MICROBIAL GENOMICS

RESEARCH ARTICLE

Top *et al.*, *Microbial Genomics* 2020;6

DOI 10.1099/mgen.0.000488

Genomic rearrangements uncovered by genome-wide co-evolution analysis of a major nosocomial pathogen, *Enterococcus faecium*

Janetta Top[1,*], Sergio Arredondo-Alonso[1], Anita C. Schürch[1], Santeri Puranen[2,3], Maiju Pesonen[2,3]†, Johan Pensar[3]‡, Rob J. L. Willems[1] and Jukka Corander[3,4,5]

Abstract

Enterococcus faecium is a gut commensal of the gastro-digestive tract, but also known as nosocomial pathogen among hospitalized patients. Population genetics based on whole-genome sequencing has revealed that *E. faecium* strains from hospitalized patients form a distinct clade, designated clade A1, and that plasmids are major contributors to the emergence of nosocomial *E. faecium*. Here we further explored the adaptive evolution of *E. faecium* using a genome-wide co-evolution study (GWES) to identify co-evolving single-nucleotide polymorphisms (SNPs). We identified three genomic regions harbouring large numbers of SNPs in tight linkage that are not proximal to each other based on the completely assembled chromosome of the clade A1 reference hospital isolate AUS0004. Close examination of these regions revealed that they are located at the borders of four different types of large-scale genomic rearrangements, insertion sites of two different genomic islands and an IS*30*-like transposon. In non-clade A1 isolates, these regions are adjacent to each other and they lack the insertions of the genomic islands and IS*30*-like transposon. Additionally, among the clade A1 isolates there is one group of pet isolates lacking the genomic rearrangement and insertion of the genomic islands, suggesting a distinct evolutionary trajectory. *In silico* analysis of the biological functions of the genes encoded in three regions revealed a common link to a stress response. This suggests that these rearrangements may reflect adaptation to the stringent conditions in the hospital environment, such as antibiotics and detergents, to which bacteria are exposed. In conclusion, to our knowledge, this is the first study using GWES to identify genomic rearrangements, suggesting that there is considerable untapped potential to unravel hidden evolutionary signals from population genomic data.

DATA SUMMARY

Raw core-genome alignment (1.1 MB, Harvest suite v1.1.2), including the 1644 clade A isolates and the complete *E. faecium* AUS0004 (accession number CP003351) as a reference, is available under the following GitLab repository https://gitlab.com/sirarredondo/efm_gwes.

INTRODUCTION

Enterococcus faecium are commensals of the gastrointestinal tract but are now recognized as a major causative agent of healthcare-associated infections [1]. The transition from commensal to nosocomial pathogen coincided with increased resistance to antibiotics [2]. Since the early 1980s, *E. faecium* first gained high-level resistance to ampicillin, followed by resistance to aminoglycosides, fluoroquinolones

Received 14 May 2020; Accepted 16 November 2020; Published 30 November 2020

Author affiliations: [1]Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, the Netherlands; [2]Department of Computer Science, Aalto University, FI-00076 Espoo, Finland; [3]Department of Mathematics and Statistics, Helsinki Institute of Information Technology (HIIT), FI-00014 University of Helsinki, Finland; [4]Pathogen Genomics, Wellcome Trust Sanger Institute, Cambridge CB10 1SA, UK; [5]Department of Biostatistics, University of Oslo, 0317 Oslo, Norway.

***Correspondence:** Janetta Top, j.top@umcutrecht.nl

Keywords: *Enterococcus faecium*; genome-wide co-evolution analysis; genomic rearrangement.

Abbreviations: DUF, domains of unknown function; GWES, genome-wide co-evolution study; LD, linkage disequilibrium; SNP, single nucleotide polymorphism; SuperDCA, direct coupling analysis tool; WGS, whole genome sequencing.

†Present address: Oslo Centre for Biostatistics and Epidemiology (OCBE), Oslo University Hospital Research Support Services, Oslo, Norway

‡Present address: Department of Mathematics, University of Oslo, 0316 Oslo, Norway.

Raw core-genome alignment is available under the following https://gitlab.com/sirarredondo/efm_gwes.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Four supplementary tables and two supplementary figures are available with the online version of this article.

and glycopeptides, particularly vancomycin [3, 4]. Previous whole-genome sequencing (WGS) studies identified that the *E. faecium* population can be divided into two lineages, including a hospital-associated clade (clade A) and a community-related clade (clade B) [5, 6]. Later, clade A was subdivided into clade A1, representing the majority of hospital-associated isolates, and clade A2 for animal-related isolates [7], although two studies that included a larger collection of animal-related isolates suggested that these isolates clustered in polyphyletic groups and not in a distinct clade A [8, 9]. In a recent study also including 55 isolates from Latin America, phylogenomic analysis suggested that the animal isolates represent multiple lineages that diverged prior to the emergence of the clinical subclades in clade A [10].

Recently, we determined the plasmid content (plasmidome) of 1644 *E. faecium* isolates from different sources, countries and years using short- and long-read whole-genome sequencing technologies in combination with machine-learning classifiers [9, 11]. This analysis revealed that the hospital-associated isolates carried a larger number of plasmid sequences compared to isolates from other sources and different configurations of plasmidome populations in the hospital environment. In addition, the source specificity was determined for the chromosomal, plasmid and whole-genome components. It was concluded that plasmid sequences make the greatest contribution to source specificity.

In this paper, WGS of the same 1644 isolates was used with the aim of identifying signals of selection acting to shape co-evolution of single-nucleotide polymorphisms (SNPs). Co-evolved SNPs can facilitate adaptation to different environments when sequentially selected mutations in adaptive elements decrease the actual fitness costs of individual mutations (antagonistic epistasis) and thus have a beneficial effect on fitness [12]. For this purpose, we used the Direct Coupling Analysis tool (SuperDCA) as previously described for *Streptococcus pneumoniae*, where previously undetected epistatic interactions related to e.g. survival of the pneumococcus at lower temperatures were discovered [13]. For this study, SuperDCA was applied on the core-genome alignment of *E. faecium* to identify likely candidates of sequentially selected or coupled mutations not in strong linkage disequilibrium due to chromosomal proximity [13].

Impact Statement

Enterococcus faecium has emerged as an important nosocomial pathogen around the world. Population genetics revealed that clinical *E. faecium* strains form a distinct clade, designated clade A1, and that plasmids are major contributors to the emergence of nosocomial *E. faecium*. Here, the adaptive evolution of *E. faecium* was further explored using an unsupervised machine learning method (SuperDCA) to identify genome-wide co-evolving single-nucleotide polymorphisms (SNPs). We identified three genomic regions harbouring large numbers of SNPs in tight linkage that are separated by a large chromosomal distance in a clinical clade A1 reference isolate, but appeared adjacent to each other in non-clade A1 isolates. We identified four different types of large-scale genomic rearrangements and in all cases we found insertion of two different genomic islands and an insertion element at the border. In contrast, no genomic rearrangement and insertions were identified among a group of clade A1 pet isolates, suggesting a distinct evolutionary trajectory. Based on the *in silico* predicted biological functions, we found a common link to a stress response for the genes encoded in three regions. This suggests that these rearrangements may reflect adaptation to the stringent conditions in the hospital environment, such as antibiotics and detergents, to which bacteria are exposed.

RESULTS

Co-evolution of core genomic variation

SuperDCA was applied to a core-genome alignment generated using the Harvest suite, including 1644 clade A isolates [Table S1 (available in the online version of this article), accession number PRJEB28495] [9] and the complete *E. faecium* genome of strain AUS0004 (accession number CP003351) as reference [14]. All locus tags mentioned below refer to AUS0004. SuperDCA identified 262877 significant couplings between SNPs with a distance of >1 kbp (Table S2). We generated a top-10 list of SNP positions with the highest number of links using non-overlapping bins of 100 bp (Fig. S1a, b). To obtain an insight into the distribution on the AUS0004 core genome, all links for each of these 10 SNP positions were plotted (Fig. S1c). In this paper, we will mainly focus on the SNP positions from bins 3 and 10 (Fig. S1b) representing genes *EFAU004_02176* and *EFAU004_02173*, because we were able to identify a possible/plausible biological explanation. For both bins, we observed a similar pattern of coupled SNPs with the genes *EFAU004_00665* to *EFAU004_00675* located at a large chromosomal distance of around 500 kbp in the AUS0004 core genome (Fig. S1c). Detailed examination of the total list of coupled SNPs revealed similar patterns of coupled SNPs for a larger region encompassing genes *EFAU004_02173* to *EFAU004_02178,* referred to as region-1 (Fig. 1a). There were in total 2323 coupled SNP positions, i.e. between 181 different SNP positions from region-1 and 131 different SNP positions in genes *EFAU004_00665* to *EFAU004_00675*, referred to as region-2 (Figs 1a and S1c, Table S3). In addition, all coupled SNPs for region-1 and region-2 were retrieved from the total list. These 17236 linked SNPs were plotted on the core genome to obtain an insight into the distribution of the links (Fig. 1a). This revealed a cluster of a total of 142 linked SNPs, represented by 32 different SNP positions in region-2 and 25 SNP positions in genes *EFAU004_02092* to *EFAU004_02100,* representing region-3, located close to region-1 (Fig. 1a, Tables S3 and S4).

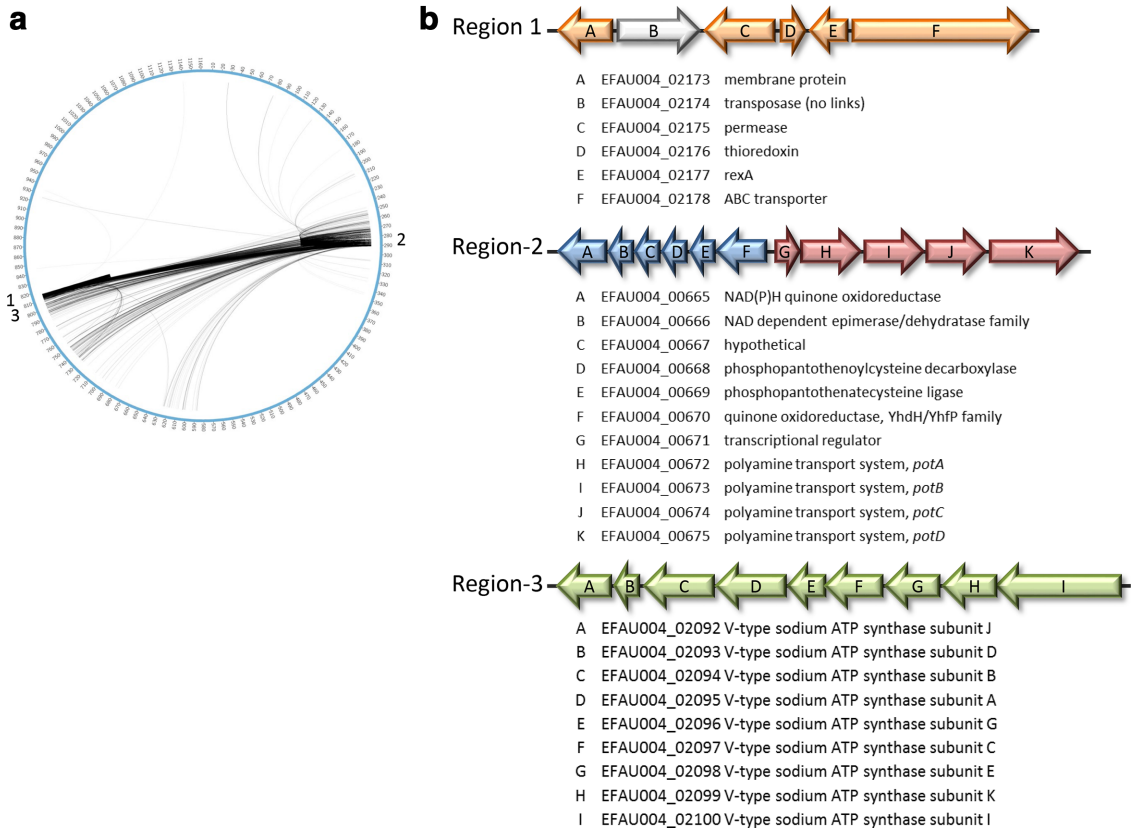

Fig. 1. (a) All 17236 epistatic links of the genes contained in region-1, region-2 and region-3. (b) Genomic organization and annotation of genes from regions 1–3 in *E. faecium* AUS0004.

In order to elucidate a possible biological explanation for the identified links between region-1, region-2 and region-3, we first determined putative biological functions for the proteins in these regions using a homology search for similarity with other proteins and investigated domains with known function.

Putative biological functions for proteins in region-1

Region-1 contains five protein-encoding genes and a transposase with no links (Fig. 1b). For two proteins a function prediction is challenging, i.e. EFAU004_02173 is annotated as a membrane protein, but there are only domains of unknown function (DUF), while EFAU004_02175 is annotated as a permease with unknown specificity. In contrast, EFAU004_02176 contained conserved domains belonging to the thioredoxin-like family (Trx-like) and EFAU004_02177 contained conserved domains belonging to the Rex (Rex-like) family of transcriptional regulators (Fig. 1b). Trx-like and Rex-like are likely involved in redox homeostasis of the bacterial cell, which was found to be critical for DNA synthesis and defence against oxidative stress [15]. The exact function of the ABC transporter (EFAU004_02178) is difficult to predict despite the presence of several domains. The protein contains two so-called AAA domains and a leucine-zipper (bZIP) domain. Proteins with AAA domains are members of a conserved family of ATP-hydrolyzing proteins with all kind of activities in many cellular pathways, including replication, DNA and protein transport, transcriptional regulation, ribosome biogenesis, membrane fusion and protein disaggregation or degradation [16]. bZIP domains are known to be involved in transcriptional regulation, e.g. in stress conditions such as heat in *Salmonella* [17, 18], or abiotic stress in eukaryotes, e.g. plants [19]. Furthermore, homology search against the GenBank database to non-*Enterococcaceae* revealed co-localization of a similar ABC transporter and *rexA* gene in other species such as *S. pneumoniae* (strain 2842STDY5753625, accession number FEGY01000003), *Streptococcus agalactiae* (strain DK-PW-092, accession number LBKE01000082) and *Listeria monocytogenes* (strain CFSAN060067, accession number AABAZK010000087), suggesting that this ABC transporter might also be involved in some kind of stress response.

Putative biological functions for proteins in region-2

The genes from region-2 can be split into two distinct gene clusters based on their putative biological functions (Fig. 1b). The first cluster encompasses genes *EFAU004_00665* to *EFAU004_00670*, which, based on their annotation, are putatively involved in respiration and coenzyme A biosynthesis

(Fig. 1b). Although SNPs linked with region-1 were observed with all these genes, we will focus on two specific genes, i.e. *EFAU004_00665* and *EFAU004_00670*, as they contained the majority of linked SNPs (Tables S3 and S4). Very similar conserved domains for NAD(P)H : quinone oxidoreductase were identified in both proteins. EFAU004_00670 contains transmembrane helices and is therefore likely membrane bound, while EFAU004_00665 lacks these transmembrane helices and is therefore likely to be soluble. The conserved domains in EFAU004_00670 had a high similarity to YhdH of *Escherichia coli* (overall amino acid similarity of 48%). The conserved domains in EFAU004_00665 were very similar to QorEc of *E. coli* (overall amino acid similarity of 32%) and QorTt *Thermus thermophiles* (overall amino acid similarity of 32%). In *Staphylococcus aureus*, expression of a gene cluster containing two *qor*-like genes was induced under oxidative stress conditions [20]. These two Qor proteins (SA1988 and SA1989) are predicted to be soluble and SA1988 contained similar conserved domains with an overall amino acid similarity of 26% with EFAU004_00665. This might suggest that this protein could also be involved in stress response.

The second gene cluster contains genes *EFAU004_00671* to *EFAU004_00675* and is likely organized as an operon (Fig. 1b). EFAU004_00671 contains a helix–turn–helix motif and is therefore predicted to be a transcriptional regulator. Its location upstream of *EFAU004_00672* to *EFAU004_00675* suggests that it regulates the expression of these genes. EFAU004_00672 to EFAU004_00675 display similarity with a polyamine transport system, as described for *S. pneumoniae,* including an ATP-binding protein, PotA, two permeases, PotB and PotC, and a substrate-binding protein, PotD [21]. Polyamines are polycationic molecules and are required for optimal growth in both eukaryotic and prokaryotic cells and are implicated in pathogenicity of *S. pneumoniae*. In *S. pneumoniae* polyamines are pivotal in survival strategies in the host when bacteria are confronted with stress conditions such as temperature shock, oxidative stress, or choline limitation [21].

Putative biological functions for proteins in region-3

Genes *EFAU004_02092* to *EFAU004_02100* are annotated to encode a membrane bound V-type ATPase (Fig. 1b). An identical V-type ATPase has been studied in detail in *Enterococcus hirae* (100% amino acid identity) [22]. The V-type ATPase belongs to the family of proton pumps and is involved in the translocation of Na+ or H+ over the cell membrane by using the energy of ATP. In *E. hirae* the V-type ATPase appeared to be highly expressed under stress conditions such as high pH and plays an important role in sodium homeostasis under these conditions.

Genomic rearrangements in completed *E. faecium* genomes

For 48 of 1644 isolates, a fully assembled circular chromosome was available (accession number PRJEB28495) as described previously [11]. This allowed us to assess the genomic organization in these isolates. Using the non-clade A1 isolate E0139 as the reference, 4 different types of large chromosomal rearrangements were observed among 38 isolates. Type-1 (n=1) involves an inversion of 0.73 Mbp; type-2 (n=34) an inversion of 1.2 Mbp; type-3 (n=2) two inversions of 0.38 Mbp; and type-4 (n=1) an inversion of 0.12 Mbp (Figs 2a–d). Except for the type-1 inversion, all these genomic rearrangements were exclusively observed in clade A1 isolates (Fig. 3). In clade A1 isolates we detected two previously described genomic islands, inserted adjacent to the genomic rearrangements that were described as being enriched among clade A1 hospital isolates (Fig. 2b–d) [23, 24]. One genomic island, putatively encoding a carbohydrate transport and metabolism pathway [23], was always found to be located downstream of the ABC transporter (*EFAU004_02178*) of region-1 (Fig. 4a). The other island encoding a phosphotransferase system (PTS) [24] was always found to be located downstream *potD* (*EFAU004_00675*) of the polyamine transport system of region-2 (Fig. 4a). In all cases, the genomic rearrangement was flanked by an IS*30*-like transposon downstream of a methionine synthase (*metE*) and a pyridine nucleotide-disulfide oxidoreductase (*pyr*) (Fig. 4a). In the 10 isolates for which a fully assembled circular chromosome was available and that lacked the genomic rearrangement, insertions of both genomic islands and IS*30*-like transposon, region-1 and region-2 are located adjacent to each other as schematically indicated for the non-clade A1 isolate *E. faecium* E0139 (Fig. 4b).

We determined the presence of the two genomic islands and their insertion sites in the draft genomes of the remaining 1596 isolates and confirmed that the genomic islands were inserted at the same position in 93% of the clade A1 isolates (Fig. 3). In addition, the insertions were also identified in 57 (14%) non-clade A1 isolates, which mainly represented dog isolates (n=43) (Fig. 3). In contrast, a branch in clade A1 consisting of 68 genetically closely related isolates, including 55 pet, 9 isolates from hospitalized patients and 4 from non-hospitalized persons, lacked both genomic island insertions (Fig. 3). Based on these draft genomes, we cannot determine whether or which genomic rearrangement could have occurred, though we always observed a contig break downstream of *metE* and *pyr*, strongly suggesting the presence of an IS element on that position. In isolates that lacked insertions of the two genomic islands the genomes were organized in a manner comparable to the non-clade A1 isolate E0139, i.e. without inversion (Figs 3 and 4b).

DISCUSSION

In this study, we identified large genomic rearrangements, enriched in *E. faecium* hospital-associated clade A1 strains, which were uncovered by the appearance of co-evolved SNPs in three chromosomal regions; region-1, region-2 and region-3. These chromosomal regions were located on the borders of the genomic rearrangement and therefore adjacent to each other in the majority of non-clade A1 isolates, but at a large distance in clade A1 isolates. In all cases, the

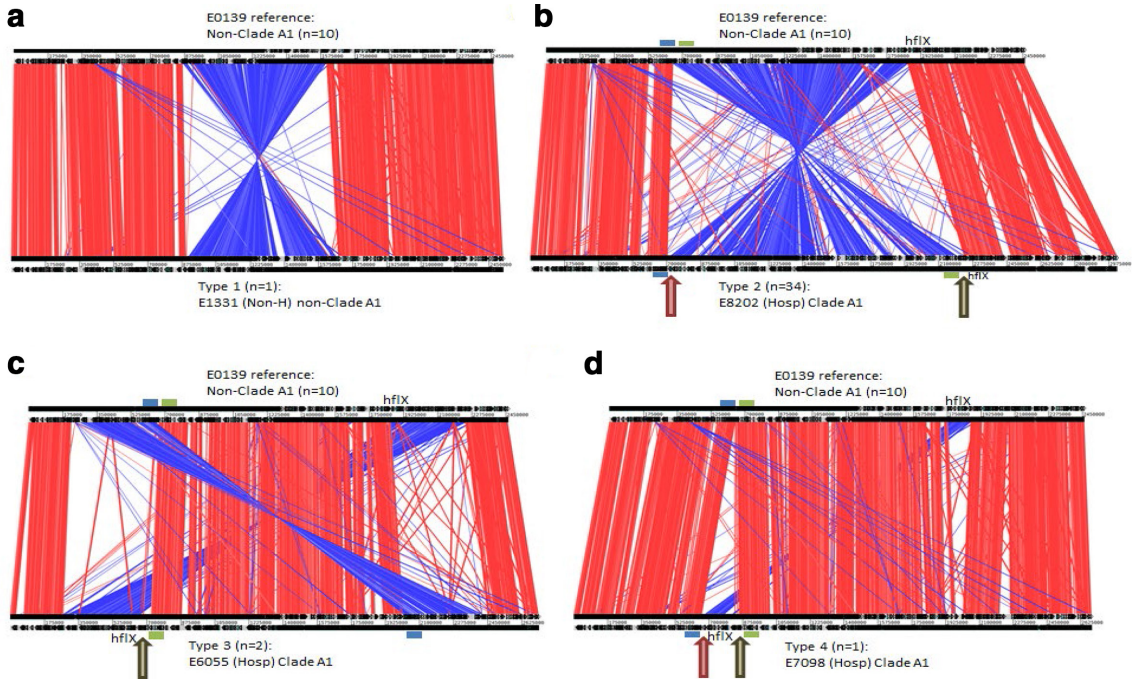

Fig. 2. Genome comparisons using strain E0139 (non-clade A1) as a reference. (a) E0139 compared to type 1 strain E1334; (b) E0139 compared to type 2 strain E8202; (c) E0139 compared to type 3 strain E6055; (d) E0139 compared to type 4 strain E7098. 'n' indicates the number of isolates with a similar genomic organization and for which a complete chromosome was available. Same-strand DNA similarity is shaded red, while reverse similarity is shaded blue. Blue bar, position of the *pot* operon; green bar, position of the ABC transporter; red arrow, insertion site for the phosphotransferase system (PTS)-encoding genomic island [24]; brown arrow, insertion site for the carbohydrate transport system-encoding genomic island [23].

genomic rearrangement is flanked by an IS*30*-like transposon downstream from *metE* and *pyr*, which suggests that the rearrangement is the result of recombination between two IS*30*-like elements. In addition, upstream of *metE* and *pyr*, we identified the insertion of two, clade A1 enriched, genomic islands, encoding a PTS and carbohydrate transport system, respectively [23, 24]. For both genomic islands, it has been suggested that they may provide a fitness advantage, which was confirmed for PTS in an *in vivo* mouse colonization model [24]. Recently, Yan *et al.* reconstructed the evolution of the marine photosynthetic micro-organism *Prochlorococcus* based on genome rearrangements and identified rearrangement hotspots that were all in the vicinity of genomic islands [25]. In fact, the authors suggest that the genomic islands serve as hotspots that induce genome rearrangement. In addition, they observed that different clades shared a conserved backbone, but also contained clade-specific regions, which were associated with ecological adaptations. Further, for *Pseudomonas putida*, a Gram-negative bacterium that can be found in different environments, it was shown that horizontal gene transfer played a key role in adaptation process, as many of the niche-specific functions were found to be encoded on clearly defined genomic islands and were linked to genomic rearrangements [26]. The observation that among the clade A1 isolates only the pet isolates do not contain the two genomic islands and genomic rearrangement may suggest that these elements are not advantageous in the niche represented by the gut of pet animals.

The predominant chromosomal rearrangement, type 2, identified in 71% of the fully assembled genomes from the current study, is also present in other publicly available completed genomes, such as *E. faecium* DO (NC_017960.1) [27], *E. faecium* Aus0085 (NC_021994.1) [28] or *E. faecium* E39 (NZ_CP011281.1). For the isolates in our study for which we only have a draft genome it is difficult to determine whether genomic rearrangements have occurred. However, in isolates in which *metE* or *pyr* were not adjacent to each other, as in the non-clade A1 configuration, we always identified an IS*30*-like element downstream from the *metE* or *pyr* gene at the border of a contig, suggesting a potential genomic rearrangement. A chromosomal rearrangement in *E. faecium* was also observed in the first completely sequenced isolate AUS0004, where it was hypothesized that it occurred between two phage elements, resulting in a replichore imbalance [14]. However, we did not observe a replichore imbalance in any of the 38 completely assembled genomes with large chromosomal rearrangements (data not shown), which suggests that AUS0004 is an exception in that regard.

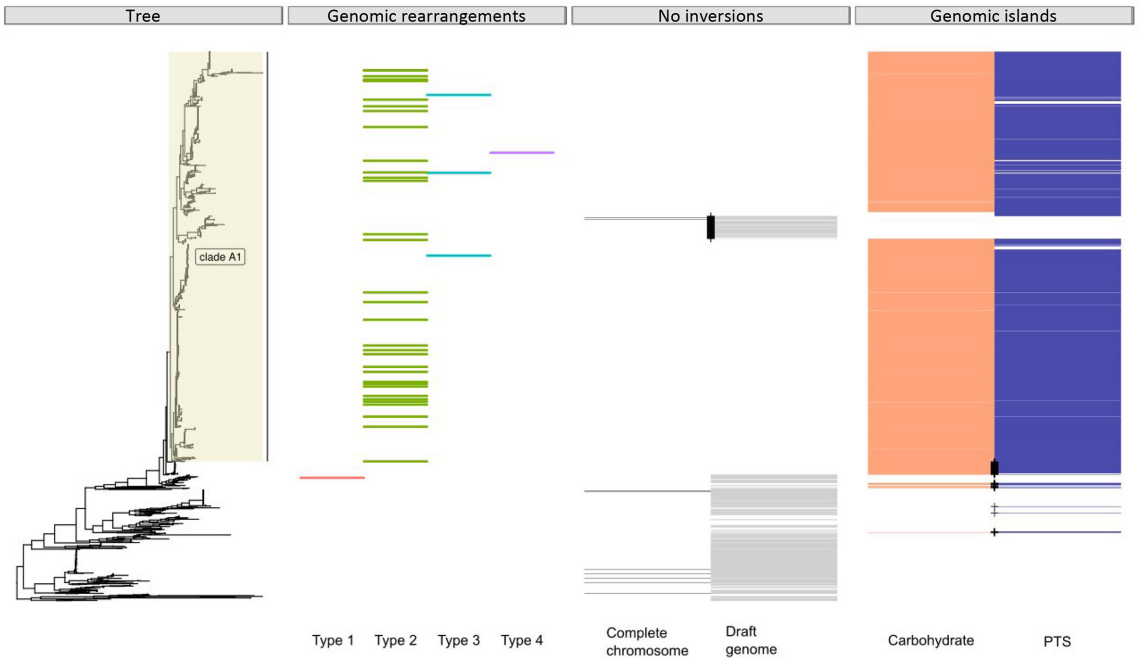

Fig. 3. Core-genome tree based on 1644 strains with the 3 metadata panels: (i) distribution of genomic rearrangements types among 38 complete chromosomes, (ii) indication of strains lacking a genomic rearrangement among 10 complete chromosomes and draft genomes and (iii) indication of strains with insertion of a carbohydrate transport system encoding genomic island, orange [23], or a phosphotransferase system (PTS) encoding genomic island, purple [24]. Black cross, indication of clade A1 dog isolates that lack the insertions/inversion or non-clade A1 dog isolates that do contain insertions/inversion.

Genomic rearrangements or inversions are not rare among prokaryotic genomes, as was investigated by Repar *et al.* [29]. The authors developed an alignment-based approach to systematically assess inversion symmetry between pairs of prokaryotic genomes, including the genera *Bacillus* and *Clostridium*, which belong to the phylum Firmicutes like *E. faecium*. *Bacillus* and *Clostridium* are dominated by X-shaped inversions that are symmetric around the *ori–ter* axis, similar to the type 1–3 inversions from this study.

The gene clusters located in the chromosomal regions-1, -2 and -3 that harboured co-evolved SNPs clearly encode common functions. The thioredoxin pathway, Rex-like (region-1) and the NAD(P)H-dependent oxidoreductase/epimerase proteins (region-2) are all dependent on the oxidation/reduction of NAD+/NADH and involved in redox homeostasis, while the polyamine transport system (region-2) and V-type ATPase (region-3) are both involved in ATP hydrolysis/synthesis. The fact that in non-clade A1 isolates chromosomal regions-1,-2

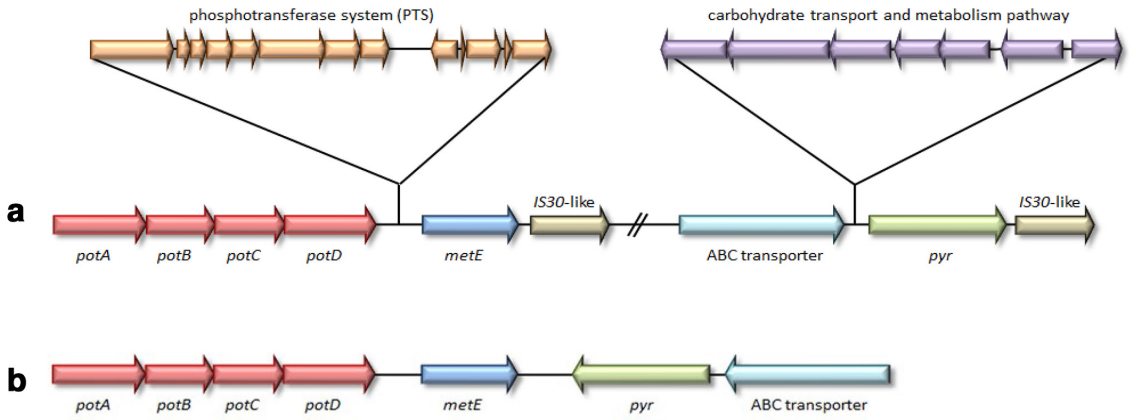

Fig. 4. (a) Genomic organization of the most predominant type 2 genomic rearrangement in strain E8202 with indication of insertion and inversion sites. (b) Genomic organization for reference strain E0139.

and -3 are located adjacent to each other explains the high number of statistically significantly linked SNPs between the genes in these two regions, as they likely have co-evolved in fairly tight linkage disequilibrium (LD). Conversely, this finding illustrates that the co-evolutionary analysis of SNPs can uncover inversions in the chromosome. Enterococci are rich in mobile genetic elements harbouring IS elements and transposons, many of which can integrate into the chromosome [3, 30] and may drive the observed rearrangements. These rearrangements may play a role in adaptation to new environments, more particularly adaptation to stress conditions, like the hospital environment, where bacteria are exposed to different stressors, including high concentrations of antibiotics, detergents and antiseptics. Furthermore, it is possible that the rearrangements are associated with the rapid adaptation of the organism to survive in the perturbed microbiota of hospitalized patients and that survival in the overall context of the intestinal microbiome may be an important driver of evolution. The lack of comparable selection pressure is a plausible explanation for the observed absence of genomic arrangements among the pet isolates from clade A1.

To our knowledge, identification of genomic rearrangements by direct coupling analysis of SNPs has not previously been described in the literature. The expected rapid increase in the availability of long-read sequences combined with large population-based collections of short-read sequenced genomes that are amenable to a statistical LD analysis such as performed here will thus facilitate the development of new methods to identify candidates for inversions associated with selective advantage.

METHODS

Genomic DNA sequencing and assembly

A detailed description of Illumina and ONT sequencing has been provided previously and includes a full description of ONT selection of *E. faecium* isolates (n=62) [9, 11] and consecutive hybrid assembly using Unicycler [31]. The R package ggtree (version 1.14.6) [32] was used to plot distinct metadata panels together with a previously described core genome tree of 1644 *E. faecium* strains [9].

Co-evolution analysis

A core-genome alignment (1.1 MB) was generated using the Harvest suite v1.1.2, including the 1644 clade A isolates and the complete *E. faecium* AUS0004 (accession number CP003351 [14]) as reference. SuperDCA was performed with the default settings as described by Puranen *et al.* [13] using the alignment of 1644 isolates, which was unfiltered for recombination events. Circos was used for the visualization of coupled SNPs on the genome [33].

Prediction of putative biological functions of proteins from regions 1–3

To predict putative biological functions of proteins, the protein sequences were compared with the non-redundant public database using the National Center for Biotechnology Information (NCBI) BLASTP server (www.ncbi.nlm.nih.gov/BLAST/) [34]. In addition, a graphical summary was used to determine whether putative conserved domains were identified and, if present, this graphical summary was selected to investigate the list of conserved domains in more detail in order to make a prediction regarding the putative biological function of the protein.

E. faecium genome organization

To unravel rearrangements in the chromosome, we considered a non-clade A1 isolate from a non-hospitalized person (E0139) with a complete chromosome sequence as reference. We only considered isolates with a complete circular chromosome (n=48). Pairwise chromosomal comparisons were computed using BLASTN (version 2.7.1+) and alignments were visualized using the Artemis Comparison tool (version 17.0.1). The insertions of two genomic islands encoding a PTS [24] and carbohydrate transport system [23] and their putative co-localization with the *pot* operon and ABC transporter, respectively, were determined by BLAST. Additionally, for each chromosomal rearrangement, we generated dotplots using Gepard (version 1.40) [35] with the non-clade A1 isolate E0139 as a reference. Clone manager 9 was used to visualize the genomic organization of regions 1–3.

Funding information
S. A. and R. J. L. W.: this study was supported by the Joint Programming Initiative in Antimicrobial Resistance (JPIAMR Third call, STARCS, JPIAMR2016-AC16/00039). J. C. was funded by the European Research Council (grant no. 742158).

Author contributions
Conceptualization, J. C. and R. J. L. W.; software, S. P., M. P. and J. P.; formal analysis, S. A., J. T. and A. C. S.; writing of original draft, J. T., S. A., A. C. S., J. C. and R. J. L. W.; funding acquisition, R. J. L. W. and J. C.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Weiner LM, Webb AK, Limbago B, Dudeck MA, Patel J *et al*. Antimicrobial-Resistant pathogens associated with healthcare-associated infections: summary of data reported to the National healthcare safety network at the centers for disease control and prevention, 2011-2014. *Infect Control Hosp Epidemiol* 2016;37:1288–1301.
2. Lebreton F, Willems RJL, Gilmore MS. Enterococcus diversity, origins in nature, and gut colonization. In: Enterococci: from commensals to leading causes of drug resistant infection [Internet] 2014:1–56.
3. Gilmore MS, Lebreton F, van Schaik W. Genomic transition of enterococci from gut commensals to leading causes of multidrug-resistant hospital infection in the antibiotic era. *Curr Opin Microbiol* 2013;16:10. ––16p..
4. Guzman Prieto AM, van Schaik W, Rogers MRC, Coque TM, Baquero F *et al*. Global emergence and dissemination of enterococci as nosocomial pathogens: attack of the clones? *Front Microbiol* 2016;7:788.
5. Galloway-Peña J, Roh JH, Latorre M, Qin X, Murray BE. Genomic and SNP analyses demonstrate a distant separation of the hospital and community-associated clades of *Enterococcus faecium*. *PLoS One* 2012;7:e30187.

6. Palmer KL, Godfrey P, Griggs A, Kos VN, Zucker J *et al*. Comparative genomics of enterococci: variation in *Enterococcus faecalis*, clade structure in *E. faecium* and defining characteristics of *E.gallinarum* and *E.casseliflavus*. *MBio* 2012;3:1–11.
7. Lebreton F, van Schaik W, McGuire AM, Godfrey P, Griggs A *et al*. Emergence of epidemic multidrug-resistant *Enterococcus faecium* from animal and commensal strains. *mBio* 2013;4:e00534-13 [Epub ahead of print 20 Aug 2013].
8. Raven KE, Reuter S, Reynolds R, Brodrick HJ, Russell JE *et al*. A decade of genomic history for healthcare-associated *Enterococcus faecium* in the United Kingdom and Ireland. *Genome Res* 2016;26:1388–1396.
9. Arredondo-Alonso S, Top J, McNally A, Puranen S, Pesonen M *et al*. Plasmids shaped the recent emergence of the major nosocomial pathogen *Enterococcus faecium*. *mBio* 2020;11:1–17.
10. Rios R, Reyes J, Carvajal LP, Rincon S, Panesso D *et al*. Genomic epidemiology of Vancomycin-Resistant *Enterococcus faecium* (VREfm) in Latin America: revisiting the global VRE population structure. *Sci Rep* 2020;10:5636 [Epub ahead of print Available from].
11. Arredondo-Alonso S, Rogers MRC, Braat JC, Verschuuren TD, Top J *et al*. mlplasmids: a user-friendly tool to predict plasmid- and chromosome-derived sequences for single species. *Microb Genom* 2018;4 [Epub ahead of print 01 11 2018].
12. Lagator M, Colegrave N, Neve P. Selection history and epistatic interactions impact dynamics of adaptation to novel environmental stresses. *Proc Biol Sci* 2014;281:20141679 [Epub ahead of print Available from].
13. Puranen S, Pesonen M, Pensar J, Xu YY, Lees JA *et al*. SuperDCA for genome-wide epistasis analysis. *Microb Genom* 2018;4 [Epub ahead of print 29 05 2018].
14. Lam MMC, Seemann T, Bulach DM, Gladman SL, Chen H *et al*. Comparative analysis of the first complete *Enterococcus faecium* genome. *J Bacteriol* 2012;194:2334–2341.
15. Wang E, Bauer MC, Rogstam A, Linse S, Logan DT *et al*. Structure and functional properties of the *Bacillus subtilis* transcriptional repressor Rex. *Mol Microbiol* 2008;69:466–478.
16. Elsholz AKW, Birk MS, Charpentier E, Turgay K. Functional diversity of AAA+ protease complexes in *Bacillus subtilis*. *Front Mol Biosci* 2017;4:1–15.
17. Hurme R, Berndt KD, Namork E, Rhen M. DNA binding exerted by a bacterial gene regulator with an extensive coiled-coil domain. *J Biol Chem* 1996;271:12626–12631.
18. Hurme R, Berndt KD, Normark SJ, Rhen M. A proteinaceous gene regulatory thermometer in *Salmonella*. *Cell* 1997;90:55–64.
19. Alves MS, Dadalto SP, Gonçalves AB, de Souza GB, Barros VA *et al*. Transcription factor functional protein-protein interactions in plant defense responses. *Proteomes* 2014;2:85–106.
20. Maruyama A, Kumagai Y, Morikawa K, Taguchi K, Hayashi H *et al*. Oxidative-stress-inducible *qorA* encodes an NADPH-dependent quinone oxidoreductase catalysing a one-electron reduction in *Staphylococcus aureus*. *Microbiology* 2003;149:389–398.
21. Shah P, Romero DG, Swiatlo E. Role of polyamine transport in *Streptococcus pneumoniae* response to physiological stress and murine septicemia. *Microb Pathog* 2008;45:167–172.
22. Murata T, Kawano M, Igarashi K, Yamato I, Kakinuma Y. Catalytic properties of Na(+)-translocating V-ATPase in *Enterococcus hirae*. *Biochim Biophys Acta* 2001;1505:75–81.
23. Heikens E, van Schaik W, Leavis HL, Bonten MJM, Willems RJL. Identification of a novel genomic island specific to hospital-acquired clonal complex 17 *Enterococcus faecium* isolates. *Appl Environ Microbiol* 2008;74:7094–7097.
24. Zhang X, Top J, de Been M, Bierschenk D, Rogers M *et al*. Identification of a genetic determinant in clinical *Enterococcus faecium* strains that contributes to intestinal colonization during antibiotic treatment. *J Infect Dis* 2013;207:1780–1786.
25. Yan W, Wei S, Wang Q, Xiao X, Zeng Q *et al*. Genome rearrangement shapes *Prochlorococcus* ecological adaptation. *Appl Environ Microbiol* 2018;84:e01178–18.
26. Wu X, Monchy S, Taghavi S, Zhu W, Ramos J *et al*. Comparative genomics and functional analysis of niche-specific adaptation in *Pseudomonas putida*. *FEMS Microbiol Rev* 2011;35:299–323.
27. Qin X, Galloway-Peña JR, Sillanpaa J, Roh JH, Nallapareddy SR *et al*. Complete genome sequence of *Enterococcus faecium* strain TX16 and comparative genomic analysis of *Enterococcus faecium* genomes. *BMC Microbiol* 2012;12:1.
28. Lam MMC, Seemann T, Tobias NJ, Chen H, Haring V *et al*. Comparative analysis of the complete genome of an epidemic hospital sequence type 203 clone of vancomycin-resistant *Enterococcus faecium*. *BMC Genomics* 2013;14:1.
29. Repar J, Warnecke T. Non-Random inversion landscapes in prokaryotic genomes are shaped by heterogeneous selection pressures. *Mol Biol Evol* 2017;34:1902–1911.
30. Hegstad K, Mikalsen T, Coque TM, Werner G, Sundsfjord A. Mobile genetic elements and their contribution to the emergence of antimicrobial resistant *Enterococcus faecalis* and *Enterococcus faecium*. *Clin Microbiol Infect* 2010;16:541–554 [Epub ahead of print Available from].
31. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13:e1005595–22.
32. Yu G, Smith DK, Zhu H, Guan Y, Lam Tommy Tsan-Yuk. ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 2017;8:28–36.
33. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R *et al*. Circos: an information aesthetic for comparative genomics. *Genome Res* 2009;19:1639–1645.
34. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.
35. Krumsiek J, Arnold R, Rattei T. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* 2007;23:1026–1028.