Evolutionary origins of the emergent ST796 clone of vancomycin resistant *Enterococcus faecium*

Andrew H Buultjens [1], **Margaret M C Lam** [1], **Susan Ballard** [2], **Ian R Monk** [1], **Andrew A Mahony** [3], **Elizabeth A Grabsch** [3], **M Lindsay Grayson** [3], **Stanley Pang** [4,5], **Geoffrey W Coombs** [4,5], **J Owen Robinson** [4,5], **Torsten Seemann** [6], **Paul D R Johnson** [Corresp., 3,7], **Benjamin P Howden** [Corresp., 2], **Timothy P Stinear** [Corresp. 1]

[1] Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, Victoria, Australia

[2] Microbiology Diagnostic Unit, Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, Victoria, Australia

[3] Infectious Diseases Department, Austin Health, Heidelberg, Victoria, Australia

[4] School of Veterinary and Life Sciences, Murdoch University, Murdoch, Western Australia, Australia

[5] Department of Microbiology, Fiona Stanley Hospital, Murdoch, Western Australia, Australia

[6] Victorian Life Sciences Computation Initiative, University of Melbourne, Carlton, Victoria, Australia

[7] Department of Medicine, University of Melbourne, Heidelberg, Victoria, Australia

Corresponding Authors: Paul D R Johnson, Benjamin P Howden, Timothy P Stinear
Email address: Paul.Johnson@austin.org.au, bhowden@unimelb.edu.au, tstinear@unimelb.edu.au

From early 2012, a novel clone of vancomycin resistant *Enterococcus faecium* (assigned the multi locus sequence type ST796) was simultaneously isolated from geographically separate hospitals in south eastern Australia and New Zealand. Here we describe the complete genome sequence of Ef_aus0233, a representative ST796 *E. faecium* isolate. We used PacBio single molecule real-time sequencing to establish a high quality, fully assembled genome comprising a circular chromosome of 2,888,087 bp and five plasmids. Comparison of Ef_aus0233 to other *E. faecium* genomes shows Ef_aus0233 is a member of the epidemic hospital-adapted lineage and has evolved from an ST555-like ancestral progenitor by the accumulation or modification of five mosaic plasmids and five putative prophage, acquisition of two cryptic genomic islands, accrued chromosomal single nucleotide polymorphisms and a 80kb region of recombination, also gaining Tn*1549* and Tn*916*, transposons conferring resistance to vancomycin and tetracycline respectively. The genomic dissection of this new clone presented here underscores the propensity of the hospital *E. faecium* lineage to change, presumably in response to the specialized conditions of hospital and healthcare environments.

Evolutionary origins of the emergent ST796 clone of vancomycin resistant *Enterococcus faecium*

Andrew H. Buultjens [1], Margaret M. C. Lam [1], Susan Ballard [2,3], Ian. R. Monk [1], Andrew A. Mahony [3], Elizabeth A. Grabsch [3], M. Lindsay Grayson [3], Stanley Pang[4,5], Geoffrey W. Coombs [4,5], J. Owen Robinson [4,5], Torsten Seemann [6], Paul D. R. Johnson [1, 3, 7, #], Benjamin P. Howden [1, 2, 7, #] and Timothy P. Stinear [1, #]

Address: [1] Department of Microbiology and Immunology, Doherty Institute for Infection and Immunity, The University of Melbourne, Victoria, 3010, Australia; [2] Microbiology Diagnostic Unit, Department of Microbiology and Immunology, Doherty Institute for Infection and Immunity, The University of Melbourne, Victoria, 3010, Australia; [3] Infectious Diseases Department, Austin Health, Heidelberg, Victoria, 3084, Australia; [4] School of Veterinary and Life Sciences, Murdoch University, Western Australia, 6150, Australia; [5] Department of Microbiology, PathWest Laboratory Medicine – WA, Royal Perth Hospital, Fiona Stanley Hospital, Murdoch, Western Australia, 6150, Australia; [6] Victorian Life Sciences Computation Initiative, Carlton, Victoria, 3053, Australia; [7] Department of Medicine, University of Melbourne, Heidelberg, Victoria, 3084, Australia.

[#]Joint senior authors

ABSTRACT

From early 2012, a novel clone of vancomycin resistant *Enterococcus faecium* (assigned the multi locus sequence type ST796) was simultaneously isolated from geographically separate hospitals in south eastern Australia and New Zealand. Here we describe the complete genome sequence of Ef_aus0233, a representative ST796 *E. faecium* isolate. We used PacBio single molecule real-time sequencing to establish a high quality, fully assembled genome comprising a circular chromosome of 2,888,087 bp and five plasmids. Comparison of Ef_aus0233 to other *E. faecium* genomes shows Ef_aus0233 is a member of the epidemic hospital-adapted lineage and has evolved from an ST555-like ancestral progenitor by the accumulation or modification of five mosaic plasmids and five putative prophage, acquisition of two cryptic genomic islands, accrued

chromosomal single nucleotide polymorphisms and a 80kb region of recombination, also gaining Tn*1549* and Tn*916*, transposons conferring resistance to vancomycin and tetracycline respectively. The genomic dissection of this new clone presented here underscores the propensity of the hospital *E. faecium* lineage to change, presumably in response to the specific conditions of hospital and healthcare environments.

INTRODUCTION

Enterococcus faecium is a human and animal gastrointestinal tract (GIT) commensal but a lineage within the species has rapidly evolved to become a significant opportunistic pathogen (Coombs et al. 2014c; Deshpande et al. 2007; Lebreton et al. 2013b; Pinholt et al. 2014).

Early genotyping methods such as amplified fragment length polymorphism (Willems et al. 2000), restriction endonuclease analysis (Quednau et al. 1999), multi-locus sequence typing (MLST) (Leavis et al. 2006; Top et al. 2008) and more recent analyses using whole genome datasets (Lebreton et al. 2013), have shown the *E. faecium* population separates into two major clades largely according to source origin, designated as clades A and B. Clade B strains are community-associated and mostly of non-clinical origin while clade A strains are hospital-associated and mostly of clinical origin (Galloway-Peña et al. 2012; Leavis et al. 2007; Palmer et al. 2012). Clade A has been found to further divide into clade A1, which contains epidemic hospital strains, and clade A2, which encompasses animal strains and strains linked to sporadic human infections (Lebreton et al. 2013).

Clade A1 or clonal complex 17 (CC17, a MLST designation) has adapted to the hospital environment and is adept at GIT colonization with the potential to cause invasive disease (Top et al. 2008; Willems et al. 2005). Members of clade A1 are characterised by larger genomes and harbour a greater abundance of virulence factors and genes conferring antibiotic resistance compared to non-A1 *E. faecium* lineages, a reflection of adaptation to healthcare environments (Galloway-Peña et al. 2012; Guzman Prieto et al. 2016; Top et al. 2008).

In Australia, as in other countries, we have observed the sequential emergence of new *E. faecium* clones within the clade A1 hospital lineage which spread rapidly and displace previously endemic clones. For example, from 1994 to 2005, Australian hospital acquired *E. faecium* VRE was uncommon and mostly caused by ST17 strains. The situation changed suddenly from 2005 when there was a nationwide wave of by *E. faecium* ST203 blood stream infections (BSI), a significant and rising proportion of which are *vanB* VRE (Coombs et al. 2014; Johnson et al. 2010; Lam et al. 2012). Previous work comparing ST17 and ST203 genomes revealed that ST203 possesses 40 unique genes with inferred functions of riboflavin metabolism, ion transport and phosphorylation, and harboured a larger vancomycin resistance-conferring Tn*1549* transposon (Lam et al. 2013).

At the Austin Hospital in Melbourne, improved cleaning protocols following our local ST203 outbreak were associated with a reduction in VRE BSI between 2009 and 2011 (Grabsch et al. 2012). However, despite retaining these protocols we once again observed an abrupt increase in *vanB* VRE *E. faecium* BSI from 2012 onwards that was caused by a completely new ST. We originally recognised the change in strain using PFGE and a high-resolution melt method (Tong et al. 2011) but have now switched to whole genome sequencing for epidemiological typing. We lodged the alleles of the new ST with the MLST Database and received the new designation ST796 in September 2012 (Mahony et al. 2014). ST796 was unknown before 2011 but by 2013 *vanB* ST796 *E. faecium* had caused a large outbreak of colonisation in a Melbourne Neonatal Intensive Care (Lister et al. 2015) and in the same year was responsible for 40% of *E. faecium* VRE BSI in 5 geographically separate Melbourne hospitals, largely replacing its ST203 predecessor strains. In 2015, ST796 *vanB E. faecium* was responsible for 62 of 117 (53%) of all patient episodes of all *E. faecium* bacteraemia in Melbourne Hospitals, compared with 10 of 117 (8.5%) for ST203.

In the current study, we used single molecule real-time sequencing to establish a high quality, fully assembled genome sequence of ST796 *E. faecium* isolate Ef_aus0233, a representative of

this emerging clone and then employed population based comparative genomics to better understand the genetic changes that have accompanied the emergence.

METHODS

Bacterial strains. A list of the isolates examined in the study is provided (Table S1). *E. faecium* were cultured as previously described (Johnson et al. 2010).

Whole Genome Sequencing. Short fragment DNA libraries were generated using the Illumina NexteraXT DNA preparation kit and fragment sequencing was undertaken with the Illumina NextSeq 500 platform using 2x150 bp chemistry. Highly intact and high quality genomic DNA was extracted from Ef_aus0233 and subjected to Pacific Biosciences SMRT sequencing according to the manufacturer's instructions and sequenced with two SMRT cells on the RS II platform (Pacific Biosciences) using P5-C3 chemistry. Genome assembly was performed using the SMRT Analysis System v2.3.0.140936 (Pacific Biosciences). Raw sequence data were *de novo* assembled using the HGAP v3 protocol with a genome size of 3 Mb. Polished contigs were error corrected using Quiver v1. The resulting assembly was then checked using BridgeMapper v1 in the SMRT Analysis System, and the consensus sequence corrected with short-read Illumina data, using the program Snippy (https://github.com/tseemann/snippy). The final chromosome assembly was validated by reference to a high-resolution NcoI optical map using MapSolver (version 3.10; OpGen, Maryland USA). Common bacterial DNA base modifications and methyltransferase motifs were assessed using the protocol, RS_Modification_and_Motif_Analysis in the SMRT Analysis System v2.3.0.140936 (Pacific Biosciences).

Plasmid Copy Number. The approximate number of plasmid copies per cell for the Ef_aus0233 genome was inferred using differences in Illumina sequence read depth. The read depth of plasmid sequences was compared to the average chromosomal coverage to estimate copy number multiplicity.

Comparison of completed genomes. Artemis Comparison Tool (Carver et al. 2005) was used to align the chromosomes of four fully assembled *E. faecium* genomes. BRIG (Alikhan et al. 2011) was used to visualize DNA:DNA comparisons using BLASTn for comparisons of the Ef_aus0233 chromosome against other fully assembled *E. faecium* chromosomes.

***De novo* assembly and genome annotation.** Illumina sequence reads were *de novo* assembled into contigs using Spades v3.6.1 (Nurk et al. 2013). The closed Ef_aus0233 genome and Spades contigs were annotated using Prokka (v1.12b) (Seemann 2014). Multilocus sequence types (STs) were determined using an *in silico* tool (https://github.com/tseemann/mlst). CRISPR databases were used to search for CRISPR sequences (http://crispi.genouest.org and http://crispr.u-psud.fr/Server/) (accessed 19th of May 2016). Sequence files were uploaded to the web based ISsaga (Varani et al. 2011)(accessed 11th of February 2016) to detect both the abundance and diversity of insertion elements. Phage discovery was undertaken using the web based resource PHAST (accessed 15th of February 2016) (Zhou et al. 2011).

Variant detection and Bayesian population clustering. Snippy was used to map short read data against the full-assembled Ef_aus0233 genome to call core genome single nucleotide polymorphism (SNP) differences. Hierarchical Bayesian clustering was performed upon a core SNP alignment to assign genomes into discrete populations using hierBAPS with BAPS6 (a prior of 10 depth levels and a maximum of 20 clusters were specified) (Cheng et al. 2013). Nested clustering analyses were undertaken upon subsets of the original SNP alignment to a total depth of three levels or until no further clustering could be achieved.

Recombination and phylogenomic analysis. Recombination within the core genome was inferred using ClonalFrameML v1.7 (Didelot & Wilson 2015) using the whole genome alignment generated by Snippy. The ML tree generated with FastTree v2.1.8 was used as a guide tree for ClonalFrameML. Positions in the reference genome that were not present in at least one genome (non-core) were omitted from the analysis using the “ignore_incomplete_sites true”

option and providing ClonalFrameML with a list of all non-core positions. Maximum likelihood trees with bootstrap support were constructed using a recombination free SNP alignment with the program FastTree (Price et al. 2010). Bootstrap support was derived from comparisons between the original tree against 1000 trees that were built upon pseudo-alignments (sampled from the original alignment with replacement).

Pan genome analysis. Orthologous proteins were identified through reciprocal blast using Proteinortho5 v5.11 (Lechner et al. 2011). A blast cutoff of 95% identity and alignment coverage of 30% were used. The resulting matrix of ortholog presence and absence was visualized using Fripan (https://github.com/drpowell/FriPan) (downloaded on the 28th of April 2016). The matrix of pan genome content has been deposited in Figshare (10.6084/m9.figshare.405479, and 10.6084/m9.figshare.4007760).

Sequence alignment and visualization. The alignment of homologous sequences was undertaken using Mauve (Darling et al. 2004). Sequences and alignments were visualized using Geneious Pro (version 8.1.8, Biomatters Ltd. [www.geneious.com]).

RESULTS AND DISCUSSION

Genome overview. Assembly of the 158,885 sequence reads from PacBio SMRT sequencing of Ef_aus0233 (N50 read length 8,952 bp) resulted in reconstruction of a 3,272,427 bp genome, comprising a circular chromosome and five circular plasmids (Table 1). Remaining 1bp hompolymer insertion errors were corrected using Illumina reads. The structural integrity of the chromosome assembly was confirmed correct by reference to a *NcoI* optical map (Fig. 1A). DNA base modification analysis indicated an absence of adenine methylation.

Antimicrobial resistance gene content. One of the major drivers behind the success of *E. faecium* in the clinical environment is their ability to acquire genes conferring antibiotic resistance (Handwerger et al. 1993; Iwen et al. 1997; Murray 2000). *In silico* antibiotic

resistance screening of the Ef_aus0233 genome confirmed the presence of seven loci, conferring resistance to many major classes of antibiotics including trimethoprim and vancomycin (Table 2). Vancomycin resistance in Ef_aus0233 is conferred by the Tn*1549* transposon, harbouring the *vanB* operon that has integrated into the chromosome. Here, Tn*1549* was the larger (57 kB) of the two reported versions and inserted into signal peptidase 1 gene, an insertion site that was previously reported in a comparative analysis of ST203 genomes (Howden et al. 2013; Lam et al. 2013a). Of particular interest, the Tn*1549* transposon in Ef_aus0233 and that of the fully assembled ST203 representative genome (Ef_aus0085) (Lam et al. 2013), shared 100% pairwise nucleotide identity across the full length of the element, implying a common Tn*1549* origin for these two clones. The majority of ST796 isolates exhibit vancomycin resistance. However, three ST796 vancomycin susceptible enterococci (VSE) have been isolated to date, one of which (Ef_aus1016) is included in our study and is discussed later.

Virulence gene content. In addition to antimicrobial resistance genes, virulence related genes are particularly enriched among hospital adapted *E. faecium* strains and are thought to enhance fitness in the hospital environment (Rice et al. 2003). *In silico* comparative analysis (see methods) of the Ef_aus0233 genome revealed the presence of several genes associated with virulence including collagen-binding adhesin (Rice et al. 2003) (chromosome coordinates 2,235,651-2,233,486), enterococcal surface protein (Shankar et al. 1999; Van Wamel et al. 2007) (chromosome coordinates 2,786,822-2,780,895), hemolysin (Cox et al. 2005) (chromosome coordinates 1,025,987-1,027,363), all of which are also present in the fully assembled genomes of ST17 and ST203 isolates (Lam et al. 2012; Lam et al. 2013). The Ef_aus0233 enterococcal surface protein shared identity with the ortholog in Ef_aus0004, a fully assembled ST17 genome (Lam et al. 2012), however it had a 358 bp insertion compared to Ef_aus0085, a fully assembled ST203 genome (Lam et al. 2013). In comparison, *esp* from Ef_aus0004=4938nt (6 Rib repeats), Ef_aus0085=5199nt (7 Rib repeats), Ef_aus0233=5928nt (9 rib repeats). The hemolysin of Ef_aus0233 shared complete identity with orthologs in Ef_aus0004 and Ef_aus0085, suggesting that this CDS may be under strong selection. Several other genes encoding factors that associate with *E. faecium* virulence were absent in

Ef_aus0233 and included gelatinase (Hancock & Perego 2004), aggregation substance (Olmsted et al. 1991), hemagglutinin (Elsner et al. 2000) and hyaluronidase (Rice et al. 2003).

Insertion sequence content. The Ef_aus0233 chromosome was found to contain 80 distinct elements (9 families) while Ef_aus0233_p1 had 41 (6 families), Ef_aus0233_p3 had 8 (5 families), Ef_aus0233_p4 had 8 (2 families) and no IS elements were detected on Ef_aus0233_p2 or Ef_aus0233_p5. Several of these IS families have been found not only in enterococci but additionally in species of other genera, including *Carnobacterium* and *Lysinibacillus*, reflecting the ease to which *E. faecium* can acquire exogenous DNA (Guzman Prieto et al. 2016).

CRISPR content. Akin to an adaptive immune system, the clustered regularly interspaced short palindromic repeats (CRISPR) systems of prokaryotes function as a sequence-specific security to defend genomes against viral predation and exposure to invading nucleic acid (Horvath & Barrangou 2010). Unlike members of the community-associated *E. faecium* lineage, genomes belonging to the CC17 *E. faecium* have been found to lack CRISPR systems (van Schaik et al. 2010). Given the advantages associated with the acquisition of extraneous DNA carrying antimicrobial resistance genes, CRISPRs are thought to be under negative selection among multi-drug resistant enterococci (Palmer & Gilmore 2010). Despite this, two distinct CRISPR loci were detected on Ef_aus0233_p1 (chromosome coordinates 168197- 168396) and Ef_aus0233_p2 (chromosome coordinates 2630- 2860), both containing three spacers and imperfect direct repeats. A single CRISPR associated gene (cas2) was detected. However, no cas1 ortholog was detected. Due to the practical necessity of cas1 for the operation of CRISPR systems (Yosef et al. 2012), it is unlikely that these detected CRISPR systems are functional.

Prophage content. The Ef_aus0233 genome was found to contain five putative prophages. Prophages Ef_aus0233_chr_phage-1 (chromosome coordinates 260,208 - 308,229: 52 CDS), Ef_aus0233_chr_phage-2 (chromosome coordinates 916,578 - 956,914: 61 CDS), Ef_aus0233_chr_phage-3 (chromosome coordinates 2,366,359 - 2,425,249: 78 CDS) and

Ef_aus0233_chr_phage-4 (chromosome coordinates 2,601,769 - 2,627,494: 19 CDS) were located on the chromosome (Fig. 1B) while Ef_aus0233_p1_phage-1 (plasmid-1 coordinates 62,059 - 94,736: 30 CDS) was identified on Ef_aus0233_p1. Alignment of these prophage elements signified that several common blocks of co-linearity existed (Fig. S1), however an overall lack of prophage genome conservation implies that these phage represent five distinct elements. Prophage gene content among a diverse collection of *E. faecium* genomes is discussed below.

Comparisons with other completed *E. faecium* genomes. In addition to diversity within the core and accessory genome, structural rearrangements represent an additional layer of genomic variation that may contribute to *E. faecium* phenotypic differences (Lam et al. 2012; Lam et al. 2013; Matthews & Maloy 2010). To assess how the genomic organization of the Ef_aus0233 chromosome compared to that of other *E. faecium* genomes, a whole chromosome alignment of Ef_aus0233, Ef_aus0004, Ef_aus0085 and DO was undertaken. The BLASTn based alignment revealed substantial conservation of genome content (Fig. 1B) and chromosomal architecture (Fig. 2A). Like Ef_aus0085 and DO, Ef_aus0233 does not exhibit the replichore inversion observed in Ef_aus0004 (Lam et al. 2012; Lam et al. 2013).

***E. faecium* population genomic comparisons.** In order to contextualize the 21 ST796 genomes within the global diversity of *E. faecium* as a species, we compared these genome data with a diverse collection of 89 published, fully assembled and draft *E. faecium* genomes (Table S1). To investigate the structure and evolutionary relatedness of the strains, we employed an unsupervised Bayesian clustering technique (BAPS) to distinguish distinct genomic populations and estimated a rooted phylogenomic tree using maximum likelihood.

Here, we found that our BAPS groups unambiguously classified genomes into the two previously reported A and B clades (Lebreton et al. 2013) (Fig. 3A). BAPS-1 corresponded with clade A and BAPS-2 corresponded with clade B, while BAPS-1, BAPS-1.1-4 and BAPS-1.5 overlapped with clades A1 and A2 (Fig. 3B). All ST796 isolates clustered within clade A1.

When inspecting the phylogeny of the ST796 genomes, it was noted that ST796 and ST555 share a most recent common ancestor (MRCA) (Fig. 3C). Almost without exception, ST555 and ST796 genomes formed distinct monophyletic clades, although the single vancomycin sensitive (VSE) ST796 included in this study clustered among the ST555 clade. Despite this phylogenomic incongruence, the BAPS groupings of ST555 and ST796 genomes (BAPS 1.3.3 and 1.3.4, respectively) were in harmony with the MLST designations. The ST555 clone is another recently-emerged hospital adapted ST, however its discovery in the hospital environment predates that of ST796 (Coombs et al. 2014). Another major difference between these two STs is that ST796 appears to have been localized to south east Australia and New Zealand (Carter et al. 2016; Coombs et al. 2014), while ST555 has been reported nationally in the Northern Territory, South Australia and Western Australia (Coombs et al. 2014), in China (Liu et al. 2011) and among wild birds in the United States (Oravcova et al. 2014). The phylogenetic position of the ST555 clade being basal to the VRE ST796 clade, suggests that VRE ST796 emerged from an ST555-like ancestral progenitor. The national and international pervasiveness of ST555 and relatively limited geographical dispersal of ST796 in southeast Australia is consistent with a scenario in which the evolutionary emergence of ST555 predates that of ST796.

Recombination analyses indicated that both the ST555-796 and ST796 MRCAs have evolved in part by recombination. Inspection of the inferred recombining segments for these two ancestors revealed a single hotspot of 170kb that contained two overlapping clusters of increased SNP density (ST555-796_MRCA: chromosome coordinates 1,783,249-1,953,029, ST796_MRCA: chromosome coordinates 1,857,926-1,937,284) (Fig. 2B). The spatial clustering of these inferred ancestral recombination events suggests that this region may contain non-clonally derived alleles (particularly those in the ST796 MRCA) that may have been under positive selection and perhaps contributed to the emergence of ST796. Inspection of this region revealed a preponderance of cell-wall associated transport CDS, including CDS encoding putative copper and cadmium-translocating P-type ATPases, amino acid permeases, OxaA-like

membrane protein, as well as housekeeping CDS such as Glycyl-tRNA synthetase subunits, RecO DNA repair proteins and a GTP-binding protein.

In order to assess the evolutionary divergence between the ST555 and ST796 clades, a ST555-796 specific core genome was established and pairwise SNP distances were calculated. As we were primarily interested in SNPs derived through clonal evolution, we removed SNPs within the above mentioned 170kb region of ancestral recombination. Inspection of SNP distribution in VSE Ef_aus1016 revealed several dense clusters, indicating substantial recombination, the majority of which was not detected by ClonalFrameML. As the recombination was so extensive and limited to Ef_aus1016, this genome was removed from the inter-clade comparisons. Consistent with two distinct groups, within clade comparisons revealed smaller mean SNP differences (within-ST555: 11 SNPs, within-ST796: 49 SNPs) than that between clades (ST555-796: 151 SNPs) (Fig. 3D).

Ancestral single nucleotide polymorphisms. Forty-one core-genome SNPs differentiated the ST555-ST796 MRCA from its predecessors, while only two core-genome SNPs were predicted in the ST796 MRCA compared to ST555 (Table S2). Analysis of these SNPs showed a range of CDS impacted. Among the 41 SNPs, 22 were predicted to change amino-acid sequence and potentially alter protein function, including non-synonymous mutations in four CDS encoding putative regulatory proteins. While the function of these regulatory proteins and the consequence of the predicted mutations is unknown, such changes can have profound impacts on phenotype (Howden et al. 2011; Howden et al. 2008).

Accessory gene content comparisons. The clustering of all predicted CDSs into orthologous groups allowed for inter-ST comparisons at the gene-content level. In total there were 10,740 orthologous clusters among the 110 genomes, of which 1,437 were core and 9,303 were variably present (accessory) - representing the *E. faecium* pan genome (Table S3). Using this approach, orthologous clusters that were diagnostic of the ST555-ST796 and ST796 populations were identified.

Lineage specific genomic islands. The acquisition of genomic islands has been reported in previous studies that compared the genomes of hospitalized and non-hospital derived isolates, suggesting that such novel elements may offer possessing strains a competitive advantage (Heikens et al. 2008). In this study, subsets of the accessory genome that were associated with ST555-ST796 and ST796 genomes were found to cluster on the Ef_aus0233 chromosome (Fig. 2C). The contiguous location of these CDS and conserved inheritance patterns, indicated that these elements collectively formed larger genomic islands and were likely to have been acquired through horizontal gene transfer events. Given the ancestral relationships among the genomes and the conservation of ortholog presence amongst these lineages, it is reasonable to infer that these events occurred at various stages along the evolutionary paths of the ST555-ST796 and ST796 ancestries.

Two genomic islands were conserved among ST555-ST796 genomes while being almost entirely absent from genomes of other STs (ST555-ST796_GI-1, chromosome coordinates 39,093-53,122: 13 CDS, and ST555-ST796_GI-2 chromosome coordinates 2,316,643-2,373,309: 63 CDS), suggesting that these elements are likely to have been acquired by the ST555-ST796 MRCA. Assessment of the CDS annotation for ST555-ST796_GI-1 (56kb) suggests it is a mosaic integrative element. Two 3kb regions spanning a site-specific tyrosine recombinase and excisonase at the 5' end of this element and replication proteins at the 3' extreme were identical to a previously described *Enterococcus faecalis* pathogenicity island (Shankar et al. 2002). A 13kb region harbouring a putative beta-galactosidase and other sugar modifying CDS was identical to a region in ST203 Ef_aus0085 (Lam et al. 2013). The function of CDS in the remaining 33kb was more difficult to infer with few database matches to indicate function. However, a role for this region in cell wall modification (potentially DNA transfer) was suggested by the presence of CDS encoding cell-wall binding proteins, peptidases, ATP/GTP-binding proteins and peptidoglycan-binding proteins. ST555-ST796_GI-2 (14kb) is another integrative element with CDS encoding a site-specific tyrosine recombinase, replication proteins, and putative sugar kinases, hydrolases and permeases. This potential carbohydrate

utilization/transport locus shared complete nucleotide sequence identity with a region of the *Enterococcus gallinarum* genome (strain ID: FDA ARGOS_163, NCBI BioProject: PRJNA231221). The phage CDS content and their predicted products is provided (Table S4)

Two other elements were found to be exclusively present among ST796 genomes in this comparison. One of these was the Tn*916*-like transposon (chromosome coordinates 635,179 - 658,601: 19 CDS), carrying tetracycline resistance (Franke & Clewell 1981) and the second was the Tn*1549* transposon (chromosome coordinates 803,567- 861,054: 61 CDS), carrying vancomycin resistance (Garnier et al. 2000). Given the phylogeny, it appears likely that the ST796 MRCA acquired these two transposons and then spread (Fig. 2C). The exception to this pattern was the single VSE ST796 genome (Ef_aus1016) that lacked these elements that were universally conserved among VRE ST796 genomes. Ef_aus1016 exhibits the same genomic island presence/absence profile that is observed among the ST555 genomes. It appears that Ef_aus1016 may actually be a ST555 genome that has recombined with ST796. This explains why BAPS (when run on a non-recombination filtered alignment) clusters Ef_aus1016 as an ST796 and why the phylogeny (which was built upon a recombination filtered alignment) clusters this genome into the ST555 clade (Fig. 3C). This hypothesis also provides an explanation as to why Ef_aus1016 exhibits an ST555-like genomic island presence and absence profile (Fig. 2C). An alternative explanation would be that Ef_aus1016 might be an extant descendent of a ST796 evolutionary intermediate that had not yet acquired the ST796 specific GIs, such as Tn*1549*. The horizontal acquisition of Tn*1549* has been demonstrated (Launay et al. 2006) and evidence for a VSE version of an emergent *E. faecium* clone to precede the VRE version has been previously documented with the emergence of ST17, ST203 and ST252 (Johnson et al. 2010).

Prophage gene content comparisons. Alignment of the orthologs found within the prophages that were identified in the Ef_aus0233 genome revealed the extent to which these elements are conserved among the greater *E. faecium* population (Fig. 4). The vast majority of orthologs within these prophages were found to exist in non-ST796 genomes, suggesting that at the gene-

content level, such prophages are not unique to ST796. Prophages Ef_aus0233_chr_phage-2, Ef_aus0233_chr_phage-4 and Ef_aus0233_p1_phage-1 showed the greatest degree of ortholog conservation with non-ST796 genomes. Prophages Ef_aus0233_chr_phage-1 and Ef_aus0233_chr_phage-3 did contain orthologs present in non-ST796 genomes, however the presence of these orthologs outside of ST796 genomes was limited. Overall the prophages in Ef_aus0233 form a substantial contribution to the accessory genome but do not contain CDS that are unique to ST796 (Table S4).

Plasmid gene content comparisons. Plasmids form an important component of the *E. faecium* accessory genome that can spread horizontally through a population and carry genetic elements that may confer enhanced fitness (Fiedler et al. 2016). Approximately 12% of the Ef_aus0233 genome (384,340 bp) is comprised of plasmid DNA. In order to assess the conservation of plasmid gene content among ST796 genomes and across the greater *E. faecium* population, the presence and absence of plasmid genes within the ortholog clusters were inspected (Fig. 5). Patterns of individual ortholog presence and absence demonstrated that all plasmid orthologs were found in non-ST796 genomes, however in varying degrees. Plasmids Ef_aus0233_p1, Ef_aus0233_p4 and Ef_aus0233_p5 contain orthologs found outside ST796, however they are rarely seen elsewhere in their entirety. A list of all plasmid CDS content and their predicted products is provided (Table S5)

Given these gene content patterns and the aforementioned phylogenomic relationships between ST555 and ST796 genomes, it appears likely that the ST555-ST796 MRCA acquired these plasmids, as they are not observed in their entirety in surrounding clades. Interestingly, Ef_aus0233_p2 was not only scarce among non-ST796 genomes but lacked conservation among the ST796 genomes. Overall, no single plasmid ortholog was specific to ST796, however these plasmids in their entirety are diagnostic of the ST555-ST796 lineage. Furthermore, the intra-ST796 differences in plasmid gene content, particularly in Ef_aus0233_p2, indicate there are appreciable amounts of diversity within the ST796 accessory genome, variation that might be useful during outbreak investigations involving this clone (Lister et al. 2015).

406

407

408 CONCLUSION

409 The hospital environment presents a challenging ecological niche for the adaptation of bacterial
410 pathogens. Historically, *E. faecalis* was the leading causative agent of enterococcal nosocomial
411 infections, however *E. faecium* infections have escalated in the last decade (Galloway-Peña et
412 al. 2009; Guzman Prieto et al. 2016; Leavis et al. 2006; Willems et al. 2011; Willems & Van
413 Schaik 2009). Following this apparent interspecies replacement, population-based studies have
414 observed substantial intraspecies dynamics with clonal replacement of *E. faecium* STs in
415 hospitals (Bender et al. 2016; Johnson et al. 2010). Here we have described the genomic basis
416 for the emergence of a new highly hospital adapted *E. faecium* ST early in its evolutionary
417 history. The preparation of a fully assembled ST796 genome facilitated a comprehensive
418 genomic analysis of this lineage and enabled detailed comparisons among other clinically
419 relevant draft and fully assembled *E. faecium* genomes.

420

421 We demonstrate that the emergence of ST796 was preceded by several genomic events
422 including the acquisition of two genomic islands, plasmid and phage activity, modest SNV
423 accumulation and recombination. These analyses highlight genetic elements within the *E.*
424 *faecium* core and accessory genome that may have been important drivers for the evolution of
425 the ST555-ST796 and ST796 lineages. Given the likely significance of genomic island acquisition
426 for the emergence of CC17 (Heikens et al. 2008), the GIs presented in this study presumably
427 reflect adaptive responses to the clinical environment, either through acquired antibiotic
428 resistance or perhaps enhanced capacity to utilize carbohydrates and thus augment
429 gastrointestinal colonization. Our finding that ST796 evolved from an ST555-like ancestral
430 progenitor is another example of newly emergent VRE arising from a VSE MRCA, although
431 ST796 itself is almost exclusively VRE when identified in human BSIs unlike ST555 which causes
432 both VSE and VRE BSI in Australian hospitals in about equal proportions (Coombs et al. 2014).

433

This analysis focused upon providing an overview of the first fully assembled ST796 genome and genomic differences that were assessed at the inter-ST population level. In order to explore specific diversity within the ST796 lineage, an intra-ST population study focusing upon diversity among a large collection ST796 genomes is currently underway. Our observation of substantial variation within the ST796 accessory genome, in particular plasmid presence and absence, suggest a means for effective intra-ST796 genotyping that could potentially be more useful than core genome analysis in the tracking of outbreaks.

In this study we have described the genomic events that have shaped the evolution of *E. faecium* ST796. While the extent to which each genomic event has contributed to the ST796 emergence is not yet understood, our findings lay a foundation for testing specific hypotheses that have arisen from this work.

REFERENCES

Alikhan N-F, Petty NK, Zakour NLB, and Beatson SA. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:1.

Arthur M, Reynolds P, and Courvalin P. 1996. Glycopeptide resistance in enterococci. *Trends Microbiol* 4:401-407. 10.1016/0966-842X(96)10063-9

Bender JK, Kalmbach A, Fleige C, Klare I, Fuchs S, and Werner G. 2016. Population structure and acquisition of the vanB resistance determinant in German clinical isolates of *Enterococcus faecium* ST192. *Sci Rep* 6:21847. 10.1038/srep21847

Burdett V, Inamine J, and Rajagopalan S. 1982. Heterogeneity of tetracycline resistance determinants in Streptococcus. *Journal of bacteriology* 149:995-1004.

Carter GP, Buultjens AH, Ballard SA, Baines S, Takehiro T, Strachan J, Johnson PDR, Ferguson JK, Seemann T, Stinear TP, and Howden BP. 2016. Emergence of endemic MLST non-typeable vancomycin-resistant *Enterococcus faecium*. *The Journal of antimicrobial chemotherapy* In press.

Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, and Parkhill J. 2005. ACT: the Artemis Comparison Tool. *Bioinformatics* 21:3422-3423. bti553 [pii] 10.1093/bioinformatics/bti553

Cheng L, Connor TR, Sirén J, Aanensen DM, and Corander J. 2013. Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Mol Biol Evol* 30:1224-1228.

Coombs GW, Pearson JC, Daley DA, Le T, Robinson OJ, Gottlieb T, Howden BP, Johnson PD, Bennett CM, Stinear TP, Turnidge JD, and Australian Group on Antimicrobial R. 2014.

Molecular epidemiology of enterococcal bacteremia in Australia. *J Clin Microbiol* 52:897-905. 10.1128/JCM.03286-13

Cox CR, Coburn PS, and 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.

Darling AC, Mau B, Blattner FR, and Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* 14:1394-1403. 10.1101/gr.2289704

14/7/1394 [pii]

Didelot X, and Wilson DJ. 2015. ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. *PLoS Comput Biol* 11:e1004041-e1004041.

Elsner HA, Sobottka I, Mack D, Claussen M, Laufs R, and Wirth R. 2000. Virulence factors of *Enterococcus faecalis* and *Enterococcus faecium* blood culture isolates. *Eur J Clin Microbiol Infect Dis* 19:39-42.

Fiedler S, Bender JK, Klare I, Halbedel S, Grohmann E, Szewzyk U, and Werner G. 2016. Tigecycline resistance in clinical isolates of Enterococcus faecium is mediated by an upregulation of plasmid-encoded tetracycline determinants tet(L) and tet(M). *The Journal of antimicrobial chemotherapy* 71:871-881. 10.1093/jac/dkv420

Franke AE, and Clewell DB. 1981. Evidence for a chromosome-borne resistance transposon (Tn916) in *Streptococcus faecalis* that is capable of "conjugal" transfer in the absence of a conjugative plasmid. *J Bacteriol* 145:494-502.

Galloway-Peña J, Roh JH, Latorre M, Qin X, and 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.

Galloway-Peña JR, Nallapareddy SR, Arias CA, Eliopoulos GM, and Murray BE. 2009. Analysis of clonality and antibiotic resistance among early clinical isolates of *Enterococcus faecium* in the United States. *Journal of Infectious Diseases* 200:1566-1573.

Garnier F, Taourit S, Glaser P, Courvalin P, and Galimand M. 2000. Characterization of transposon Tn1549, conferring VanB-type resistance in Enterococcus spp. *Microbiology* 146:1481-1489.

Grabsch EA, Mahony AA, Cameron DR, Martin RD, Heland M, Davey P, Petty M, Xie S, and Grayson ML. 2012. Significant reduction in vancomycin-resistant enterococcus colonization and bacteraemia after introduction of a bleach-based cleaning-disinfection programme. *J Hosp Infect* 82:234-242. 10.1016/j.jhin.2012.08.010

Guzman Prieto AM, van Schaik W, Rogers MR, Coque TM, Baquero F, Corander J, and Willems RJ. 2016. Global Emergence and Dissemination of Enterococci as Nosocomial Pathogens: Attack of the Clones? *Front Microbiol* 7:788. 10.3389/fmicb.2016.00788

Hancock LE, and Perego M. 2004. The *Enterococcus faecalis fsr* two-component system controls biofilm development through production of gelatinase. *J Bacteriol* 186:5629-5639. 10.1128/JB.186.17.5629-5639.2004

Handwerger S, Raucher B, Altarac D, Monka J, Marchione S, Singh KV, Murray BE, Wolff J, and Walters B. 1993. Nosocomial outbreak due to *Enterococcus faecium* highly resistant to vancomycin, penicillin, and gentamicin. *Clinical Infectious Diseases* 16:750-755.

513 Heikens E, van Schaik W, Leavis HL, Bonten MJ, and Willems RJ. 2008. Identification of a novel
514 genomic island specific to hospital-acquired clonal complex 17 *Enterococcus faecium*
515 isolates. *Applied and environmental microbiology* 74:7094-7097.
516 Horvath P, and Barrangou R. 2010. CRISPR/Cas, the immune system of bacteria and archaea.
517 *Science* 327:167-170.
518 Howden BP, McEvoy CR, Allen DL, Chua K, Gao W, Harrison PF, Bell J, Coombs G, Bennett-Wood
519 V, Porter JL, Robins-Browne R, Davies JK, Seemann T, and Stinear TP. 2011. Evolution of
520 multidrug resistance during *Staphylococcus aureus* infection involves mutation of the
521 essential two component regulator WalKR. *PLoS Pathog* 7:e1002359.
522 10.1371/journal.ppat.1002359
523 PPATHOGENS-D-11-01054 [pii]
524 Howden BP, Stinear TP, Allen DL, Johnson PD, Ward PB, and Davies JK. 2008. Genomic analysis
525 reveals a point mutation in the two-component sensor gene *graS* that leads to
526 intermediate vancomycin resistance in clinical *Staphylococcus aureus*. *Antimicrob*
527 *Agents Chemother* 52:3755-3762. AAC.01613-07 [pii]
528 10.1128/AAC.01613-07
529 Iwen PC, Kelly DM, Linder J, Hinrichs SH, Dominguez EA, Rupp ME, and Patil KD. 1997. Change in
530 prevalence and antibiotic resistance of Enterococcus species isolated from blood
531 cultures over an 8-year period. *Antimicrobial agents and chemotherapy* 41:494.
532 Johnson PD, Ballard SA, Grabsch EA, Stinear TP, Seemann T, Young HL, Grayson ML, and
533 Howden BP. 2010. A sustained hospital outbreak of vancomycin-resistant *Enterococcus*
534 *faecium* bacteremia due to emergence of *vanB E. faecium* sequence type 203. *The*
535 *Journal of infectious diseases* 202:1278-1286. 10.1086/656319
536 Lam MM, Seemann T, Bulach DM, Gladman SL, Chen H, Haring V, Moore RJ, Ballard S, Grayson
537 ML, Johnson PD, Howden BP, and Stinear TP. 2012. Comparative analysis of the first
538 complete *Enterococcus faecium* genome. *J Bacteriol* 194:2334-2341. 10.1128/JB.00259-
539 12
540 Lam MM, Seemann T, Tobias NJ, Chen H, Haring V, Moore RJ, Ballard S, Grayson LM, Johnson
541 PD, Howden BP, and Stinear TP. 2013. Comparative analysis of the complete genome of
542 an epidemic hospital sequence type 203 clone of vancomycin-resistant *Enterococcus*
543 *faecium*. *BMC Genomics* 14:595. 10.1186/1471-2164-14-595
544 Launay A, Ballard SA, Johnson PD, Grayson ML, and Lambert T. 2006. Transfer of vancomycin
545 resistance transposon Tn1549 from *Clostridium symbiosum* to *Enterococcus* spp. in the
546 gut of gnotobiotic mice. *Antimicrobial agents and chemotherapy* 50:1054-1062.
547 Leavis HL, Bonten MJM, and Willems RJL. 2006. Identification of high-risk enterococcal clonal
548 complexes: global dispersion and antibiotic resistance. *Cur Opin Microbiol* 9:454-460.
549 Leavis HL, Willems RJ, Van Wamel WJ, Schuren FH, Caspers MP, and Bonten MJ. 2007. Insertion
550 sequence–driven diversification creates a globally dispersed emerging multiresistant
551 subspecies of *E. faecium*. *PLoS Pathog* 3:e7.
552 Lebreton F, van Schaik W, McGuire AM, Godfrey P, Griggs A, Mazumdar V, Corander J, Cheng L,
553 Saif S, Young S, Zeng Q, Wortman J, Birren B, Willems RJ, Earl AM, and Gilmore MS.
554 2013. Emergence of epidemic multidrug-resistant *Enterococcus faecium* from animal
555 and commensal strains. *MBio* 4. 10.1128/mBio.00534-13

556 Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, and Prohaska SJ. 2011. Proteinortho:
557 detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics* 12:124.
558 10.1186/1471-2105-12-124
559 Lister DM, Kotsanas D, Ballard SA, Howden BP, Carse E, Tan K, Scott C, Gillespie EE, Mahony AA,
560 Doherty R, Korman TM, Johnson PD, and Stuart RL. 2015. Outbreak of vanB vancomycin-
561 resistant *Enterococcus faecium* colonization in a neonatal service. *Am J Infect Control*
562 43:1061-1065. 10.1016/j.ajic.2015.05.047
563 Liu Y, Cao B, Gu L, and Wang H. 2011. Molecular characterization of vancomycin-resistant
564 enterococci in a Chinese hospital between 2003 and 2009. *Microbial Drug Resistance*
565 17:449-455.
566 Mahony AA, Grabsch EA, Ballard SA, Wang J, Xie S, Roberts SA, Stuart RL, Bak N, Stinear TP, and
567 Johnson PDR. 2014. Vancomycin-resistant *Enterococcus faecium* sequence type 796, the
568 new trans-Tasman epidemic clone. Australian Society for Antimicrobials - Antimicrobials,
569 Melbourne; 2014. Melbourne: Australian Society for Antimocrobials.
570 Matthews TD, and Maloy S. 2010. Fitness effects of replichore imbalance in *Salmonella*
571 *enterica*. *J Bacteriol* 192:6086-6088. 10.1128/JB.00649-10
572 Murray BE. 2000. Vancomycin-resistant enterococcal infections. *New England Journal of*
573 *Medicine* 342:710-721.
574 Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Prjibelski AD, Pyshkin
575 A, Sirotkin A, Sirotkin Y, Stepanauskas R, Clingenpeel SR, Woyke T, McLean JS, Lasken R,
576 Tesler G, Alekseyev MA, and Pevzner PA. 2013. Assembling single-cell genomes and
577 mini-metagenomes from chimeric MDA products. *J Comput Biol* 20:714-737.
578 10.1089/cmb.2013.0084
579 Olmsted S, Kao S, Van Putte L, Gallo J, and Dunny G. 1991. Role of the pheromone-inducible
580 surface protein Asc10 in mating aggregate formation and conjugal transfer of the
581 *Enterococcus faecalis* plasmid pCF10. *J Bacteriol* 173:7665-7672.
582 Oravcova V, Zurek L, Townsend A, Clark AB, Ellis JC, Cizek A, and Literak I. 2014. American crows
583 as carriers of vancomycin-resistant enterococci with *vanA* gene. *Environ Microbiol*
584 16:939-949.
585 Palmer KL, and Gilmore MS. 2010. Multidrug-resistant enterococci lack CRISPR-cas. *MBio*
586 1:e00227-00210.
587 Palmer KL, Godfrey P, Griggs A, Kos VN, Zucker J, Desjardins C, Cerqueira G, Gevers D, Walker S,
588 and Wortman J. 2012. Comparative genomics of enterococci: variation in *Enterococcus*
589 *faecalis*, clade structure in *E. faecium*, and defining characteristics of *E. gallinarum* and
590 *E. casseliflavus*. *MBio* 3:e00318-00311.
591 Patterson JE, and Zervos MJ. 1990. High-level gentamicin resistance in Enterococcus:
592 microbiology, genetic basis, and epidemiology. *Review of Infectious Diseases* 12:644-
593 652.
594 Portillo A, Ruiz-Larrea F, Zarazaga M, Alonso A, Martinez JL, and Torres C. 2000. Macrolide
595 Resistance Genes in Enterococcus spp. *Antimicrobial agents and chemotherapy* 44:967-
596 971.
597 Price MN, Dehal PS, and Arkin AP. 2010. FastTree 2–approximately maximum-likelihood trees
598 for large alignments. *PLoS One* 5:e9490.

599 Quednau M, Ahrne S, and Molin G. 1999. Genomic Relationships between *Enterococcus*
600 *faecium* Strains from Different Sources and with Different Antibiotic Resistance Profiles
601 Evaluated by Restriction Endonuclease Analysis of Total Chromosomal DNA Using EcoRI
602 andPvuII. *Appl Environ Microbiol* 65:1777-1780.
603 Rice LB, Carias L, Rudin S, Vael C, Goossens H, Konstabel C, Klare I, Nallapareddy SR, Huang W,
604 and Murray BE. 2003. A potential virulence gene, hylEfm, predominates in *Enterococcus*
605 *faecium* of clinical origin. *The Journal of infectious diseases* 187:508-512.
606 10.1086/367711
607 Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*:btu153.
608 Sekiguchi J-i, Tharavichitkul P, Miyoshi-Akiyama T, Chupia V, Fujino T, Araake M, Irie A, Morita
609 K, Kuratsuji T, and Kirikae T. 2005. Cloning and characterization of a novel trimethoprim-
610 resistant dihydrofolate reductase from a nosocomial isolate of *Staphylococcus aureus*
611 CM. S2 (IMCJ1454). *Antimicrobial agents and chemotherapy* 49:3948-3951.
612 Shankar N, Baghdayan AS, and Gilmore MS. 2002. Modulation of virulence within a
613 pathogenicity island in vancomycin-resistant *Enterococcus faecalis*. *Nature* 417:746-750.
614 10.1038/nature00802
615 Shankar V, Baghdayan AS, Huycke MM, Lindahl G, and Gilmore MS. 1999. Infection-derived
616 *Enterococcus faecalis* strains are enriched in esp, a gene encoding a novel surface
617 protein. *Infect Immun* 67:193-200.
618 Tong SY, Xie S, Richardson LJ, Ballard SA, Dakh F, Grabsch EA, Grayson ML, Howden BP, Johnson
619 PD, and Giffard PM. 2011. High-resolution melting genotyping of Enterococcus faecium
620 based on multilocus sequence typing derived single nucleotide polymorphisms. *PLoS*
621 *One* 6:e29189. 10.1371/journal.pone.0029189
622 Top J, Willems R, and Bonten M. 2008. Emergence of CC17 *Enterococcus faecium*: from
623 commensal to hospital-adapted pathogen. *FEMS Immunol Med Microbiol* 52:297-308.
624 FIM383 [pii]
625 10.1111/j.1574-695X.2008.00383.x
626 Trieu-Cuot P, Poyart-Salmeron C, Carlier C, and Courvalin P. 1990. Nucleotide sequence of the
627 erythromycin resistance gene of the conjugative transposon Tn1545. *Nucleic Acids*
628 *Research* 18:3660-3660.
629 van Schaik W, Top J, Riley DR, Boekhorst J, Vrijenhoek JE, Schapendonk CM, Hendrickx AP,
630 Nijman IJ, Bonten MJ, and Tettelin H. 2010. Pyrosequencing-based comparative genome
631 analysis of the nosocomial pathogen *Enterococcus faecium* and identification of a large
632 transferable pathogenicity island. *BMC Genomics* 11:239.
633 Van Wamel WJ, Hendrickx AP, Bonten MJ, Top J, Posthuma G, and Willems RJ. 2007. Growth
634 condition-dependent Esp expression by *Enterococcus faecium* affects initial adherence
635 and biofilm formation. *Infect Immun* 75:924-931. 10.1128/IAI.00941-06
636 Varani AM, Siguier P, Gourbeyre E, Charneau V, and Chandler M. 2011. ISsaga is an ensemble of
637 web-based methods for high throughput identification and semi-automatic annotation
638 of insertion sequences in prokaryotic genomes. *Genome Biol* 12:R30. 10.1186/gb-2011-
639 12-3-r30
640 Willems RJ, Hanage WP, Bessen DE, and Feil EJ. 2011. Population biology of Gram-positive
641 pathogens: high-risk clones for dissemination of antibiotic resistance. *FEMS Microbiol*
642 *Rev* 35:872-900.

643 Willems RJ, Top J, van Belkum A, Endtz H, Mevius D, Stobberingh E, van den Bogaard A, and van
644 Embden JD. 2000. Host specificity of vancomycin-resistant *Enterococcus faecium*.
645 *Journal of Infectious Diseases* 182:816-823.
646 Willems RJ, Top J, van Santen M, Robinson DA, Coque TM, Baquero F, Grundmann H, and
647 Bonten MJ. 2005. Global spread of vancomycin-resistant *Enterococcus faecium* from
648 distinct nosocomial genetic complex. *Emerg Infect Dis* 11:821-828.
649 10.3201/eid1106.041204
650 Willems RJ, and Van Schaik W. 2009. Transition of *Enterococcus faecium* from commensal
651 organism to nosocomial pathogen. *Future Microbiol* 4:1125-1135.
652 Yosef I, Goren MG, and Qimron U. 2012. Proteins and DNA elements essential for the CRISPR
653 adaptation process in *Escherichia coli*. *Nucleic Acids Res* 40:5569-5576.
654 10.1093/nar/gks216
655 Zhou Y, Liang Y, Lynch KH, Dennis JJ, and Wishart DS. 2011. PHAST: a fast phage search tool.
656 *Nucleic Acids Res* 39:W347-352. 10.1093/nar/gkr485
657

TABLES

Table 1: Characteristics of *E. faecium* ST796 Ef_aus0233 complete genome

	Length (bp)	% G+C	Copy number	No of CDS	No of non-parologous CDS	No of tRNA	No of rRNA
Chromosome	2,888,087	38.2	1	2,726	2,644	70	6
Ef_aus0233_p1	197,153	35.4	1	210	197	-	-
Ef_aus0233_p2	79,293	33.8	1	96	96	-	-
Ef_aus0233_p3	77,977	35.2	1	84	84	-	-
Ef_aus0233_p4	22,080	35.6	2	28	23	-	-
Ef_aus0233_p5	7,837	33.5	8	8	8	-	-

666 **Table 2:** Antibiotic resistance genes and mutations present in Ef_aus0233 and other ST796

Resistance	Product	Gene	Location (nucleotide positions)	Reference
Trimethoprim	Dihydrofolate reductase	*dfrG*	331,475 - 331,972 (chromosome)	(Sekiguchi et al. 2005)
Tetracycline	Tetracycline resistance protein	*tetM* (Tn*916*)	652,734 - 654,653 (chromosome)	(Burdett et al. 1982)
Macrolides	ABC transporter protein	msrC	2,711,468 - 2,712,946 (chromosome)	(Portillo et al. 2000)
	rRNA adenine N-6-methyltransferase	*ermB*	13,080 - 13,842 (plasmid 4)	(Trieu-Cuot et al. 1990)
Aminoglycosides	Bifunctional aminoglycoside modifying enzyme	aac(6')-aph2''	60,698 - 62,008 (plasmid 1) 60,366 - 61,805 (plasmid 3)	(Patterson & Zervos 1990)
Vancomycin	VanB ligase	*vanB* (Tn*1549*)	803,567- 861,054 (chromosome)	(Arthur et al. 1996)

667

668

669

670

FIGURES

Fig. 1

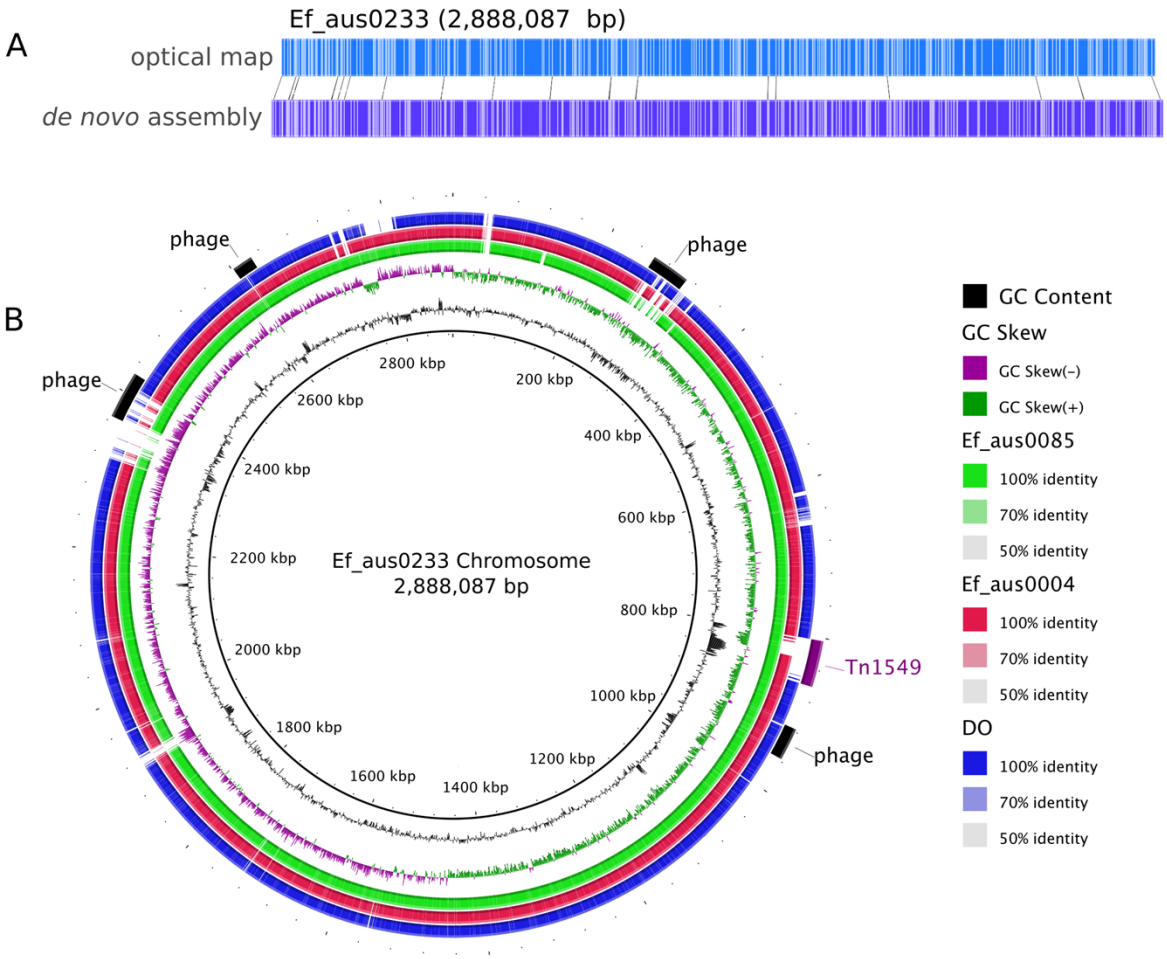
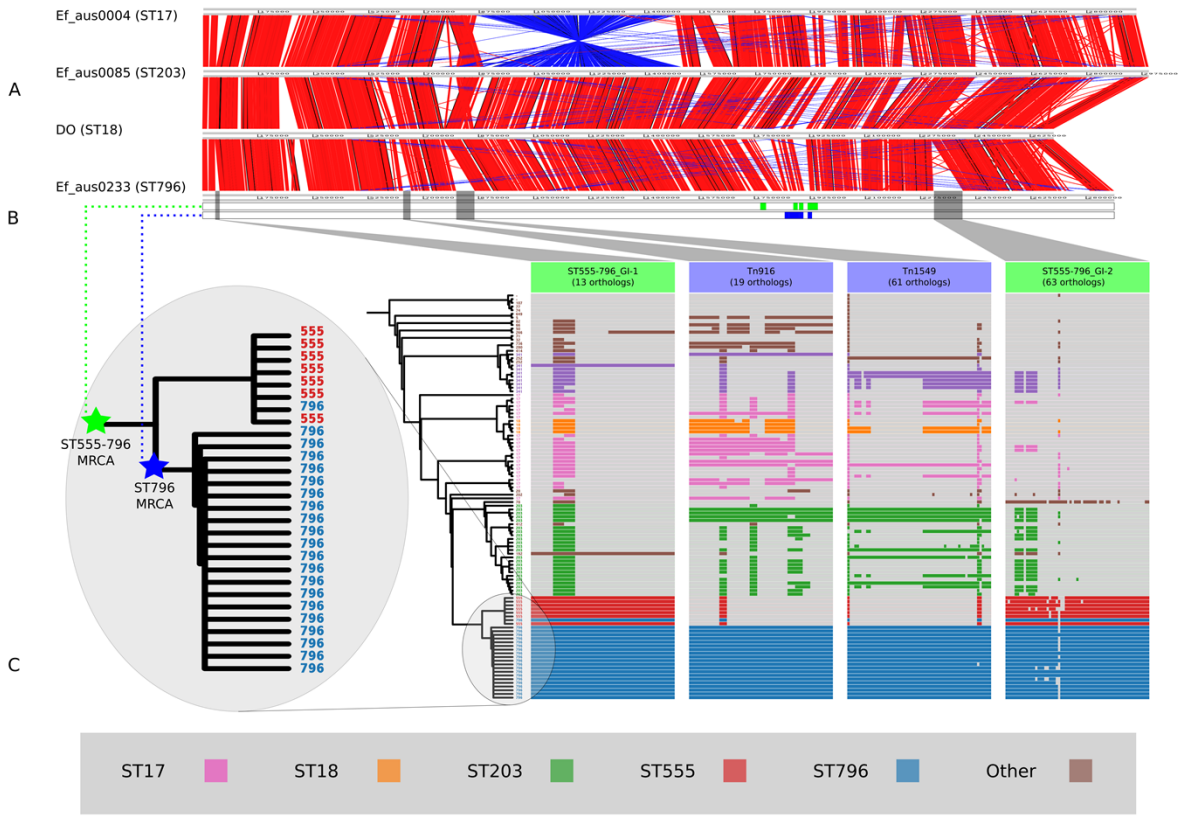


Fig. 1: Ef_aus0233 chromosomal optical map and BRIG plot. (A) Optical map of the Ef_aus0233 chromosome. (B) Referenced based alignment of blast hits of Ef_aus0085, Ef_aus0004 and DO genomes against the aus0233 chromosome. Prophage elements and the Tn*1549* VanB containing transposon are annotated as arcs in the outermost ring.

681 **Fig. 2**



682

683

684 **Fig 2:** Comparisons of chromosomal architecture, genomic islands and recombining segments

685 associated with the ST555-796 and ST796 clades. (A) Alignment of fully assembled

686 chromosomes of Ef_aus0233, Ef_aus0085, Ef_aus0004 and DO. (B) Recombining segments in

687 the ST555-796 and ST796 MRCAs. (C) Core genome phylogeny aligned with gene content blocks

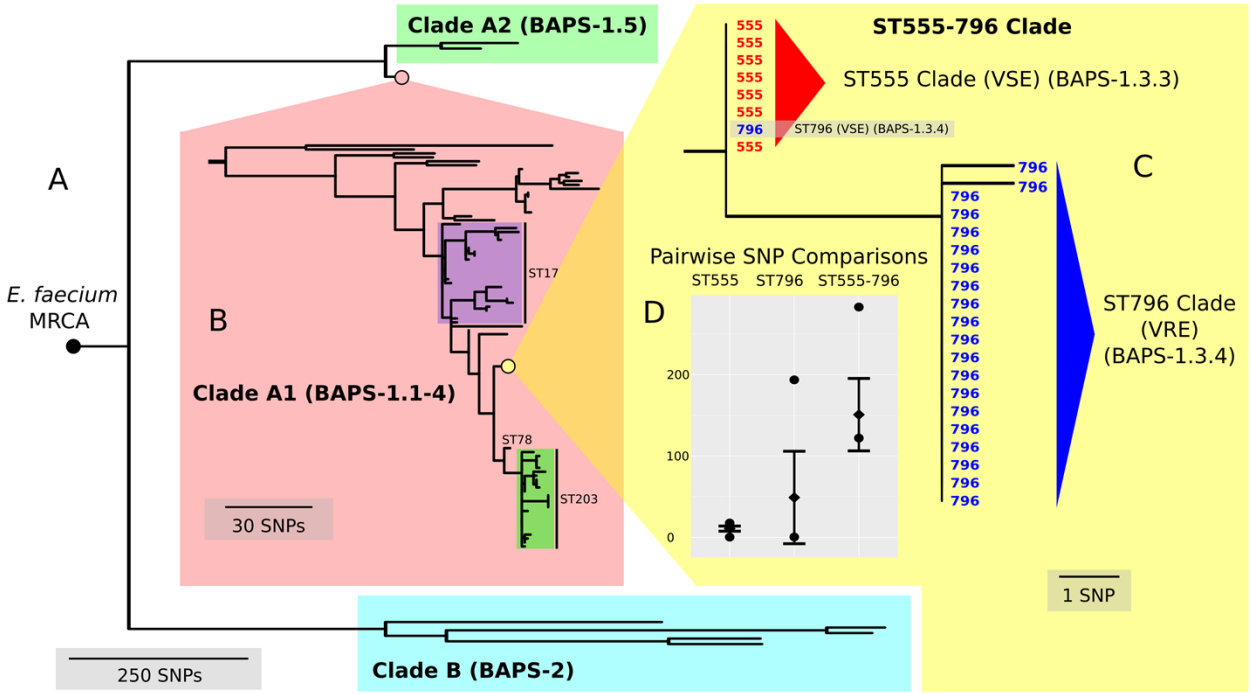
688 for identified genomic islands. Colours indicate the MLST designations.

689

690

691

692 **Fig. 3**



693

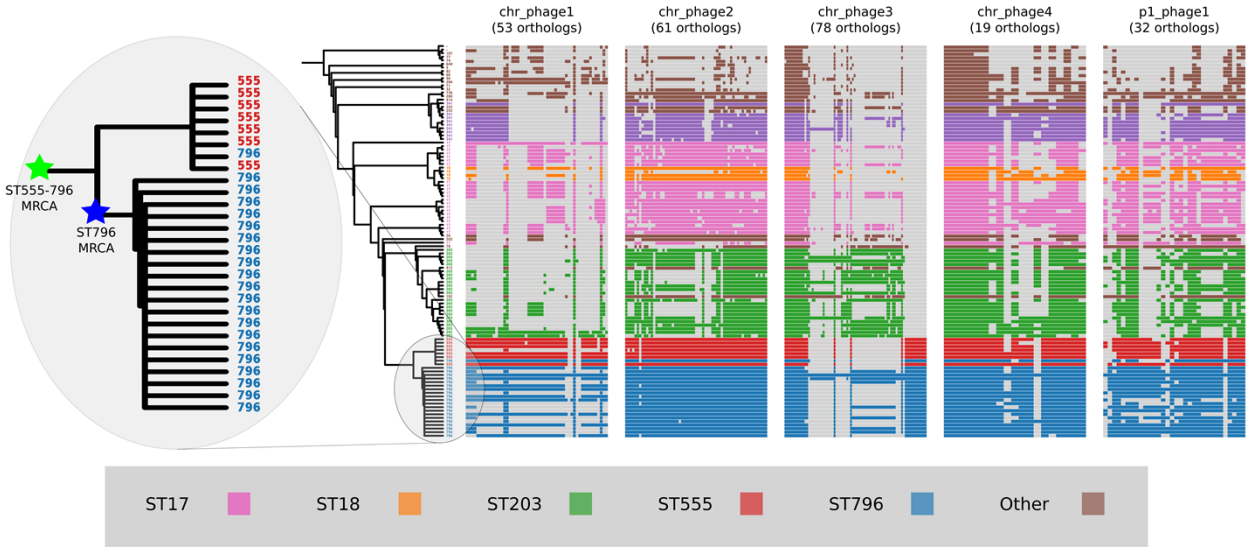
694

695 **Fig. 3:** Nested core genome SNP phylogeny of the greater E. faecium population. (A) *E. faecium*
696 population tree containing the major division between the community and hospital associated
697 clades (B and A). (B) Nested tree focusing on the sub-clade containing the ST555 and ST796
698 genomes. (C) ST555-796 pairwise core SNP differences. (D) Pairwise SNP comparisons of within
699 and between core genome nucleotide diversity of the ST555 and ST796 clades. Y-axis depicts
700 the number of SNP differences, error bars indicate one standard deviation above and below the
701 mean and points represent the minimum and maximum values.

702

703

704 **Fig. 4**



705

706 **Fig 4:** Prophage gene content comparisons: the presence and absence of orthologs within each
707 of the five phages that were identified in the Ef_aus0233 genome. The phylogeny depicts the
708 evolutionary relationships among the genomes. Colours indicate the MLST designations.

709

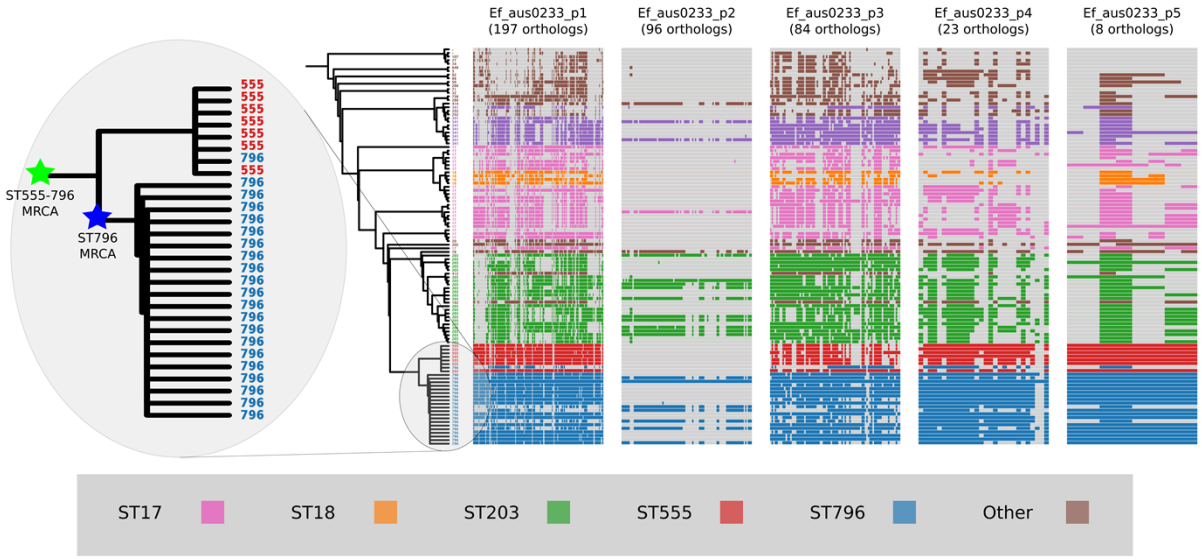
710

711

712 **Fig. 5**

713

714



715

716 **Fig 5:** Plasmid gene content comparisons: the presence and absence of orthologs within each of
717 the five plasmids that were identified in the Ef_aus0233 genome. The phylogeny depicts the
718 evolutionary relationships among the genomes. Colours indicate the MLST designations.

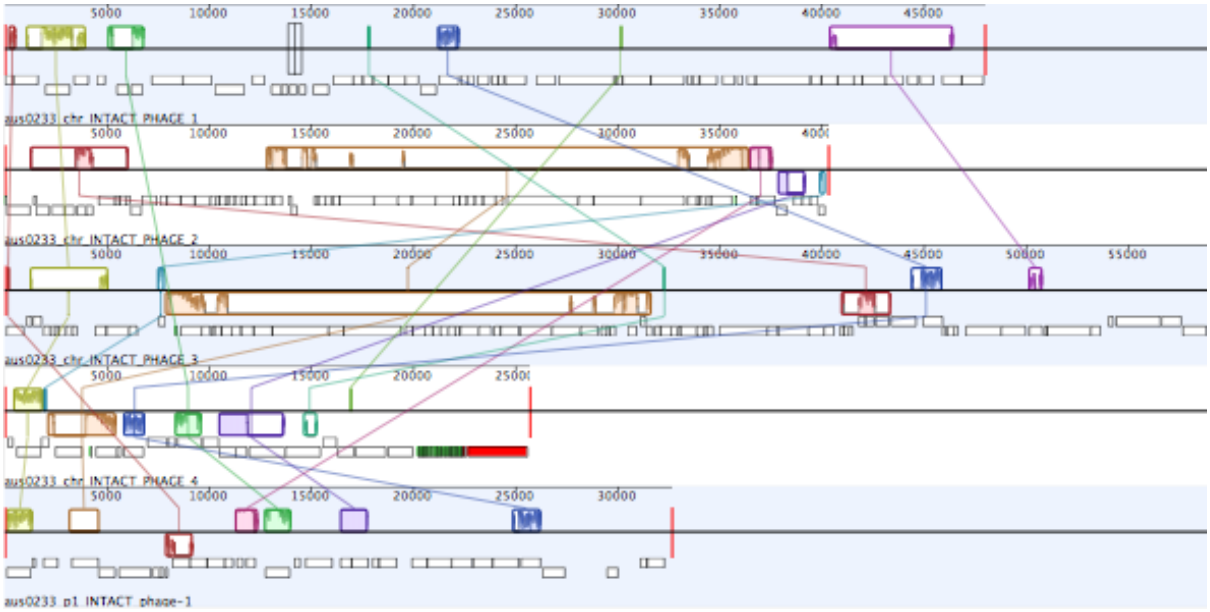
719

720

721

722 **Supplementary Fig. S1**

723



724

725 **Supplementary Fig. 1**: Mauve alignment of prophages detected in the Ef_aus0233 genome.

726

727