipase, and hemagglutinin (16), are absent in *E. faecium* strain Aus0004. As previously reported, Aus0004 does not contain hyaluronidase (*hyl*) (32).

**Insertion sequence elements.** The Aus0004 genome is rich with repetitive DNA, highlighted by the presence of multiple insertion sequence elements (ISEs). At least 21 different ISEs were

detected, ranging in copy number from 1 to 13, representing 76 distinct copies and distributed around the chromosome (Fig. 1A) and plasmids (Table 2). The most frequently observed ISE types were from the ISL3 and ISEf1 families. Notably, ISEf1 is also common in *E. faecalis* strains and was first identified as a predominant ISE type in the sequenced *E. faecalis* strain V583. Interestingly, none of the ISEs directly disrupted any CDS, although impacts on gene expression through ISE promoter adjacency may be occurring.

Plasmids *Enterococcus* species have been reported to contain several plasmids that often mediate resistance to antimicrobials and particular heavy metals and contribute to enhanced virulence and/or DNA repair mechanisms (3, 21, 22, 24). The plasmid copy number per cell was estimated by observing Illumina sequence read coverage distribution and indicated 1 copy of Aus0004\_p1, 24 copies of Aus0004\_p2, and 86 copies of Aus0004\_p3. Comparison of their *repA* genes indicates that Aus0004\_p1 belongs to the rep17 family while Aus0004\_p2 and Aus0004\_p3 belong to the rep14 family (31). Plasmid Aus0004\_p1 harbors genes that encode a toxin-antitoxin system, a conjugation system, and several proteins with no known function. Toxin-antitoxin systems are frequently reported in *E. faecium* strains (40), and the Aus0004 chro-

TABLE 2 Insertion sequences identified in *E. faecium* Aus0004

Insertion sequence	No. of copies	Locus tag(s) in Aus0004 <sup>a</sup>
ISEnfa3	4	EFAU004_00019+00020, EFAU004_01902+01903+01904, EFAU004_01898+01899+01900, EFAU004_00210
ISEfa11	8	EFAU004_00021, EFAU004_00594+00595, EFAU004_00659, EFAU004_01043+01044, EFAU004_01114, EFAU004_01490, EFAU004_01891, EFAU004_01919
ISEf1	13	EFAU004_00043, EFAU004_00157+00158, EFAU004_00187, EFAU004_00318, EFAU004_01145, EFAU004_01263, EFAU004_01527+01528+01529, EFAU004_01764+01765, EFAU004_01797, EFAU004_01901, EFAU004_01984, EFAU004_02700, EFAU004_02709
ISEfa10	1	EFAU004_00477
ISEfa7	5	EFAU004_00580, EFAU004_01762, EFAU004_02708, EFAU004_01780, EFAU004_01781
ISEfa8	3	EFAU004_00848+00849+00850, EFAU004_01409+01410+01411, EFAU004_01840+01841
ISEfm1	6	EFAU004_00289, EFAU004_02446, EFAU004_00867, EFAU004_00300, EFAU004_01982, EFAU004_01779
IS16	3	EFAU004_01890, EFAU004_02358, EFAU004_02683
IS1380	1	EFAU004_01897
IS1476	1	EFAU004_01884
IS1485	2	EFAU004_02218+02219, EFAU004_02710+02711
IS6770	2	EFAU004_01022, EFAU004_01956
ISL3 family	1	EFAU004_01362
IS4 family	1	EFAU004_01844
IS66 family	6	EFAU004_00014, EFAU004_01985, EFAU004_01825, EFAU004_01957, EFAU004_01958, EFAU004_01589+01590
IS116 family	6	EFAU004_01259, EFAU004_01518, EFAU004_01893, EFAU004_01894, EFAU004_02174, EFAU004_02661
IS1251-like	8	EFAU004_00913, EFAU004_01298, EFAU004_01364, EFAU004_01404, EFAU004_01760, EFAU004_01885, EFAU004_02480, EFAU004_02481
IS200/IS605 family	1	EFAU004_02276
Transposase-like	1	EFAU004_02756
Transposase-like	2	EFAU004_02757, EFAU004_02758
Transposase-like	1	EFAU004_02811

<sup>a</sup> Locus tags linked by "+" represent multiple transposases belonging to a single IS.

mosome additionally harbors at least five such systems. Their presence is thought to ensure the maintenance of the plasmid in the bacterial population (25, 40). Plasmids Aus0004\_p2 and Aus0004\_p3 each have shared segments of high DNA sequence similarity, 88% and 97%, respectively, with previously described 6,038-bp cryptic plasmid pRI1 identified in isolates of *E. faecium* from various animals and human origin (21). Excluding those of the replication initiation protein and plasmid recombination enzyme, the functions of the proteins encoded by the other CDS present on the plasmids are unknown. Although an ap-

parent biological role appears to be lacking, filter-mating experiments have demonstrated the cotransfer of plasmid pRI1 with antimicrobial resistance-encoding plasmids, suggesting that cryptic plasmids may play some role in the transfer of antimicrobial resistance between bacterial strains (21).

**MLST gene distribution.** We plotted the locations of the seven genes used for MLST analysis of *E. faecium*. Their distribution around the chromosome is shown in Fig. 1A and summarized in Table 3. We observed variations in the sequence of published primer sequences versus Aus0004 target sequences for *gyd* and

TABLE 3 Analysis of MLST loci in *E. faecium* Aus0004

Gene name	Locus tag in Aus0004	MLST PCR product <sup>a</sup>			Published length (bp) (28)	Oligonucleotide primer sequence differences
		Locus start	Locus end	Length in Aus0004 (bp)		
<i>adk</i>	EFAU004_00116	108013	108542	529	437	None
<i>atpA</i>	EFAU004_02059	2086115	2086768	653	556	None
<i>ddl</i>	EFAU004_00216	210951	211494	543	465	None
<i>gdh</i>	EFAU004_00777	810740	811399	659	530	gdh1: GCGCACT- <u>AAAGAT</u> ATGGT Aus004: GCGCACT <u>TAAAGAC</u> ATGGT gdh2: CCAAGATTG <u>GCA</u> ACTTCTGCCA Aus004: CCAAGATTG <u>AGCG</u> ACTTCTGCCA
<i>gyd</i>	EFAU004_01020	1042017	1042396	379	395	gyd1: CAAACTG <u>CTTAGC</u> TCCAATGGC Aus004: CAAACTG <u>TTTAGC</u> ACCTATGGC gyd2: <u>CATTT</u> CGTTGTGCAT <u>ACCA</u> AGC Aus004: <u>GATTT</u> GACGGTCAT <u>CATAC</u> CC
<i>purK</i>	EFAU004_01116	1145293	1145951	658	492	None
<i>pstS</i>	EFAU004_01886	1910865	1911495	630	583	None

<sup>a</sup> Underlining indicates sequence differences between the oligonucleotide primers used for MLST and the genome sequence of Aus0004.

*gdh*, as well as differences in the predicted PCR product sizes for various loci (Table 3).

**Conclusion.** While partially assembled draft genomes have provided some important insights into the diversity and evolution of *E. faecium*, complete and closed genomes will further exploit the capacity of genomics to understand this important human pathogen. The complete *E. faecium* Aus0004 genome has enabled us to examine the architecture of the genome and more precisely describe the genomic features of this strain. Using these data we have concisely defined an *E. faecium* core genome, and we also describe the genomic relationships between all the publicly available partially assembled *E. faecium* genomes. We have highlighted an abundance of mobile genetic elements, including multiple prophage and insertion sequence elements, plasmids, two large genomic islands, and the location of the Tn1549-like vancomycin resistance-encoding transposon, in our sequenced strain. As with *E. faecalis*, a significant portion of Aus0004 is comprised of mobile DNA that has led to the accumulation of a relatively large accessory genome (38%). In comparison, *Staphylococcus aureus* has an accessory genome of 11 to 18% among diverse strains (14). Further comparative and functional genomic research is required to pinpoint specific genetic features that explain the emergence of hospital epidemic and endemic *E. faecium* strains such as Aus0004. The complete *E. faecium* Aus0004 genome sequence described here represents an important resource in our efforts to understand this species and control the emergence of yet another hospital superbug.

#### ACKNOWLEDGMENTS

We thank Elizabeth Grabsch for assistance with characterization of the strain.

The National Health and Medical Research Council (NHMRC) of Australia supported this work.

#### REFERENCES

- Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402.
- Arduino RC, Murray BE, Rakita RM. 1994. Roles of antibodies and complement in phagocytic killing of enterococci. *Infect. Immun.* 62:987–993.
- Arias CA, Panesso D, Singh KV, Rice LB, Murray BE. 2009. Cotransfer of antibiotic resistance genes and a hylEfm-containing virulence plasmid in *Enterococcus faecium*. *Antimicrob. Agents Chemother.* 53:4240–4246.
- Arias CA, et al. 2007. Failure of daptomycin monotherapy for endocarditis caused by an *Enterococcus faecium* strain with vancomycin-resistant and vancomycin-susceptible subpopulations and evidence of in vivo loss of the *vanA* gene cluster. *Clin. Infect. Dis.* 45:1343–1346.
- Arthur M, Reynolds P, Courvalin P. 1996. Glycopeptide resistance in enterococci. *Trends Microbiol.* 4:401–407.
- Baldassarri L, Bertuccini L, Ammendolia MG, Gherardi G, Creti R. 2001. Variant *esp* gene in vancomycin-sensitive *Enterococcus faecium*. *Lancet* 357:1802.
- Barrangou R, et al. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315:1709–1712.
- Bell JM, Paton JC, Turnidge J. 1998. Emergence of vancomycin-resistant enterococci in Australia: phenotypic and genotypic characteristics of isolates. *J. Clin. Microbiol.* 36:2187–2190.
- Billstrom H, Lund B, Sullivan A, Nord CE. 2008. Virulence and antimicrobial resistance in clinical *Enterococcus faecium*. *Int. J. Antimicrob. Agents* 32:374–377.
- Bonfield JK, Smith K, Staden R. 1995. A new DNA sequence assembly program. *Nucleic Acids Res.* 23:4992–4999.
- Bonten MJ, Willems R, Weinstein RA. 2001. Vancomycin-resistant enterococci: why are they here, and where do they come from? *Lancet Infect. Dis.* 1:314–325.
- Carias LL, Rudin SD, Donskey CJ, Rice LB. 1998. Genetic linkage and cotransfer of a novel, *vanB*-containing transposon (Tn5382) and a low-affinity penicillin-binding protein 5 gene in a clinical vancomycin-resistant *Enterococcus faecium* isolate. *J. Bacteriol.* 180:4426–4434.
- Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28:464–469.
- Chua KY, et al. 2011. The dominant Australian community-acquired methicillin-resistant *Staphylococcus aureus* clone ST93-IV [2B] is highly virulent and genetically distinct. *PLoS One* 6:e25887.
- Cox CR, Coburn PS, Gilmore MS. 2005. Enterococcal cytolysin: a novel two component peptide system that serves as a bacterial defense against eukaryotic and prokaryotic cells. *Curr. Protein Pept. Sci.* 6:77–84.
- Elsner HA, et al. 2000. Virulence factors of *Enterococcus faecalis* and *Enterococcus faecium* blood culture isolates. *Eur. J. Clin. Microbiol. Infect. Dis.* 19:39–42.
- Endtz HP, van den Braak N, Verbrugh HA, van Belkum A. 1999. Vancomycin resistance: status quo and quo vadis. *Eur. J. Clin. Microbiol. Infect. Dis.* 18:683–690.
- Fisher K, Phillips C. 2009. The ecology, epidemiology and virulence of *Enterococcus*. *Microbiology* 155:1749–1757.
- Fox J. 2000. DOE Joint Genome Institute feat: one-day sequencing of *E. faecium*. *ASM News* 66:388.
- Galloway-Pena J, Roh JH, Latorre M, Qin X, Murray BE. 2012. Genomic and SNP analyses demonstrate a distant separation of the hospital and community-associated clades of *Enterococcus faecium*. *PLoS One* 7:e30187.
- Garcia-Migura L, Hasman H, Jensen LB. 2009. Presence of pRI1: a small cryptic mobilizable plasmid isolated from *Enterococcus faecium* of human and animal origin. *Curr. Microbiol.* 58:95–100.
- Garcia-Migura L, Liebana E, Jensen LB. 2007. Transposon characterization of vancomycin-resistant *Enterococcus faecium* (VREF) and dissemination of resistance associated with transferable plasmids. *J. Antimicrob. Chemother.* 60:263–268.
- Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35:W52–W57.
- Hasman H, et al. 2006. Copper resistance in *Enterococcus faecium*, mediated by the *trcB* gene, is selected by supplementation of pig feed with copper sulfate. *Appl. Environ. Microbiol.* 72:5784–5789.
- Hayes F. 2003. Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. *Science* 301:1496–1499.
- Heikens E, et al. 2009. Enterococcal surface protein Esp is not essential for cell adhesion and intestinal colonization of *Enterococcus faecium* in mice. *BMC Microbiol.* 9:19.
- Hendrickson H, Lawrence JG. 2007. Mutational bias suggests that replication termination occurs near the *dif* site, not at Ter sites. *Mol. Microbiol.* 64:42–56.
- Homan WL, et al. 2002. Multilocus sequence typing scheme for *Enterococcus faecium*. *J. Clin. Microbiol.* 40:1963–1971.
- Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23:254–267.
- Iwen PC, et al. 1997. Change in prevalence and antibiotic resistance of *Enterococcus* species isolated from blood cultures over an 8-year period. *Antimicrob. Agents Chemother.* 41:494–495.
- Jensen LB, et al. 2010. A classification system for plasmids from enterococci and other Gram-positive bacteria. *J. Microbiol. Methods* 80:25–43.
- Johnson PD, et al. 2010. A sustained hospital outbreak of vancomycin-resistant *Enterococcus faecium* bacteremia due to emergence of *vanB E. faecium* sequence type 203. *J. Infect. Dis.* 202:1278–1286.
- Kamarulzaman A, Tosolini FA, Boquest J, Geddes JE, Richards MJ. 1995. Vancomycin-resistant *Enterococcus faecium* in a liver transplant recipient. *ANZ J. Med.* 25:560.
- Launay A, Ballard SA, Johnson PD, Grayson ML, Lambert T. 2006. Transfer of vancomycin resistance transposon Tn1549 from *Clostridium symbiosum* to *Enterococcus* spp. in the gut of gnotobiotic mice. *Antimicrob. Agents Chemother.* 50:1054–1062.
- Leavis H, et al. 2004. A novel putative enterococcal pathogenicity island linked to the *esp* virulence gene of *Enterococcus faecium* and associated with epidemicity. *J. Bacteriol.* 186:672–682.
- Leavis HL, et al. 2007. Insertion sequence- and gene-driven diversification creates a



- globally dispersed emerging multiresistant subspecies of *E. faecium*. *PLoS Pathog.* 3:e7.
37. Matthews TD, Maloy S. 2010. Fitness effects of replicore imbalance in *Salmonella enterica*. *J. Bacteriol.* 192:6086–6088.
  38. McBride SM, Fischetti VA, Leblanc DJ, Moellering RC, Jr, Gilmore MS. 2007. Genetic diversity among *Enterococcus faecalis*. *PLoS One* 2:e582.
  39. Metan G, Zarakolu P, Unal S. 2005. Rapid detection of antibacterial resistance in emerging Gram-positive cocci. *J. Hosp. Infect.* 61:93–99.
  40. Moritz EM, Hergenrother PJ. 2007. Toxin-antitoxin systems are ubiquitous and plasmid-encoded in vancomycin-resistant enterococci. *Proc. Natl. Acad. Sci. U. S. A.* 104:311–316.
  41. Nallapareddy SR, Singh KV, Murray BE. 2006. Construction of improved temperature-sensitive and mobilizable vectors and their use for constructing mutations in the adhesin-encoding *acm* gene of poorly transformable clinical *Enterococcus faecium* strains. *Appl. Environ. Microbiol.* 72:334–345.
  42. Nallapareddy SR, Weinstock GM, Murray BE. 2003. Clinical isolates of *Enterococcus faecium* exhibit strain-specific collagen binding mediated by *Acm*, a new member of the MSCRAMM family. *Mol. Microbiol.* 47:1733–1747.
  43. Padiglione AA, et al. 2003. Risk factors for new detection of vancomycin-resistant enterococci in acute-care hospitals that employ strict infection control procedures. *Antimicrob. Agents Chemother.* 47:2492–2498.
  44. Palmer KL, et al. 2010. High-quality draft genome sequences of 28 *Enterococcus* sp. isolates. *J. Bacteriol.* 192:2469–2470.
  45. Pultz NJ, Shankar N, Baghdayan AS, Donskey CJ. 2005. Enterococcal surface protein Esp does not facilitate intestinal colonization or translocation of *Enterococcus faecalis* in clindamycin-treated mice. *FEMS Microbiol. Lett.* 242:217–219.
  46. Rambaut A. 2009. FigTree, tree figure drawing tool, v. 1.3.1.
  47. Rice LB, et al. 2009. Transferable capacity for gastrointestinal colonization in *Enterococcus faecium* in a mouse model. *J. Infect. Dis.* 199:342–349.
  48. Rumble SM, et al. 2009. SHRiMP: accurate mapping of short color-space reads. *PLoS Comput. Biol.* 5:e1000386.
  49. Solheim M, Aakra A, Snipen LG, Brede DA, Nes IF. 2009. Comparative genomics of *Enterococcus faecalis* from healthy Norwegian infants. *BMC Genomics* 10:194.
  50. Sorek R, Kunin V, Hugenholtz P. 2008. CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.* 6:181–186.
  51. Stinear TP, et al. 2004. Giant plasmid-encoded polyketide synthases produce the macrolide toxin of *Mycobacterium ulcerans*. *Proc. Natl. Acad. Sci. U. S. A.* 101:1345–1349.
  52. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer. *Bioinformatics* 27:1009–1010.
  53. Top J, Willems R, Bonten M. 2008. Emergence of CC17 *Enterococcus faecium*: from commensal to hospital-adapted pathogen. *FEMS Immunol. Med. Microbiol.* 52:297–308.
  54. Top J, Willems R, van der Velden S, Asbroek M, Bonten M. 2008. Emergence of clonal complex 17 *Enterococcus faecium* in the Netherlands. *J. Clin. Microbiol.* 46:214–219.
  55. van Schaik W, et al. 2010. Pyrosequencing-based comparative genome analysis of the nosocomial pathogen *Enterococcus faecium* and identification of a large transferable pathogenicity island. *BMC Genomics* 11:239.
  56. van Schaik W, Willems R. 2010. Genome-based insights into the evolution of enterococci. *Clin. Microbiol. Infect.* 16:527–532.
  57. Vernikos GS, Parkhill J. 2006. Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics* 22:2196–2203.
  58. Willems RJ, et al. 2005. Global spread of vancomycin-resistant *Enterococcus faecium* from distinct nosocomial genetic complex. *Emerg. Infect. Dis.* 11:821–828.
  59. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. *Nucleic Acids Res.* 39:W347–W352.