UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE FARMACIA

Departamento de Microbiología II

TESIS DOCTORAL

Enterococcus faecalis: nuevas perspectivas sobre la estructura poblacional y el impacto de los elementos genéticos móviles en la evolución

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Ricardo León Sampedro

Directora

María Teresa Coque González

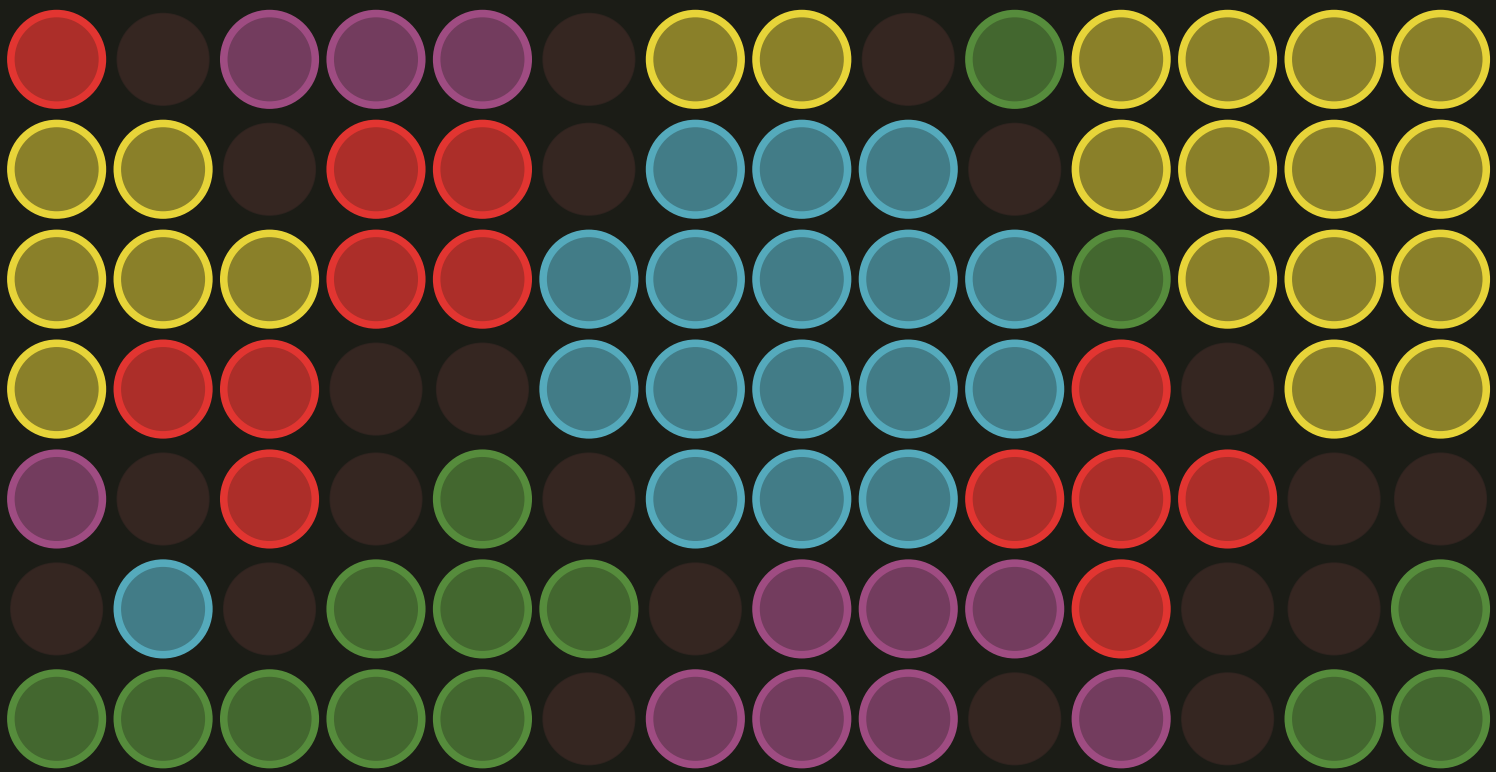
Madrid, 2018

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE FARMACIA
Departamento de Microbiología II

Enterococcus faecalis: nuevas perspectivas sobre la estructura poblacional y el impacto de los elementos genéticos móviles en la evolución

Ricardo León Sampedro

Directora:
Dra. Teresa M. Coque González



Enterococcus faecalis: new insights about population structure and the impact of mobile genetic elements on evolution

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE FARMACIA

Departamento de Microbiología II

Enterococcus faecalis: nuevas perspectivas sobre la estructura poblacional y el impacto de los elementos genéticos móviles en la evolución

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Ricardo León Sampedro

Directora

Dra. María Teresa Coque González

Madrid 2017

PI12-01581

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE FARMACIA

Departamento de Microbiología II

Enterococcus faecalis: new insights about population structure and the impact of mobile genetic elements on evolution

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Ricardo León Sampedro

Directora

Dra. María Teresa Coque González

Madrid 2017

PI12-01581

A mi madre, a mi hermano, a mi padre.

A mi abuela.

A Julián.

A mi guía y directora, Teresa.

Index of contents

Index of contents

Abstract

Abstract

Enterococcus faecalis is a generalist bacterial species that inhabitant a variety of disparate hosts such as mammals, reptiles, insects and birds, being a potential cause of severe infections in all these hosts. This species is also one of the major nosocomial pathogens and antibiotic resistance threat.

To define the structure of *E. faecalis* populations is important to identify main clonal lineages causing infections in relevant hosts. The available *E. faecalis* population structure reflects a cluster of a limited number of genotypes and a mismatch distribution dominated by mutation, with a slight increase of identical genotypes pairs due to local hospital transmission. The overrepresentation of *E. faecalis* isolates of hospitalized patients in gene databases precludes defining the population structure of those species colonizing different host organisms of poorly defined and disparate ecological patches. Most of the works published to date are based on MLST data analyzed by using goEBURST, Bayesian Analysis Population Structure (BAPS) software or cgMSLT and revealed an epidemic structure with a high recombination /mutation ratio. Such studies highlight the difficulties to establish the population structure and to elucidate the evolutionary history of some lineages. The limited number of genomes from non-human hosts and the tools to analyze species with a high recombination potential preclude confirming such hypothesis.

Wild animals are a good subject for expanded the available knowledge on population structure of *E. faecalis*. They are subjected to a variety of selective pressures (antibiotics, heavy metals, biocides) and have a high rate of network of contacts promoting the exchange and flow between different bacterial communities. In addition, novel tools to analyse accessory genomes and typing mobile genetic elements developed in our group or the consortia in which we participated would allow a comprehensive search of the *E. faecalis* pan-genome.

Moreover, extrachromosomal elements play important roles in the adaptation and evolution of bacterial populations of enterococci. They include a high diversity of elements although plasmids, integrative conjugative elements (ICEs) and conjugative transposons (CTn). In fact, paradigms of conjugation systems, namely those of Tn*916* and plasmid families Inc18 and RepA_N pheromone responsive elements were firstly recovered in *E. faecalis*.

The extended use of some antibiotics, as is the case of tetracycline, seems to have resulted in an evolutionary bottleneck in the population structure of some genera of Firmicutes of interest in human health. Tetracycline resistance in major Gram positive human opportunistic pathogens is mainly caused by the acquisition of integrative and conjugative elements of the Tn*916* family carrying the *tet*(M) gene. These Tn*916*-like elements include a wide variety of ICE with a common synteny but differing in the integration and transfer mechanisms, the insertion site specificity and the host range. Tn*5801*, a Tn*916*-like transposon, was considered a site specific element that nonetheless, had been detected in different genus of clinical interest for a very long time (staphylococci, streptococci, esnterococci).

The main objective of this dissertation is to get insights in the population structure of *E. faecalis* and in the role of the accessory genome and horizontal gene transfer in shaping the population structure and evolution of this species. For the analysis of the population structure, we studied a collection of *E. faecalis* from migratory birds in order to feed the databases with isolates recovered from non-human hosts and to compare core and accessory genome of the species to identify niche-specific accessory protein clusters. In order to understand the impact of horizontal gene transfer in the evolution of different lineages (genome structure, fitness cost and evolvability) we will take as example the evolution of Tn*5801,* analysing its variability in several Firmicutes genus by the analysis of the WGS database. Finally, we characterize the transfer events involved in the dynamics of this conjugative transposon among *E. faecalis* and their fitness in the presence of tetracycline. The thesis is divided in three main chapters focused in these objectives.

Chapter I gives new insights about the structure of *E. faecalis* populations and their niche-specific adaptations. It also provides a detailed analysis of the pangenome of this species (core and accessory genome). We confirm the limited specialization in *E. faecalis* of previous works by using BAPS and comparative genomics. However, we also demonstrated the small core and large accessory genome that would support the Black Queen hypothesis according which the species could rapidly adapt to different environments by acquisition of a large pangenome and the arsenal of mobile genetic elements that enable different horizontal gene transfer events. In this scenario, niche specific convergence between *E. faecalis* clones from different host species of wild birds emerges and resulted in clonal complexes restricted to specific lifestyles like CC82 that seems to be confined to wild birds. Within the accessory genome, the plasmids showed a high degree of mosaicism further demonstrating their ability to recombine and construct

chimeras. This high mosaicism and the lack of plasmids in some isolates suggest differences in the adaptation between hosts.

Chapter II comprehensively characterize the diversity, evolution and mobilization of Tn*5801,* a scarcely explored member of the Tn*916* family which is a paradigm of conjugative transposons. Transferability of Tn*5801* was demonstrated between different *E. faecalis* backgrounds, even if the strain already harbored a Tn*5801* element. The finding is relevant as this transposon has a site specific tyrosine recombinase that recognizes the 3´-end of guanosine monophosphate (GMP) synthase gene (*guaA*), an insertion hotspot of genomic islands in Firmicutes and can be functional at secondary integration sites.

Chapter III fully analyzed the genetic context of the Tn*5801* transconjugants to understand the mobilization events of Tn*5801* and the influence of acquiring this CTn in the bacterial fitness of *E. faecalis.* The study not only proved the unexpected transferability and integration of more than one copy of this CTn but also define a novel 53,3kb composite element designed as ICE50CE comprising Tn*5801*.B25 and a new 30kb ICE called ICEEfasa1. The work highlights the diversity of horizontal gene transfer (HGT) events, especially those of tandem accretion after site-specific recombination. Such strategy would enable the selection and persistence of some pathogenic clones.

Resumen en español

Resumen en español

Enterococcus faecalis: nuevas perspectivas sobre la estructura poblacional y el impacto de los elementos genéticos móviles en la evolución

Enterococcus faecalis es una especie bacteriana generalista que habita en una amplia variedad de hospedadores como mamíferos, reptiles insectos y aves, siendo la causa potencial de graves infecciones en todos ellos. Esta especie es también uno de los principales patógenos nosocomiales y uno de los mayores vehículos de transmisión de genes de resistencia a antibióticos.

Para definir la estructura poblacional de *E. faecalis* es importante identificar las principales líneas clonales causantes de infecciones en hospedadores relevantes. Existen estudios basados en el análisis de datos de MLST mediante la aplicación de herramientas como eBURST (Based Upon Related Sequence Types) o BAPS (Análisis Bayesiano de la Estructura Poblacional) así como de cgMLST (Core Genome Multilocus sequencing), los cuales revelan una estructura epidémica con una alta tasa de recombinación/ mutación. Estos estudios destacan las dificultades para establecer una estructura poblacional y para elucidar la historia evolutiva de algunos de los linajes clonales. Además el limitado número de genomas provenientes de hospedadores no-humanos y las escasas herramientas para analizar especies con alta recombinación impide la confirmación de esta hipótesis.

Estudiar cepas de animales salvajes es una buena aproximación para expandir el conocimiento disponible sobre la estructura poblacional de *E. faecalis*. Estos hospedadores están sometidos a una gran variedad de presiones selectivas (como antibióticos, metales pesados, biocidas u otros compuestos) y tienen una amplia red de contactos en los distintos ambientes que promueven el intercambio y el flujo entre diferentes comunidades bacterianas. Además, recientemente hemos creado nuevas herramientas en nuestro grupo y con los consorcios en los que participamos que nos permiten un estudio profundo del pangenoma de *E. faecalis*.

Adicionalmente, los elementos extracromosómicos juegan un papel importante en la adaptación y evolución de poblaciones bacterianas de enterococos. Estos elementos abarcan una gran diversidad entre los que se incluyen plásmidos, elementos conjugativos e integrativos y transposones conjugativos. Los paradigmas de los sistemas conjugativos son el transposón

Tn*916* y las familias de plásmidos tanto Inc18 como los plásmidos inducibles por feromonas RepA_N fueron encontrados por primera vez en *E. faecalis*.

El uso extendido de algunos antibióticos, como es el caso de la tetraciclina, parece haber provocado un cuello de botella evolutivo en la estructura poblacional de algunos géneros de Firmicutes de interés clínico. La resistencia a tetraciclina en la mayoría de las bacterias Gram positivas patógenas de humanos es causada principalmente por la adquisición de elementos integrativos y conjugativos de la familia Tn*916* portadores del gen *tet(M).* Estos elementos Tn*916* incluyen una gran variedad de transposones con una sintenia común pero difiriendo en los mecanismos de integración y transferencia, en el sitio específico de inserción y en el rango de hospedador. Tn*5801*, perteneciente a la familia Tn*916*, fue considerado un elemento sitio-específico. Sin embargo, su detección en diferentes géneros de interés clínico durante mucho tiempo (staphylococci, streptococci, enterococci) sugiere eventos de transferencia horizontal.

El objetivo principal de este estudio es ampliar las perspectivas sobre la estructura poblacional de *E. faecalis* y sobre el papel del genoma accesorio y la transferencia horizontal de genes en la organización de esta estructura poblacional y en la evolución de la especie. Para ello, hemos analizado una colección de *E. faecalis* aislados de aves migratorias salvajes con el objetivo de incrementar el número de genomas de hospedadores no-humanos en las bases de datos. Además, hemos comparado el genoma "core" y el accesorio de la especie, identificando grupos de proteínas accesorias nicho-específicas. Con la finalidad de entender la influencia de la transferencia horizontal de genes en la evolución de diferentes linajes clonales (estructura genómica, coste de crecimiento y evolvabilidad) hemos tomado como modelo la evolución de Tn*5801*, analizando su variabilidad y evolución en Firmicutes a través del análisis de la base de datos de genomas completos. Finalmente, hemos caracterizado los eventos de trasferencia relacionados con este transposón en *E. faecalis* y su coste de crecimiento en la presencia de tetraciclina. La tesis está dividida en tres capítulos principales centrados en estos objetivos.

El capítulo I aporta nuevas perspectivas sobre la estructura poblacional de *E. faecalis* y sus adaptaciones a nichos específicos. También proporciona un análisis detallado del pangenoma de la especie (genoma "core" y accesorio) y destaca su relevancia en la evolución.

Hemos confirmado con este estudio la limitada especialización en *E. faecalis* utilizando los métodos de BAPS y genómica comparativa. No obstante, también demostramos el reducido

tamaño de su genoma "core" frente al gran tamaño del accesorio, lo cual apoya la hipótesis de la Reina Negra (the "Black Queen Hypothesis"), según grandes pan-genomas que contienen un arsenal de elementos genéticos móviles facilitan adaptaciones rápidas a cambios medioambientales por eventos de transferencia horizontal. En este escenario, se produce la adaptación nicho-específica entre clones de *E. faecalis* de diferentes especies de aves salvajes hospedadoras que origina complejos clonales restringidos a ciertos estilos de vida como en el caso del CC82 confinado en aves salvajes.

El capítulo II caracteriza de forma profunda la diversidad, evolución y movilización del transposón Tn*5801*, un miembro de la familia de los Tn*916* muy poco estudiado hasta el momento. La transferabilidad de Tn*5801* fue demostrada entre cepas de *E. faecalis* con distinta base epidemiológica y también en cepas que ya albergan otro miembro de la familia de Tn*5801*. La importancia de este resultado reside en que este transposón tiene un sitio de inserción con una secuencia específica en el cual se inserta su tirosina-recombinasa, que reconoce el extremo 3' de una guanosin monofosfato sintasa (GMP), el gen *guaA*, un gen que sirve como sitio de inserción a numerosas islas genómicas de Firmicutes.

El capítulo III consiste en el análisis del contexto genético de los transconjugantes obtenidos de la transferencia de Tn*5801*. El objetivo ha sido comprender los eventos de movilización que han tenido lugar y su influencia en la tasa de crecimiento en asislados de esta especie. Este estudio no prueba únicamente la inesperada transferabilidad e integración de más de una copia de estos transposones conjugativos, sino que además define un nuevo elemento móvil de 33kb que forma un tándem con la variante Tn*5801*.B15 del elemento. Este nuevo transposon recibe el nombre de ICE*Efs*Asa1. Este trabajo realza la diversidad de los eventos de transferencia horizontal de genes destacando la creación y acreción de tándems tras la recombinación sitio-específica. Esta estrategia podría permitir la selección de persistencia de algunos clones patógenos.

Introduction

Introduction

Taxonomy and ecology

Enterococcus faecalis belongs to the phylum Firmicutes, class Bacilli, order Lactobacillales and family Enterococcaceae. They are Gram-positive non-motile cocci, non-spore-forming facultative anaerobes which can grow in a wide range of temperatures (10-45ºC) with optimum growth temperature at 35ºC. Although it does not produce a catalase reaction in the presence of hydrogen peroxidase, it can produce a pseudocatalase weak reaction when grows on blood agar (1). *E. faecalis* ferments glucose without gas production and can metabolize a variety of energy sources including complex carbohydrates, glycerol, lactate, citrate, malate, amino acids such as arginine, and many alpha-keto acids. One of the most characteristics features of the species and also of the genus *Enterococcus*, is the ability to survive in the presence of a wide variety of compounds like antibiotics, heavy metals, ethanol, azide, detergents or biocides, which allow them colonizing and persisting in a wide variety of host and conditions (2, 3). They include the gastrointestinal (GI) and oral tract of humans and other mammals, reptiles and birds as well as food, water, plants, soil and other biotic and abiotic surfaces (2, 4–6).

The species *E. faecalis* is specially abundant in the intestines of cattle, pigs, dogs, horses and poultry (7, 8) and it has been frequently isolated from wild mammals, reptiles, birds, insects, and wild plants. It represents the species most commonly recovered from humans, more than *Enterococcus faecium,* although the ratio of recovery is still 10 to 1 for many series (9). In healthy humans *E. faecalis* clones are recovered from individuals of all ages, and may persist in the gut for longer periods of time in the gut than *E. faecium* clones (6, 9, 10).

This ability of *Enterococcus faecalis* to colonize and cause infection in a wide of different hosts is supported by a variety of genomic traits including bacteriocins, microbial surface proteins, cell-wall-anchored proteins, or forming- biofilms proteins (6). Bacteriocins are small peptides with the ability to inhibit the growth of specific bacteria and can be classified in two classes, namely class I or Lantibiotics and class II (commonly found in enterococal species). Bacteriocin producing strains have been isolated from a wide variety of environments but they appear to be more common in food samples, and in human and animal gastrointestinal tract (6, 11). Among the cell surface determinants stand out aggregation substance genes like *asa1*, which are a group

of *E. faecalis* proteins, normally encoded in pheromone responsive plasmids. The carriage of the enterococcal surface protein (Esp) which contribute to biofilm formation, was described in a pathogenicity island. Secreted proteins as haemolysin-cytolysin (Cyl) or gelatinase (GelE) have been traditionally considered as virulence factors in *E. faecalis* (12). Cyl is a toxin able to lyse blood cells from humans, cow and horses. The protease GelE, seems to be involved in the degradation of the host tissues and modulation of the host immune response as well as in the activation of autolysin leading to the release of DNA from the bacterial cell and also biofilm formation. Such virulence genes are frequently found among *E. faecalis* isolates from a wide variety of different niches like the hospital, animals, food and the environment (13–15). Recently, Bakshi et al suggested the founder effect of a pathogenicity island (PAI) containing different virulence factors including *asa1* and *cyl*, PAI which is associated with the pathogenicity potential of the species (16).

Clinical relevance and antibiotic resistance

Enterococci are important nosocomial pathogens and are one of the 21st century medical challenges due to the increasing prevalence of antibiotic resistance strains against all antibiotics used in healthcare and veterinary (6, 17). The Centers for Disease Control and Prevention (CDC) has determined that Vancomycin Resistant Enterococci (VRE) is a serious threat to human health (https://www.cdc.gov/drugresistance/threat-report-2013/). Among enterococci, *E. faecalis* is the most common species recovered from human infections (60%) and one of the most common causes of healthcare-associated infections (7,8%) (18).

The genus *Enterococcus* is intrinsically resistant to β-lactams, aminoglycosides, and trimethoprim-sulfamethoxazole. The species *E. faecalis* is specifically resistant to lincosamides and streptogramins. In contrast to this intrinsic resistance, Enterococci can acquire an extensive variety of resistance mechanisms by mutation or by acquisition of exogenous genes (6, 17).

Population structure

Different methods have been used to discriminate strains with a high epidemicity potential including PFGE (Pulsed Field Gel Electrophoresis), MLVA (Multilocus Variable number tandem repeats Analysis), MLEE (Multilocus Enzyme Electrophoresis), MLST (Multilocus Sequence Typing) and more recently, cgMLST and whole genome sequencing (WGS) (19, 20). Different

schemes of MLST analysed by eBURST and goeBURST constitutes the most used method to analyse the population structure of *E. faecalis* (21, 22).

The first studies of *E. faecalis* population structure using MLST identified some overrepresented clonal complexes (CCs) among isolates causing infections in humans and animals (14, 23). Clonal complexes CC2, CC9, CC16, CC21, CC87, have extensively been associated with strains causing hospital infections and antibiotic resistance although is also frequent to recover strains of these major *E. faecalis* lineages in disparate hosts (23–28). As example, CC21, CC16 and CC40 are widely isolated from food, farm animals or hospitalized and non-hospitalized environments (23, 25–27).

These observations support an epidemic population structure for *E. faecalis* (6). Analysis of the trees of individual MLST loci were incongruent for three of the six genes and revealed *E. faecalis* is a highly recombinogenic species and has a high recombination:mutation rate (6, 23). This makes difficult the identification of some high-risk clones as isolates of the same lineage can be identified as different sequence types (ST) (29).

Analysis of whole genome sequences has been used in epidemiological studies to overcome such limitations. The first phylogenomic analysis of *E. faecalis* that included 18 genomes from clinical, commensal and animal sources to show the low relationship between the topology of the phylogenetic core tree and the ecological origin of the isolates (30). It also demonstrated a limited phylogenetic diversity in comparison to that of *E. faecium*.

A recent genome-based study of 515 *E. faecalis* genomes mainly associated with clinical settings on UK, revealed three dominant lineages of hospital associated lineages, but these data shows results of specific clonal outbreaks indicating independent clonal spread (31). So, this study provides for a deep analysis of genome base data, but not could be applied for a general interpretation of *E. faecalis* population structure. It is important to highlight the differences between antibiotic resistance and pathogenicity island traits in the same *E. faecalis* lineage,.

Algorithms that use sequence data from the taxon in question serve to identify the appropriate level of sequence diversity for distinguishing groups with the dynamic properties of the species. To overcome the limitations of the current typing methods, the Bayesian Analysis of Population Structure (BAPS) is an algorithm blind to ecology which is able to establish ecotypes based on patterns of sequence clustering (32). Some authors have used BAPS to analyse the *E. faecalis* population structure based on MLST data of gut samples from hospitalized and non-hospitalized

individuals (33, 34). Using BAPS, the *E. faecalis* populations clustered in 5 BAPS groups. Any cluster showed significant association between hospitalized and particular *E. faecalis* BAPS groups, but the contrary occurs with BAPS 2, which showing positive association with non-hospitalized individuals (33). Related with this issue, the widespread and wide distribution of some STs like ST6, ST16, ST40 or ST55 able to colonize hospitalized and non-hospitalized humans and also animals (15, 25, 26), apparently precludes the possibility of recognise ecotypes associated with specific environments. This can be caused by the higher variety of ecological microniches occupied by the *E. faecalis* species.

Mobile genetic elements

Extrachromosomal elements play important roles in the adaptation and evolution of bacterial populations of enterococci. They include a high diversity of elements although plasmids, integrative conjugative elements (ICEs) including conjugative transposons (CTn) and different types or transposons have been extensively analysed. In fact, paradigms of conjugation systems, namely those of Tn*916,* and the plasmid families Inc18 and RepA_N pheromone responsive elements subgroups were firstly recovered in *E. faecalis* (35). A number of recent reviews have comprehensively analysed the diversity of mobile genetic elements in these species (35–38). In the following paragraphs we will summarize the most important aspects related to those of *E. faecalis*.

Plasmids

Plasmids are categorized on the basis of size, replication mechanisms, ability to be transferred, phenotypic and functional characteristics, and host range. Recently developed PCR typing system based on the diversity of replication initiator proteins (39) or relaxases as markers of the mobilization (MOB) systems (40). Such systems allowed defining the diversity of enterococcal plasmids from humans, animals, and environmental samples (35, 36). However, these enterococcal typing schemes accounted for approximately 70% of the total number of plasmids in genome databases and approximately 60% of the total number of plasmids visualized by S1-nuclease analyses, thus reflecting the existence of unidentified plasmids rep-types (36, 38).

Rep-typing of large collections of *E. faecalis* strains has shown a dominance of plasmids of certain families (RepA_N pheromone-responsive (rep8 and 9), pS86 (rep6)- and Inc18 (rep1 and

2) and pUSA02 (rep7)). However, other plasmid families can be sporadically detected among *E. faecalis* populations.

Four families of plasmids that replicate by a theta mechanism, three that comprise conjugative plasmids (RepA_N, Inc18, and pMG1) and one small with nonconjugative elements (Rep_3) have been described. Members of the RepA_N and Inc18 families are often enriched in insertion sequences (ISs) that facilitate co-integration and rearrangements between different plasmid families and species. Such recombination events seem to have facilitated the origin of the great mosaicism of enterococcal plasmids that often carry more than one replication initiation protein, lack some essential modules, and eventually carry more than one relaxase (36).

The predominant group into the RepA_N family is the pheromone-responsive plasmids which represent a paradigm of elements in the biology of mobile genetic elements. Pheromone responsive plasmids constituted a group of conjugative narrow host range plasmids mostly described in *E. faecalis*. These pheromone plasmids includes pAD1, pAM373 and pCF10 as the predominant examples (41, 42). Pheromone plasmids receive their name because encode responses to small peptide fragments known as pheromones, which allow the donor cells the expression of aggregation substance protein facilitates the conjugation of plasmids. They frequently carry genes coding for antibiotic resistance or virulence. Inc18 plasmids constitute a large group of plasmids with a broad host range (36). The prototypes for these plasmids are pAMβ1, encoding resistance to macrolides, and pIP501, conferring resistance to macrolides and chloramphenicol and heavy metals and, eventually, glycopeptides. Chimeras containing plasmids of different origins have been reported (36).

The RCR or rolling circle replication plasmids comprising a high copy number of small plasmids (10 to 60 copies per cell with less than 5kb). The production of single-stranded DNA and the mechanism of replication of these plasmids enhance their ability to recombine, by either homologous recombination or recombination with other RCRs and theta replicating plasmids. RCR plasmids are frequently integrated into chromosomes and often harbored one or two antibiotic resistance genes (conferring resistance to tetracycline, chloramphenicol, macrolides, and trimethoprim). Specific members are often isolated form clinical isolates (35, 36).

Transposable elements

E. faecalis poses a large variety of transposons implicated in genome rearrangements and which confer mobility to a variety of virulence, colonization and cell-to-cell communication

determinants. Conjugative transfer of chromosomal fragments (up to 857 kb) by integration of pheromone responsive conjugative plasmids seem to play an important role in the evolvability of some lineages (43). Transposons can be classified into three groups: composite transposons;, Tn*3*-family transposons, and conjugative transposons.

Composite transposons

The ability to be transfer of composite transposons is conferred by flanking identical insertion sequences (ISs), oriented as direct or as inverted repeats. The most commonly identified composite transposon are those of the Tn*5281* family encoding resistance to aminoglycosides (44).

Tn3-family

The Tn3 family of transposons lack the encoding of transfer functions, instead of that is transferred within or different replicons through a replicative mechanism promoted by a transposase (TnpA) and a resolvase (TnpR). The main transposon described here was Tn*917*, confering resistance to macrolides, lincosamides, and streptogramin (*ermB* gene) (45, 46).

Conjugative transposons

A wide diversity of conjugative transposons (CTns) have been described among enterococci since the original description of Tn*916* in the early 1980s. They display a common synteny, but differ in their integrase/excisionase sequences and the specificity for the insertion site (35). CTns encode resistance to tetracycline (Tn*916-like,* Tn*6000,* Tn*5801,* Tn*5397*), kanamycin and erythromycin (Tn*1545*), or glycopeptides (Tn*1549/*Tn*5382*).

A large number of Tn*916*/Tn*1545* family members (*tetM*) have been detected in different enterococcal species recovered from humans, pets, farm animals, wild boars, house flies from food settings, and foods (38, 47). Enterococcal CTns that lack antibiotic resistance genes are increasingly detected in sequenced genomes. As an example, a Tn*916*-like element, *efaB5*, a~49.5 kb enriched among CC2 isolates, and plays a role in niche adaptation (48, 49).

Different examples demonstrated the presence of different CTns and/or multiple copies of a given CTn. The most emblematic ones are the platform of three Tn*916*-like transposons (Tn*6084*, Tn*6085*a and Tn*6085*b) in the epidemic VanB *E. faecium* C68 clone disseminated in Cleveland, Ohio (USA) in the early 90s, and Tn*5382* (*vanB2, pbp5*), enterococcal isolates from humans and animals that contain different CTns (different combinations of Tn*916*, Tn*5397*, Tn*6000* and other ICEs) (50, 51).

CTns may contribute to genome plasticity by several mechanisms. Multiple copies of related transposable elements are substrates for homologous recombination within and between different DNA elements mediating rearrangements which often result in mosaic platforms that carry fragments of transposons, plasmids, RM systems, and self-splicing elements, as demonstrated for the Tn*6000* (50, 52, 53). Moreover, transposable elements may alter the expression of genes either through disruption of genes by insertion within the coding region or by integrations in the promoter region. They might also contribute to genome evolution by favoring large deletions (54). Mosaic platforms that contain Tn*916-like* transposons are Tn*5385* (Tn*5381*-Tn*5384*-Tn*552*), Tn*5382* (*pbp5*-CTn*1549)*, or Tn*6009* (Tn*916*-mer) (35).

Hypothesis and objectives

Hypothesis and objectives

Enterococcus faecalis colonizes the majority of hosts (mammals, birds, insects, reptiles) within a population and therefore can be considered as part of the physiological commensal microbiota of humans and many other animals. The available *E. faecalis* population structure reflects a cluster of a limited number of genotypes and a mismatch distribution dominated by mutation, with a slight increase of identical genotypes pairs due to local hospital transmission. The overrepresentation of *E. faecalis* isolates of hospitalized patients in gene databases precludes defining the population structure of those species colonizing different host organisms of poorly defined and disparate ecological patches.

Our **HYPOTHESIS** is that hallmarks of either a high rate of mutation and/or drift in the population structure of *E. faecalis* might indicate the evolutionary dynamics of a generalist organism, which experiences a high level of drift and gene flow between disparate host species. If such, we could consider that *E. faecalis* has an "endemic polyclonal structure", where endemicity is assured by a highly frequent vertical and horizontal inter-host migration resulting in a minimal adaptive stress in colonization of most hosts.

The **MAIN OBJECTIVE** of this dissertation is to get insights in the population structure of *E. faecalis* and in the role of the accessory genome and HGT in shaping the population structure and evolution of this species.

The more specific **OBJECTIVES (O)** are the following:

> **O1.** To analyse the ***population structure*** of *E. faecalis* from migratory birds in order to feed the databases with isolates recovered from non-human hosts and to compare core and accessory genome of the species to identify niche-specific accessory protein clusters. Wild birds often live under different selective pressures related to resistance to antimicrobials (antibiotics, heavy metals, biocides) and have a high rate of network of contacts. The whole population structure of the species will be revised by using different approaches (goeBURST, BAPS, WGS) to analysed MLST data and core genome respectively. Particular emphasis will be devoted to analyse the role of the accessory genome in the structure of the species using advanced bioinformatics tools, including an in-depth study of their plasmid variability using standard typing techniques recently developed.

O2. To determine the impact of ***horizontal gene transfer*** in the evolution of different lineages (genome structure, fitness cost and evolvability). We will take as example the evolution of Tn*5801*, a Tn*916*-like transposon able to insert in a site-specific chromosomal site and be transferred among different bacterial species. We will analyse the variability and evolution of Tn*5801* in Firmicutes by the analysis of the WGS database, with the aim of understand the impact of the site specific transposons as Tn*5801* in the diversification of bacterial high risk clonal complexes (HiRCC).

O3. To characterize **the transfer events involved in the dynamics** of the conjugative transposon Tn*5801* among *E. faecalis* and their fitness in the presence of tetracycline. We will investigate the ability of Tn*5801* to be mobilized and further analyse the genomes of unexpected transconjugants (size and copy number) of site specific Tn*5801* in different *E. faecalis* backgrounds in order to characterize the different transfer events produced.

Materials and methods

Materials and methods

A wide variety of methods of clinical microbiology, molecular biology, molecular epidemiology, genomics and proteomics were used in this thesis. This section summarizes the techniques learned and applied in the different parts of the thesis. Some points are comprehensively and specifically detailed in chapters I, II and III.

1. MATERIALS

1.1. Growth media

- Brain-Heart Infusion (BHI) (Conda Laboratories S.A., Madrid, Spain). Powder dissolved in water and autoclaved.
- BHI agar (Conda Laboratories S.A., Madrid, Spain). Powder dissolved in water and autoclaved.
- Muller Hinton broth and agar (MH) (Oxoid Lda., Basingstoke, Hampshire, GB). Powder dissolved in water and autoclaved.
- Blood Agar (Oxoid Lda., Basingstoke, Hampshire, GB)

1.2. Reagents

1.2.1. Antibiotics tested by disc diffusion or gradient

Antibiotics	Concentration range E-test (mg/L)[a]	Concentration of disc (µg)[b]
Ampicillin	0.125-256	-
Cefotaxim	0.125-256	30
Ceftazidime	0.06-128	30
Ciprofloxacin	0,002-32	5
Chloramphenicol	0,016-256	30
Erythromycin	0,016-256	-
Spectinomycin	-	10
Streptomycin	-	10
Gentamicin	0,016-256	10
Imipinem	0,002-32	-
Kanamycin	-	30
Tetracyclin	0,016-256	35
Trimethoprim	-	5

[a] (Oxoid Lda., Basingstoke, Hampshire, GB); [b] (bioMérieux, Marcy l'Etoile, Francia)

1.2.2. Solutions

- SDS: 20% (100ml) – Dissolve 20g in 70ml of ultra-pure water, complete the volume of 100ml and sterilize with filter.
- NaOH: 10N (100ml) – Dissolve 40g of NaOH (*Merck, GMaA, Germany*) in ultra-pure water and make up to 100ml.
- Tris-HCl: 1M, pH=7.6 (1L) – Dissolve 121.1g of Tris base (*Calbiochem*) in 800ml of ultra-pure water adjust the pH=7.6 with concentrated HCl (~60ml), complete the volume of 1L and sterilize.
- EDTA: 0.5M, pH=8 (*Merck*)
- NaCl: 5M (*Merck*)
- Brij58 10% (*Sigma*)
- Deoxycolate 10% (*Sigma*)
- Sarkosyl 20% (*Sigma*)
- Proteinase K solution (25mg/ml) (*Promega, Madison, WI, USA*)
- Ribonuclease (RNase) solution (10mg/ml) (*USB Corp., Cleveland, OH*)
- TE 1X (10 mM Tris-HCl pH=7.6, 1 mM EDTA pH=8.0)
- TBE 10X extended range buffer (Tris/Boric Acid/EDTA) (*Bio-Rad Laboratories GmbH, München*). For manual preparation (1L): 108g Tris base, 55 g boric acid and 40ml EDTA 0.5M pH=8.0, make up to 1L with ultra-pure water and autoclave.
- Ethanol (*Merck*)
- Isopropanol (*Merck*)
- Phenol-chloroform (1:1)
- Phenol-chloroform-isoamylic acid (25:24:1) (*Sigma*)
- NaOH: 0.4N (1L) – Dissolve 16g of NaOH (*Merck KGaA, Germany*) in 900ml of ultra-pure water and complete the volume of 1L. Do not sterilize.
- Tris-HCl: 0.5M pH=7 (1L) – Dissolve 60.55g of Tris base in 800ml of ultra-pure water, adjust the pH of 7.0 with HCl and complete the volume of 1L. To sterilize for stock.
- 20X SSC [3M NaCl, 0.3M $Na_3(C_3H_5O(COO)_3)$] (1L) – Dissolve 175.31g NaCl (*Merck KGaA, Germany*) and 88.23g $Na_3(C_3H_5O(COO)_3)$ (*Merck KGaA, Germany*) in 800ml of ultra-pure water and make up to 1L. Sterilize as a stocking solution and dilute 1:10 for 2X SSC.

- $NaH_2PO_4.H_2O$: 0.5M pH=7 (100ml) – Dissolve 6.9g (*Merck KGaA, Germany*) in 80ml of ultra-pure water, adjust the pH=7 with NaOH pellets, complete the volume of 100ml and sterilize for stock.
- $MgCl_2$: 1M (25ml) – Dissolve 5.08g of $MgCl_2$ (*Merck KGaA, Germany*) in ultra-pure water, complete the volume and sterilize for stock.
- Hybridization buffer [NaCl 0.5M + blocking reagent 4% (W/V)] (100ml) – 2.922g of NaCl and 4g of blocking reagent (supplied with the kit) are mixed with 100ml of hybridization buffer (supplied with the kit) and the solution is immediately placed in vigorous agitation for 2 hours at room temperature.
- Primary wash buffer (1L) – 120g of urea (*Merck KGaA, Germany*); 8.7g NaCl; 100ml of $NaH_2PO_4.H_2O$: 0.5M pH=7; 1ml $MgCl_2$: 1M; and 2g blocking reagent are dissolved in 700ml of ultra-pure water (in a boiling water bath and agitation); 1g SDS (Sodium *n*-Dodecyl Sulfate; *Calbiochem*®) is added and the volume of 1L is complete with water. This solution can be kept at 4ºC for one week.
- Secondary wash buffer (1L) – Dissolve 121g Tris base and 112g of NaCl in 800ml of ultra-pure water, adjust to the pH=10 and complete the volume of 1L. After sterilization, this stock solution can be stored at 4ºC for 4 months. The working solution must be prepared when needed by diluting the stock solution 20X in water and adding 2mL of $MgCl_2$ (for 1L).
- SDS: 0.1% (1L) – Dissolve 1g of SDS (*Calbiochem*®) in 1L of distilled water and carefully warm the solution until boiling for striping.

1.2.3. Restriction Enzymes

- ***Sma*I** – supplied with 10X NEBuffer 4; optimal activity at 25ºC *(New England Biolabs Inc.*)
- **I-*Ceu* I** - supplied with 10X NEBuffer 4 and 100X BSA; optimal activity at 37ºC *(New England Biolabs Inc.*)
- **S1 nuclease** – supplied with 10X S1 nuclease buffer; optimal activity at 37ºC *(Takara Bio Inc., Shiga, Japan*)

1.2.4. Molecular markers

- *100 bp DNA ladder* (*Promega, Madison, WI USA*)
- *λ DNA/Hind III* (*Promega, Madison, WI USA*)
- *λ-EcoT14 I/Bgl II digest* (*Takara Bio Inc., Shiga, Japan*)
- *λ-ladder PFG marker* and *Low Range PFG marker* (*New England Biolabs, Inc., Ipswich, USA*)

1.2.5. Others

- Deoxynucleoside triphosphates (dNTPs) stock solution – equimolar amounts of dATP, dTTP, dCTP, and dGTP (*PCR Nucleotide Mix, Promega, Madison, WI USA*)
- Oligonucleotide primers – acquired from *Sigma-Aldrich, UK*
- LA *Taq* polymerase for long PCRs (supplying dNTPs, reaction buffer and $MgCl_2$) (*Takara Bio Inc., Shiga, Japan*)
- Agarose D-1 Low EEO (Pronadisa, Madrid, España)
- Agarose D5 (Pronadisa, Madrid, España)
- Plug mold (*Bio-Rad Laboratories, Hercules, CA USA*)
- Lysozyme from chicken egg white (*Sigma-Aldrich, St Louis, MO USA*).

1.3. Receptor, control and reference strains

Strain	Applicability
Escherichia coli ATCC® 25922	Reference strain
Escherichia coli ATCC® 10536	Reference strain
Staphylococcus aureus ATCC® 29213	Reference strain
Enterococcus faecalis ATCC®29212	Reference strain
Staphylococcus aureus ATCC® 6538	Reference strain
Enterococcus hirae ATCC® 10541	Reference strain
Enterococcus faecium GE-1	Conjugation recipient strain
Enterococcus faecium BM4105RF	Conjugation recipient strain
Enterococcus faecium 64/3	Conjugation recipient strain
Enterococcus faecalis JH2-2	Conjugation recipient strain
Enterococcus faecalis OG1RF	Conjugation recipient strain
Enterococcus faecalis OG1SS	Conjugation recipient strain
Enterococcus faecium C68	hylEfm
Enterococcus faecalis RE25	Plasmid typing - pRE25 (rep, rel, TA genes)
Enterococcus faecalis HUMV 14.3	Plasmid typing - pEF1, pCIZ2 (rep, rel, bac genes)
Enterococcus faecium 399.F99.H8	Plasmid typing - pVEF1 (rep, TA genes)
Enterococcus faecium U37	Plasmid typing - pRUM (rep, rel, TA genes)
Enterococcus faecium pMG1	Plasmid typing - pMG1 (rep, rel genes)
Enterococcus faecalis DS16	Plasmid typing - pAD1 (rep, rel, TA genes)
Enterococcus faecalis S-86	Plasmid typing - pS86 (rep, rel genes)
Enterococcus faecium 9631160-1	Plasmid typing - pRI1-like (rep, rel genes)
Enterococcus faecium 64/3xUW2774	Plasmid typing - pLG1 (rep, rel genes)

Abbreviations: Rep, replication; Rel, relaxases; TA, toxin-antitoxin

1.4. Commercial kits

- Wizard® Genomic DNA Purification Kit (Promega)
- QIAamp® DNA Mini Kit (*Qiagen GmbH, Hilden, Germany*)
- Genomic DNA kit Qiagen GmbH, Hilden, Alemania
- Plasmid MiniPrep, Midi, Maxi Qiagen GmbH, Hilden, Alemania
- LaTaq Takara Bio Inc, Shiga, Japón
- PCR clean up kit Promega, Madison, WI
- GoTaq Flexo DNA Polymerase Promega, Madison, WI
- ExoSAP-IT® purification kit USB Corp., Cleveland, OH, EEUU
- Power SYBR Green Master Mix Applied Biosystems, Foster City, CA, EEUU
- Gene Images Alkphos Direct Labelling system GE Healthcare Life Sciences Amersham, Uppsala, Suecia
- CDP-star Detection Reagent GE Healthcare Life Sciences Amersham, Uppsala, Suecia. *Amersham GB, GE Healthcare Life Sciences UK Limited*

1.5. Equipment

- Termocyclers MJ Research PTC-100 (Global Medical Instrumentation Inc., Minnesota, EEUU)
- Qubit™ Fluorometer (Thermo Scientific, Waltham, MA, USA)
- Nanodrop 1000 (Thermo Scientific, Waltham, MA, USA).
- CHEF-DRIII – PFGE - (Bio-Rad, La Jolla, CA., EEUU)
- ChemiDoc™ XRS (Bio-rad).

1.6. Other materials

- Nylon transfer membrane – Hybond™ N+ (*GE Healthcare Life Sciences UK*)
- Blotting paper (Whatman 3MM) (*Whatman International Ltd., Maidstone England*)

1.7. Bacterial isolates

A collection of 610 enterococcal isolates collected in different countries along a 23 year period (1987-2010) from humans (n=320; hospitalized patients, HP, n=195; and healthy human volunteers, HV, n=125); farm animals (n=236; poultry, P, n=164; and swine/piggeries, SP, n=72),

wild birds (WB, n=97) and hospital/urban sewage (n=54, SW) (55–57) were analysed in this thesis Table 1.

Table 1. Bacterial isolates analysed

Species	Source type	Host/ origin	Number	Geographical origin
E. faecalis	Human	Patient	75	13 countries
		Healthy individuals	61	Portugal
	Sewage	-	17	Portugal
	Animals	Poultry	60	Portugal
		Swine	28	Portugal
		Wild birds	97	Spain
E. faecium	Human	Patient	122	23 countries
		Healthy individuals	51	Portugal
	Sewage	-	29	Portugal
	Animals	Poultry	46	Portugal
		Swine	66	Portugal
Enterococcus spp.	Human	Healthy individuals	13	Portugal
	Sewage	-	8	Portugal
	Animals	Poultry	57	Portugal
		Swine	26	Portugal

Enterococcal isolates from retail poultry samples were obtained from swabs or meat samples from the inner carcass of retail poultry (chicken and turkey lots) from two different butcher shops and one market in the Porto area. Isolates from sewage water of four Porto hospitals and urban sewage waters draining directly to the sea.

2. METHODS

2.1. Identification of *Enterococcus* species

The strains were identified by MALDI-TOF MS (Matrix assisted laser desorption ionization-time of flight mass spectrometry), a tool for rapid identification and typing of bacterial species. MALDI-TOF is an ionization technique that uses a laser energy absorbing matrix to create ions from large molecules with minimal fragmentation.

MALDI methodology is a three-step process (Figure 1). First, the sample is mixed with a suitable matrix material and applied to a metal plate. Second, a pulsed laser irradiates the sample, triggering ablation and desorption of the sample and matrix material. Finally, the analyte molecules are ionized by being protonated or deprotonated in the hot plume of ablated gases, and can then be accelerated into whichever mass spectrometer is used to analyse them. We adopted a protocol suggested by the manufacturer for clinical microbiology diagnostics (extraction procedure, matrix, sample:matrix ratio and solvents) in order to grant maximal compatibility with established workflows and easier adaptation of current protocols. Bacterial extracts were obtained according to the instructions of Bruker, the manufacturer. Briefly, overnight cultures (18 h/37 ºC) in BHI agar were suspended in HPLC water and mixed with ethanol, and the pellets were resuspended in equal parts of formic acid (70 %) and acetonitrile. One microlitre of crude extracts was spotted onto a polished target (AnchorChip™), air dried, overlaid with 1 µl of α-cyano-4-hydroxycinnamic acid (HCCA) matrix solution (50 % acetonitrile/2.5 % TFA v/v).

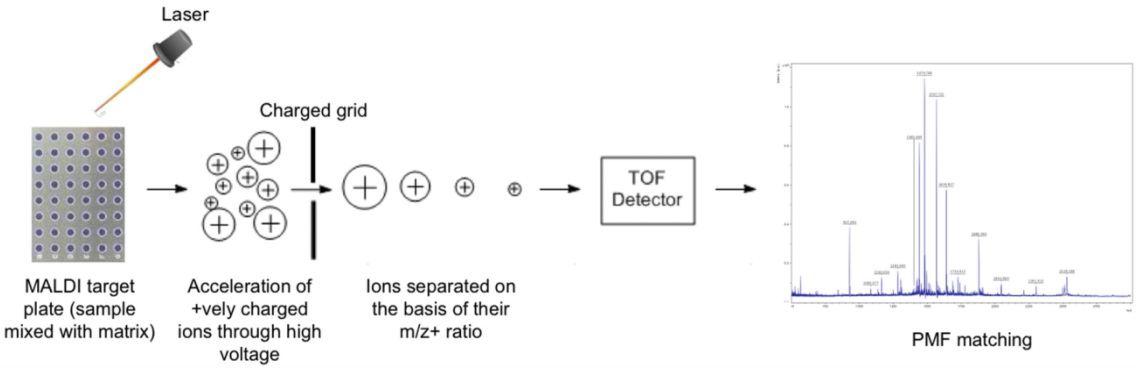


Figure 1. Schematic diagram showing the work-flow in a MALDI-TOF MS. Taken from Singhal et al, Frontiers in Microbiology 2015 (58)

2.2. Antimicrobial susceptibility tests

The antimicrobial susceptibility of bacteria to different antibiotics was performed by using disk diffusion methods (Kirby-Bauer and Epsilon test E-test) and agar dilution susceptibility test and/or the Epsilon-test (E-test).

The inoculum used in all the antibiotic susceptibility techniques was prepared from a colony bacterial suspension in 0.9% sterile saline with a 0.5 McFarland turbidity (1.5 x 10^8 CFU/ml). This suspension was diluted 1:100 in Mueller Hinton broth to obtain an inoculum of 1.5 x 10^7 CFU/ml. The inocula were used within the 15m after the preparation. Agar plates were incubated at 37ºC during 16-18h accordingly to the CLSI guidelines (10)

2.2.1. The agar dilution method

The agar dilution method is the antimicrobial susceptibility testing method that provide more accurate result (3). The lowest concentration of an antimicrobial agent that completely inhibits the visible growth of a microorganism is known as the MIC. Agar dilutions were prepared in 90 mm diameter Petri dishes which allowed testing several organisms on each plate.

2.2.2. The disk diffusion test

The Kirby-Bauer agar disk diffusion method (5) involves the use of a dried filter paper disc impregnated with a specific concentration of antibiotic in the agar used to culture the microorganisms which produce a halo of inhibition around the disc. By measuring the diameter of the halo, we can classify the strain in susceptible (S), intermediate (I) or resistant (R) following the CLSI standard guidelines (10). The rationale of these interpretative charts assumes that the relationship between the diameter of the zone and the MICs is known and that the outer limit of the zone of inhibition contains an antibiotic concentration which is similar to the MIC of that antibiotic to a specific organism.

2.2.3. The Epsilon-test (E-test)

The Epsilon Test is a quantitative method combining the principle of the agar diffusion test with the determination of the MIC value. The E-test utilises a thin inert strip impregnated with a specific antimicrobial agent present on a variable range of concentrations, combining the principle of the agar diffusion test with the determination of the MIC value. A bacteria strain is inoculated onto a Mueller Hinton agar plate and the E-test strip is laid on top. The intersection of

the inhibitory zone edge and the strip indicates the MIC value over a wide concentration range with inherent precision and accuracy.

2.3. DNA extraction

The extraction of the genomic or plasmid DNA was performed using either standard conventional techniques or commercial kits. All are based on the disruption of the cell, the removal of cell debris and proteins, and the selective precipitation of double-stranded DNA of high molecular weight and purity suitable for several applications on molecular biology (26). The conventional methods used as lytic agents heat (boiling), alkaline solutions and/or detergents while the commercial kits combine columns

2.3.1. Genomic DNA

Total DNA extraction was achieved by the classic *boiling* method or by using specific commercial kits like Wizard® Genomic DNA Purification Kit (Promega) and QIAamp® DNA Mini Kit (DNA Purification from Tissues).

For the extraction of genomic DNA *in situ* the lysis of bacteria is carried using cells embedded in agarose (25,25,37). The method requires the preparation of unsheared DNA by the incorporation of organisms in agarose, the lysis and disruption of the cells and the washing of cellular debris. Briefly, a single colony enterococci is grown overnight in 5 ml of brain heart infusion broth at 37ºC and the cells are harvested and suspended in 1.25 ml of buffer. A portion (0.4 ml) of this suspension is mixed with 0.4 ml of 2% agarose D-1 Low EEO (Pronadisa, Madrid, Spain) in water at 50 to 60°C and then pipetted into a plug mold allowed to solidify at 4ºC. For lysis, one to five plugs are placed in 4 ml of fresh lysis solution (6 mM Tris-HCl pH=7.6; 1 M NaCI; 100 mM EDTA pH=7.5; 0.5% Brij 58; 0.2% deoxycholate; 0.5% sodium lauroyl sarcosine; 20 ug/ml of RNase; 1 mg/ml of lysozyme). Following incubation overnight at 37°C with gentle shaking, this solution is replaced with 5 ml of ESP solution and then incubated overnight at 50°C with shaking. The plugs are finally washed three times for 30 min each with 5 ml of TE 1X and then stored in this buffer at 4°C.

2.3.2. Plasmid DNA

The overall plasmid content of isolates was determined using a modification of the Barton method (4). The procedure included treatment of DNA embedded in agarose plugs with 14U of

S1 nuclease at 37ºC during 15min to convert supercoiled plasmids into full-length linear molecules and PFGE with *λ-ladder PFG* (48.5kb to 1,018.5kb) and *Low Range PFG* (0.13kb to 194kb) linear DNA markers to estimate plasmid sizes. Linearized, single-copy plasmids were detected after electrophoresis as bands in a faint genomic background.

2.4. DNA amplification

In this thesis, Polymerase Chain Reaction (PCR) the standard molecular biology laboratory technique (28,36,41) was used for the detection and identification of different sequences both by simplex and multiplex schemes. Furthermore PCR mapping and long PCR were use in chapter II to determine the backbone structure of Tn*5801* transposons.

All PCR reactions were performed in termocyclers MJ Research PTC-100 (Global Medical Instrumentation Inc., Minnesota, EEUU) with appropriate controls (positive, negative and blank) and reagents [DNA polymerase, oligonucleotide primers, deoxynucleoside triphosphates (dNTPs), divalent cations ($MgCl_2$), reaction buffer (Tris-Cl pH=8-9)]. And the amplification conditions were adjusted to each experiment depending on the size of the expected amplicon. Detailed protocols were mentioned in chapter I and II.

2.5. DNA separation and detection

The separation of PCR products and plasmids or DNA extractions was achieved by horizontal agarose gel electrophoresis in an electric continuous-field. The agarose gels were prepared in different concentrations (0.8%-2%) accordingly to the amplicon size, and DNA was separated in TBE 0.5X buffer (45 mM Tris-Borate, 1mM EDTA) with different running conditions depending on the amplicon. The tracking dye loading buffer is mixed with the samples before loading onto the gel to increase the density of the sample, to simplify the loading process by adding colour to the sample, and to follow the progress of the electrophoresis based on the movement of the dye.

2.5.1. Pulsed-Field Gel Electrophoresis (PFGE)

The resolution of large DNA molecules was achieved by a method based on the principle of an electrophoretic current "pulsed" in different directions over a gradient of time intervals (25,43,44). The method includes the genomic DNA extraction in agarose molds (1.5.1.1), the DNA digestion with appropriate restriction enzymes and the separation of a range of DNA

molecules, including those of megabase size. The pulse-switching increments and the total electrophoresis run time are influenced by the range of DNA fragment sizes that must be separated as a result of specific chromosome-restriction enzyme interaction (17,17,41).

The gel was prepared with the same agarose used for preparing the DNA molds (agarose D5; Pronadisa, Madrid, Spain*)* which high-quality and characteristics allow an optimal separation of heavy DNA fragments as well as a great electrophoretic mobility adequate to further DNA transfer into membranes. Electrophoresis of digested genomic DNA was carried out on a CHEF-DRIII (Bio-Rad, La Jolla, CA., EEUU) device using the running TBE 0.5X buffer maintained at 14ºC with standardized conditions (6V/cm, 120º). In CHEF-DRIII gels, the electric field is generated from multiple electrodes that are arranged in a square or hexagonal contour around the gel and are clamped to predetermine potentials (8). We used different running settings depending on each restriction enzymatic reaction and the expected fragments size: *Sma*I [1-20s, 38h (*E. faecalis*) or 1-20s, 26h (*E. faecium*)], I-*Ceu*I (5-30s, 22h) and S1 nuclease (5-25s for 6 h followed by 30-45s for 18 h). After electrophoresis through agarose gels, DNA was staining with the GelRed™ dye and visualized by illumination with 300-nm UV light.

2.6. Clonal Relationship

The identification and discrimination of clones or clonal populations consists in the clusterization of isolates sharing identical or very similar genotypes, descendent from a single common ancestor and not influenced by HGT (horizontal gene transfer). We establish this population structure by two techniques: PFGE and/or Multilocus Sequence Typing (MLST).

2.6.1. Pulsed-Field Gel Electrophoresis (PFGE)

Similarity of whole digested DNA profiles obtained by PFGE has extensively used in the clinical setting for many years for rapidly detect hospital outbreaks

In this thesis, PFGE was used to discriminate among several enterococcal isolates originating from the same or different ecological niches. Clonal relationships were established by comparison of DNA restriction patterns (digested with *Sma*I) following the Tenover *et al.* method (49). By this method, the isolates can be identified as indistinguishable (if the restriction patterns have the same bands – isolates are considered the same strain), closely related

(restriction patterns differ in 2 or 3 bands), possibly related (restriction patterns differ in 4-6 bands) and unrelated (differ in ≥ 7 bands).

2.6.2. Multilocus Sequence Typing (MLST)

MLST is a nucleotide sequence-based typing scheme where multiple genes (loci) are sequenced to measure genetic relatedness and analyse sequence variation between alleles from many strains (32,51). *E. faecalis* MLST schemes include the identification of alleles from DNA sequences of seven housekeeping genes internal fragments. The choice of these housekeeping genes is based on their putative function, on their location within the chromosome, on their use in MLST schemes for other bacterial species and on the availability of sequence data from each enterococci species. For *E. faecalis* MLST typing, internal fragments of the following seven genes were amplified by PCR: *gdh* (glucose-6-phosphate dehydrogenase), *gyd* (glyceraldehyde-3-phosphate dehydrogenase), *pstS* (phosphate ATP-binding cassette transporter), *gki* (putative glucokinase), *aroE* (shikimate 5-dehydrogenase), *xpt* (xanthine phosphoribosyltransferase), and *yiqL* (acetyl-coenzyme A acetyltransferase).

The primers and PCR conditions used are detailed in the MLST database (http://efaecium.mlst.net and http://efaecalis.mlst.net). Each *loci* sequence was compared with that available in the database. Different sequences were assigned allele numbers, and different allelic profiles were assigned sequence types (STs) by interrogating the MLST database (http://www.mlst.net). Single-locus variants (SLVs) differ in only one of the seven housekeeping genes and double-locus variants (DLVs) differ in two of the seven housekeeping genes. Clusters of related STs differing in not more than in two of the seven loci that were thought to be descendants from a common ancestor were grouped into clonal complexes (CCs) by using eBURST (16). The website (http://www.mlst.net) contains all essential data to successfully perform MLST typing of *Enterococcus* and is hosted at Imperial College with funding from the Wellcome Trust.

2.7. Hybridization

Southern blotting involves the transfer of DNA from an electrophoresis gel to a membrane and its subsequent hybridization with a DNA probe (59). In this thesis, the identification of sequences or genes in bacterial genomes and/or mobile genetic elements (plasmids,

transposons) was performed by hybridization of DNA with specific probes generated by PCR. DNA transfer and hybridization were performed by standard procedures.

This protocol includes different steps: DNA separation; DNA transfer and fixation; DNA labelling and hybridization; post-hybridization washes; signal generation and detection. Chemiluminescence detection was carried out using a dioxetane substrate (*CDP-Star reagent*) which is catalyzed by the probe-bound alkaline phosphatase generating a rapid light output. The signal was detected by ChemiDoc™ XRS (Bio-rad). Membrane stripping: the DNA immobilized on a nylon membrane can be hybridized with different DNA probes several times without significant loss of signal strength. Between each assay, the DNA probes must be eliminated by washing the nylon membrane with a denaturing and boiling solution (SDS at 1%) during 1-2h followed by a brief wash with 2X SSC solution.

2.8. Location of resistance/virulence/plasmids genes

I-*Ceu*I digestion followed by pulsed-field gel electrophoresis is a powerful tool for determining genome structure as it cleaves the circular chromosomes at single sites in the rRNA operons thereby generating DNA fragments (6 in *E. faecium* and 4 in *E. faecalis*) suitable for the rapid mapping by PFGE (31). This method easily distinguishes between plasmid and chromosomal localizations, as I-*Ceu*I only cuts chromosomal DNA. Nevertheless, S1-PFGE has been considered the gold method to identify large plasmids and we assessed the genomic location of different genes by both I-*Ceu*I- and S1-digested genomic DNA PFGE with further hybridization. Chromosome and/or plasmid location of antibiotic resistance and virulence genes was assessed under high stringency conditions with the genes to be tested and 23S rRNA intragenic probes (13). Genomic DNA from different *Enterococcus* strains was used as positive controls for hybridization.

2.9. Conjugation assays

The most established method of DNA genetic exchange in enterococci is by conjugation. Conjugation experiments to study the *in vitro* transfer of genetic resistant elements were performed by the *filter mating* method (19) at 1:1 donor-recipient ratio using different *E.faecium* (BM4105RF, 64/3), *E. faecalis* (JH2-2, OG1RF, OG1SS) strains as recipients (all are plasmid-free).

Briefly, 500 µl of exponentially grown donor and recipient cells are placed onto a nitrocellulose filter and incubated overnight at 37°C; the filter is harvested and dilutions in physiological saline are streaked onto single selective agar plates. Transconjugants were selected on Brain Heart Infusion agar plates containing the selection antibiotic and incubated at 37ºC for 24h-48h. The transconjugants were confirmed by susceptibility tests and *Sma*I-PFGE electrophoresis analysis by comparison to recipient and donor strains.

2.10. Plasmid classification

The classification of *E. faecalis* plasmids was based on the presence of modules for the basic plasmid functions as replication (replicases), mobilization (relaxases) and stability (toxin-antitoxin and bacteriocins) using different PCR typing schemes that allowed the identification of the different plasmid families identified in enterococci and other Gram-positive genera: i) Inc18; ii) pMG1/pHTβ-derivatives; iii) Rolling-circle replicating plasmids; iv) low size theta replicating plasmids; v) pheromone-responsive plasmids (pAD1-like); v) mosaic plasmids (pEF1, pRUM, pVEF1/2/3); and vi) other conjugative plasmids from *S. aureus , Bacillus* and *Lactococcus lactis* (39). We used the different screening schemes for classifying the plasmid content: **Rep initiator proteins** developed by Jensen *et al.* (21). **Relaxases** multiplex-PCR-based relaxase typing method described by Francia *et al*. (40), **Toxin-Antitoxin systems (TA)** (23,35); (60); and **Virulence genes** (12)

2.11. Bioinformatics analysis

This thesis contains an important part of bioinformatics and computational work, from different analyses with raw whole genome sequence archives to handling large databases. For the different bioinformatics analysis we use a wide variety of tools and wrote scripts based on python programming language and in R-environment.

2.11.1. Whole genome sequencing and genome comparison

Sequencing was carried out using a standard 2 x 100 base protocol in a Genome Analyzer IIx Illumina HiSeq 2500 platform (Illumina, San Diego, CA). The “de novo” assembly of the paired-end reads was performed by SPAdes genome assembler (v.3.5.0)(76). For evaluating the quality of the genome assembly we use QUAST tool (77).

The genome comparisons were achieved by obtaining all the *E. faecalis* nucleotide and protein sequences from the NCBI database and included in the different analysis. SNIPPY tool was used to obtain the core genome and the whole genome alignment, which is a rapid variant calling and core SNP phylogeny tool that allows comparing all the genome sequences. *E. faecalis* V583 genome was used in all comparisons as reference. Multiple alignments of the concatenated SNPs extracted from each core genome were used to generate a maximum likelihood tree with FasTree v2.1.8 (78). The tree was plotted using the ggtree package of the R-environment (http://www.r-project.org).

All the sequences were annotated with Prokka v1.12 tool (Seeman T et al, 2014), and Easyfig v2.2.2 (Sullivan MJ et al, 2011) was used for drawing the schemes and comparisons. We classified some gene functions according the database of Clusters of Orthologous Groups of proteins (COGs) (http://www.ncbi.nlm.nih.gov/COG/).

The plasmid content of the sequenced genomes was inferred by two different tools: PLACNET (72) and plasmidSPAdes (v.3.5.0) (73). Recently, Orlek *et al*. widely discussed the different methods available for resolve plasmid structures from ambiguous assembly.

plasmidSPAdes (Antipov et al., 2016) works by calculating the median coverage of longer contigs to estimate chromosomal coverage; this estimate is used as a basis for filtering putative plasmid contigs from the assembly graph, but this approach assumes that chromosomal contig coverage differs from plasmid contig coverage and this does not happen under all conditions. In contrast, PLACNET uses reference sequences to find homology of contigs with known plasmids; an assembly graph can be visualized to allow manual pruning and correction. In the case of *E faecalis,* we face two main problems: the large amount of ISs present in their genome and the mosaicism of its plasmids. They preclude getting a good assembly, problem that is enlarged due to the low number of plasmid sequences deposited in the databases. To solve this analysis, we decide to combine these two different tools.

To obtain and analyze the accessory genome, we used ACCNET (61), a tool that allows inferring the accessory genome from genome sequences belonging to a population and further clustering the proteins by similarity. Using this tool, a comparison matrix was constructed to obtain a dendrogram of the accessory genome of all *E. faecalis* used.

Chapter I

Enterococcus faecalis in wild and migratory birds: new insights in the population structure and ecology of the species

The work related with this chapter was published in:

León-Sampedro R; Martins E; del Campo R; Rodriguez-Baños M; Tedim AP; Pozuelo MJ; Peixe L; Freitas,A; Novais C; Lanza VF., Baquero F; Coque TM. *Enterococcus Faecalis* in Spanish wild and migratory birds: from population structure to accessory genome. 7th FEMS Microbiology Congress 2017, Valencia (Spain). Abstract poster FEMS7-2555.

(manuscript submitted)

Chapter I: *Enterococcus faecalis* in wild and migratory birds: new insights in the population structure and ecology of the species

Abstract

Enterococcus faecalis is a commensal of mammals, reptiles, insects and birds, which might comprise reservoirs and vehicles of antimicrobial resistance genes of relevance in medicine and veterinary. Population structure reflects an epidemic structure and a high recombination/mutation ratio and suggests this species is a generalist organism under a high level of drift and gene flow between disparate host species. The lack of sequences of non-human hosts and the lack of suitable tools to analyse accessory genome, preclude the confirmation of theoretical hypothesis. The population structure of *E. faecalis* was revised after analyzing isolates recovered from migratory birds, a host barely studied to date and underrepresented in public databases. The analysis include both the core and the accessory genome using recently developed bioinformatics tools

E. faecalis isolates from 97 wild birds attended at two Spanish Veterinarian Centers of Wild Nature Conservation were analyzed. Clonal relationship was established by SmaI-PFGE and MLST analyzed by goeBURST and Bayesian Analysis of Population Structure (BAPS), Antibiotic and biocide susceptibility (microdilution in broth) was determined using standard methods (CLSI). The presence of genes encoding resistance to virulence genes, conjugative transposons (CTns) and plasmids was inferred by PCR typing schemes, hybridization and sequencing. Plasmid analysis included identification of replication initiation proteins (Rep), relaxases (Rel), and toxin-antitoxin systems (TA). Five isolates of disparate hosts but exhibiting related PFGE profiles were sequenced; different tools being used for the analysis of core and accessory genome (ACCNET) and plasmids (PLACNET).

E. faecalis clonal analysis showed 8 major PFGE clusters, diverse STs, mostly singletons not previously reported and classifying in 4 group BAPS (BAPS 1- 33,93%; BAPS 2- 46,43%; BAPS 3- 8,93% and BAPS 5- 10,71%). Isolates were resistant to tetracycline (67%), chloramphenicol (42%) erythromycin (28%) and high levels of streptomycin (26%), kanamycin (19%) and gentamicin (5%). The number of plasmids per cell ranges from 0 to 4 (3-290kb). They were

classified as RepA_N pheromone plasmids (rep9), with scarce presence of other families such as Inc18, Rep_trans-RCR and Rep3_small theta. Tn*916*-like elements (Tn916, Tn*5801*, Tn*5397*, Tn*6000*) and virulence genes were detected. A comprehensive comparative analysis of the genomes described here with those available in public databases (core genome phylogeny, accessory genome including plasmids) is given.

E. faecalis from wild birds exhibit variable susceptibility to biocides and antibiotics mediated by widespread transposable elements [*tet*(M), *erm(B)*] and in few cases by Inc18 plasmids, but with low number genes coding for heavy metal tolerance. These isolates carry a smaller proportion of plasmids than other hosts, being less likely to transfer resistance genes. It is important to highlight genomic content and evolvability of these strains of *E. faecalis* compared to other hosts and environments.

Introduction

Enterococcus faecalis is a member of the gastrointestinal flora of mammals, reptiles, insects and birds able to cause mild and severe infections in all these hosts (10, 61, 62).

The available *E. faecalis* population structure based on MLST data analyzed by *goeBURST* (23, 63) shows a cluster of a limited number of genotypes and a mismatch distribution dominated by mutation, with a slight increase of identical genotypes pairs due to local transmission in hospitals (CC2, CC9, CC87 or ST6, ST40), or farms (ST16) (63). Some major complexes are recovered from different geographical locations (64) or different hosts (15) but comparative genomics analysis suggest to discard transmission (64). Recent attempts to establish ecotypes by using a Bayesian analysis of population structure (BAPS) yielded incongruent results (28). Afterwards, Numminen et al suggested that hallmarks of either a high rate of mutation and/or drift in the population structure of *E. faecalis* indicated the evolutionary dynamics of a generalist organism, which experiences a high level of drift and gene flow between disparate host species (65). If so, we could consider that *E. faecalis* has an "endemic polyclonal structure", where endemicity is assured by a highly frequent inter-host migration (both vertical and horizontal), resulting in a minimal adaptive stress in colonization of most hosts (65). The hypothesis is tempting but the low numbers of isolates from different sources precludes sound conclusions.

Wild animals are supposed to live in areas under different selective pressures related to resistance to antimicrobials (antibiotics, heavy metals, biocides) and have a remarkable network of contacts. Although some reports suggest transmission of traits between different hosts, the number of genomes in databases is still anecdotic to support such hypothesis.

In this study, we analyzed 97 *E. faecalis* from Spanish migratory birds and revised the population structure of the species considering both the core and the accessory genome using recently developed bioinformatics tools (66). The whole genome sequencing of 5 apparently related isolates recovered from different wild bird species was analyzed in detail.

Materials and methods

Samples and bacterial strains

Fecal samples from wild birds were collected between the years 2001 and 2010 in hospitals of two Spanish veterinarian centers of nature conservation in the central and south west area of the Iberian Peninsula, the wildlife Hospital "Grupo para la Rehabilitación de la Fauna Autóctona" (GREFA) in Madrid (40°28′22″N 3°52′20″O), and in the wildlife recovery center Acción por el mundo Salvaje (AMUS) in Extremadura (38°33′41″N 6°20′21″O).

Animals sampled consist in live native wild birds, especially raptors. They entered in the centres after suffering collisions with vehicles or power poles, gunshot wounds, or also for being orphaned chicks. The bird species selection was focused in migratory birds and raptors, which represent the diversity of Iberian Peninsula birdlife (orders Falconiforme, Anseriformes, Ciconiiformes, Charadriiformes, Caprimulgiformes, Columbiformes, Passeriformes, Gruiformes, Strigiformes, Coraciformes), with different life and dietary habits (Table 1).

Individual swabs on Amies transport medium (Biomedics, Spain) were labeled and sent to the laboratory within a maximum of 24 hours of sampling and preserved at 4ºC until time of processing. Samples were screened for *E. faecalis* using selective media and further identified by MALDI-TOF. Antimicrobial susceptibility against 9 antibiotics (ampicillin, gentamicin, kanamycin, streptomycin, tetracycline, erythromycin, vancomycin, moxifloxacin and chloramphenicol) was performed by the agar dilution method accordingly to CLSI standard procedures (67). The presence of putative virulence traits enterococcal surface protein (Esp), hyaluronidase ($Hyl_{E.\ faecium}$), cytolysin/hemolysin (Cyl), gelatinase (GelE) and aggregation substance (Agg) were screened using PCR as described (12).

Population structure analysis

We analyzed plasmid diversity in 43 isolates representing the clonal diversity of the sample. Plasmid size and content was determined in all isolates using the method described by Barton et al (68). The presence of genes involved in basic plasmid functions as replication (rep-initiator proteins), mobilization (relaxases) and stability (toxin-antitoxin systems), was performed using PCR typing methods and further sequencing as previously described (69). Briefly, relaxases were sought by the PCR-based relaxase typing method previously described able to identify known plasmids and conjugative transposons of different MOB families (40); rep-initiator

proteins were screened using schemes by Jensen *et al* (39), further updated by Wardal *et al* to identify subgroups of *rep*9 pheromone plasmids (70); and toxin-antitoxin systems (TA) identified in streptococci and enterococci (Axe-Txe, ω-ε-ζ) were detected by simplex PCR (71).

The location of sequences related to pheromone plasmids was analyzed by Southern blot hybridization of S1-digested genomic DNA with rel/rep/TA probes obtained by PCR using DNA from reference plasmids.

The plasmid content of the 5 genomes sequenced was further inferred by combining the bioinformatics tools PLACNET (72) and plasmidSPAdes (v.3.5.0) (73). Then, the plasmid contigs obtained by both methods were annotated with Prokka v1.12 tool (74) and further compared with available sequences in the NCBI database using BLASTN and BLASTP analysis. Easyfig v2.2.2 (75) was used for drawing the plasmid backbones.

Conjugative transposons

Conjugative transposons identified among Firmicutes such as Tn*916*, Tn*5397*, Tn*5398*, Tn*6000,* and Tn*5801* was screened by a multiplex PCR assay based in the amplification of specific Tn integrases and excisionases (57, 76).

Plasmid characterization

We analyzed plasmid diversity in 43 isolates representing the clonal diversity of the sample. Plasmid size and content was determined in all isolates using the method described by Barton et al (68). The presence of genes involved in basic plasmid functions as replication (rep-initiator proteins), mobilization (relaxases) and stability (toxin-antitoxin systems), was performed using PCR typing methods and further sequencing as previously described (69). Briefly, relaxases were sought by the PCR-based relaxase typing method previously described able to identify known plasmids and conjugative transposons of different MOB families (40); rep-initiator proteins were screened using schemes by Jensen *et al* (39), further updated by Wardal *et al* to identify subgroups of *rep*9 pheromone plasmids (70); and toxin-antitoxin systems (TA) identified in streptococci and enterococci (*Axe-Txe*, *ω-ε-ζ*) were detected by simplex PCR (71).

The location of sequences related to pheromone plasmids was analyzed by Southern blot hybridization of *S1*-digested genomic DNA with rel/rep/TA probes obtained by PCR using DNA from reference plasmids.

The plasmid content of the 5 genomes sequenced was further inferred by combining the bioinformatics tools PLACNET (72) and plasmidSPAdes (v.3.5.0) (73). The plasmid contigs obtained by both methods were annotated with Prokka v1.12 tool (74). Comparison of contigs with known sequences in the NCBI database was performed using BLASTN and BLASTP analysis. Easyfig v2.2.2 (75) was used for drawing the plasmid backbones.

Comparative genomics

Total DNA was extracted from 5mL overnight cultures using Wizard® Genomic DNA Purification Kit (Promega) and DNA concentration was measured with Qubit™ Fluorometer and Nanodrop 1000 (Thermo Scientific, Waltham, MA, USA). Sequencing was carried out using a standard 2 x 100 base protocol in a Genome Analyzer IIx Illumina HiSeq 2500 platform (Illumina, San Diego, CA).

The paired-end reads were *de novo* assembled by SPAdes genome assembler (v.3.5.0)(77); and QUAST was used for evaluating the quality of the genome assembly (78).

All the *E. faecalis* nucleotide and protein sequences were obtained from the NCBI database and included in the analysis of the core and the accessory genomes. Core genome and the whole genome alignment were inferred using SNIPPY, a rapid variant calling and core SNP phylogeny tool, that allows comparing all the genome sequences, including the 5 genomes from wild birds sequenced in this study and using the genome of *E. faecalis* V583 as reference. Multiple alignments of the concatenated SNPs extracted from each core genome were used to generate a maximum likelihood tree with FastTree v2.1.8 (79). The tree was plotted using the *ggtree* package (80) of the R-environment (http://www.r-project.org). To obtain and analyze the accessory genome, we used ACCNET (66), a tool that allows inferring the accessory genome from genome sequences belonging to a population and further clustering the proteins by similarity. Using this tool, a comparison matrix was constructed to obtain a dendrogram of the accessory genome of all the *E. faecalis* analyzed.

Results

Epidemiological background

We recovered *E. faecalis* from 97 of the 100 samples analyzed (Table 1). The 68% of the isolates (66/97) were resistant to one or more antibiotics, with variable rates of resistance to tetracycline (67%), chloramphenicol (42%), erythromycin (28%) and high levels of streptomycin (26%), kanamycin (19%) or gentamicin (5%) (Table 2). All isolates were susceptible to glycopeptides (vancomycin and teicoplanin) and quinolones (moxifloxacin). Thirty-three of the isolates were susceptible to all antibiotic tested, and 24 were resistant only to tetracycline. The predominant resistance profile was kanamycin, streptomycin, tetracycline and erythromycin and appears in 9 (9,27%) isolates. Resistance profiles are shown in Table 2 and Figure 1.

We detected the presence of different putative virulence genes coding for gelatinase (*gelE*, 93%), aggregation substance (*asa1*, 76%), enterococcal surface protein (*esp*, 60%) and cytolysin-hemolysin (*cylA* 35%).

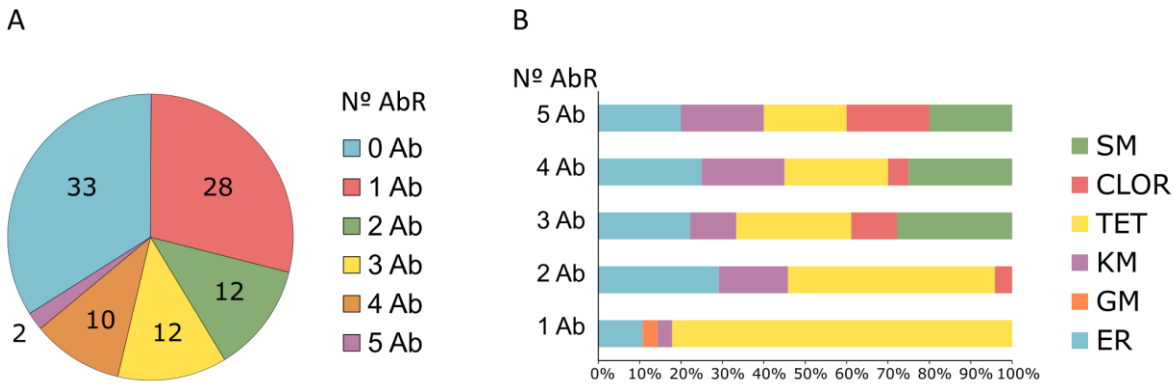


Figure 1. Antibiotic susceptibility of *E: faecalis* from wild birds. A. Pie chart Distribution of *E: faecalis* isolates from wild birds according the number of antibiotics against they are resistant showing the proportion of isolates susceptible to diferent antibiotics (0-5). B. Antibiotic profiles distributed by number of antibiotics with resistance.

Population structure of *Enterococcus faecalis*

From the analysis of the *SmaI* PFGE results, we can draw 8 major clusters (75% similarity among them) comprising isolates with diverse PFGE patterns. Within these groups, we identified 19 main subgroups of 2-5 strains each with a coefficient of similarity higher than 85-95% (Table 2).

Table 1. Classification of the wild bird samples by bird order and scientific names

Order	Name	Latin name	Number	Total
Falconiformes	Griffon Vulture	*Gyps fulvus*	6	**42**
	Cinereus Vulture, Black Vulture, Monk Vulture	*Aegypius monachus*	3	
	Red Kite	*Milvus milvus*	3	
	Black kite	*Milvus migrans*	3	
	Egyptian vulture or White Scavenger Vulture	*Neophron percnopterus*	1	
	Northern Goshawk	*Accipiter gentilis*	1	
	Western Marsh-harrier	*Circus aeruginosus*	2	
	Montagu's Harrier	*Circus pygargus*	1	
	Black-winged Kite	*Elanus caeruleus*	1	
	Sparrowhawk	*Accipiter nisus*	2	
	Booted Eagle	*Hieraaetus pennata*	1	
	Bonelli's Eagle	*Aquila fasciata*	1	
	Common Buzzard	*Buteo buteo*	1	
	Common Kestrel	*Falco tinnunculus*	9	
	Lesser Kestrel	*Falco naumanni*	7	
Anseriformes	Drakes	*Anas platyrhynchos*	9	**10**
	Drakes	*Anas platyrhynchos*	1	
Ciconiiformes	Cattle Egret	*Bubulcus ibis*	2	**19**
	White Stork	*Ciconia ciconia*	17	
Charadriiformes	Stone Curlew	*Burhinus oedicnemus*	6	**6**
	Lesser Black-backed Gull	*Larus fuscus*		
Caprimulgiforme	European Nightjar	*Caprimulgus europaeus*	1	**1**
Columbiformes	Rock Dove	*Columba livia*	2	**2**
Passeriformes	Red-billed Chough	*Pyrrhocorax pyrrhocorax*	1	**10**
	Red-billed Chough or Chough	*Pyrrhocorax pyrrhocorax*	1	
	Common raven	*Corvus corax*	3	
	House Sparrow	*Passer domesticus*	1	
	House Martin	*Delichon urbicum*	1	
	Barn Swallow	*Hirundo rustica*	1	
	Common Blackbird	*Turdus merula*	2	
Gruiformes	Great Bustard	*Otis tarda*	3	**3**
Strigiformes	Tawny Owl or Brown Owl	*Strix aluco*	1	**6**
	owl	*Different species*	1	
	Little owl	*Athene noctua*	4	
Coraciformes	Hoopoe	*Upupa epops*	1	**1**
	Unknown		3	**3**

Fifty-six isolates represented 40 different ST. Most STs (26/40=65%) were singletons (n=19; 4 ST192, 4 ST245, 2 ST314, 2 ST337, 3 ST340, 2 ST342, 2 ST344, 2 ST345, 2 ST346, 2 ST348, 2 ST349, 2 ST352, 2 ST353, 2 ST354, 2 ST356, 2 ST358, 2 ST360, 1 ST361, 1 ST362) or double locus variants (n=7; ST34, ST177, ST341, ST290, ST300, ST350, ST355), being 20 of them firstly

described here. Also, we identified previously described STs grouped into 8 clonal complexes, namely, CC81 (5 ST81, 1 ST357), CC82 [5 ST82, 1 ST170], CC4 (ST4, ST338, ST359), CC16 (ST16, ST363), CC40 (ST40), CC55 (ST55), CC58 (ST63), CC116 (ST275), CC19 (ST339). We represented a *eBURST* diagram with all the ST deposited in the database in September of 2016 (Figure 2).

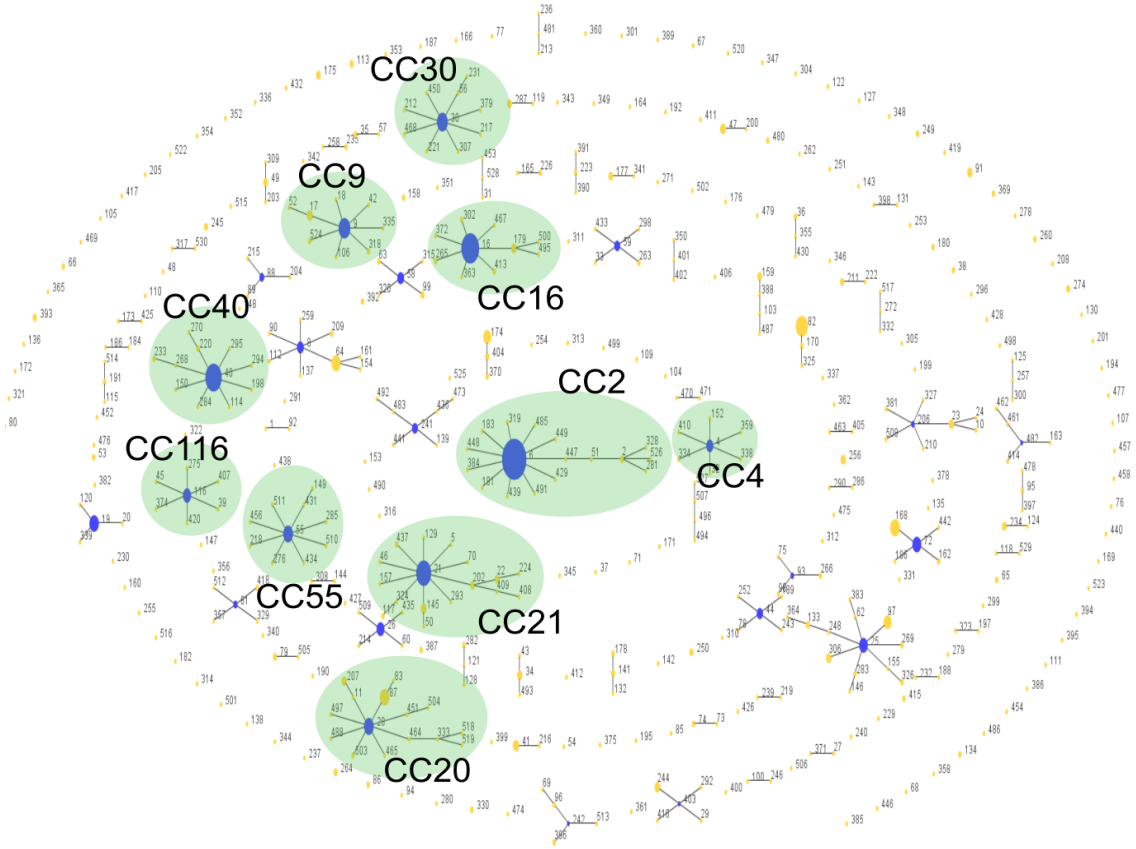


Figure 2. Population structure of *E. faecalis* by the analysis of all MLST of the database (September 2016). Green ovals represent major clonal complexes (CCs). Comparative electronic Based Upon Related Sequence Types (eBURST) analysis of the studied group of *E. faecalis* isolates from wild birds and the MLST database reference dataset (last update September 2016). Green circles represent major clonal complexes (CC); yellow circles, other STs (circle size is proportional to the number of isolates); lines, SLV links.

Table 2. Epidemiological background of *E. faecalis* from wild birds withMLST data, PFGE clusters and resistance profiles to antibiotics.

Cluster	ST	PFGE	Species	Order	Resistance profile
A	81	14	*Gyps fulvus*	Falconiformes	SM, TET, ER
	81	15	*Neophron percnopterus*	Falconiformes	ER
	81	48	*Anas platyrhynchos*	Anseriformes	TET, ER
	81	-	*Columba livia*	Columbiformes	TET
B	40	21	*Ciconia ciconia*	Ciconiiformes	KM, TET
	82	17	*Ciconia ciconia*	Ciconiiformes	TET
	82	45	*Turdus merula*	Passeriformes	KM, SM, TET, ER, CLOR
	354	22	*Anas platyrhynchos*	Anseriformes	SM, TET, CLOR
	363	24			
C	4	5.1	*Corvus corax*	Passeriformes	TET
	345	5	*Circus pygargus*	Falconiformes	KM, SM, TET
	349	5.1	*Milvus migrans*	Falconiformes	-
	55	3.2	*Ciconia ciconia*	Ciconiiformes	TET
	16	3.5	*Gyps fulvus*	Falconiformes	TET
	82	3.1	*Strix aluco*	Strigiformes	TET
	351	3	*Falco tinnunculus*	Falconiformes	TET
	170	3.2	*Ciconia ciconia*	Ciconiiformes	SM, TET, ER, CLOR
	82	3	*Falco tinnunculus*	Falconiformes	TET
	63	10.6	*Otis tarda*	Gruiformes	-
	290	10	*Anas platyrhynchos*	Anseriformes	TET
	275	11	*Pyrrhocorax pyrrhocorax*	Passeriformes	KM
	348	6	*Circus aeruginosus*	Falconiformes	TET, ER
	344	33	*Milvus milvus*	Accipitriforme	TET
	356	43	*Different species*	Strigiformes	-
D	352	13.5	*Anas platyrhynchos*	Anseriformes	TET
	352	13	*Anas platyrhynchos*	Anseriformes	TET
	359	13.2	*Hieraaetus pennata*	Accipitriforme	-
E	34	35	*Athene noctua*	Strigiformes	-
	81	48	*Burhinus oedicnemus*	Charadriiformes	-
	337	33	*Ciconia ciconia*	Ciconiiformes	-
	337	33	*Burhinus oedicnemus*	Charadriiformes	-
	340	17.4	*Milvus migrans*	Falconiformes	KM, SM, TET, ER, CLOR
	342	36	*Milvus milvus*	Accipitriforme	KM, SM, TET, ER
	340	47.3	*Columba livia*	Columbiformes	-
	245	16.3	*Accipiter nisus*	Falconiformes	KM, SM, TET, ER
	340	16.6	*Accipiter gentilis*	Falconiformes	-
	353	16	*Gyps fulvus*	Falconiformes	KM, SM, ER
	341	18	*Burhinus oedicnemus*	Charadriiformes	TET
F	177	18	*Anas platyrhynchos*	Anseriformes	GM
	245	16.3	*Accipiter nisus*	Falconiformes	KM, SM, TET, ER
	245	16.4	*Milvus milvus*	Falconiformes	TET
	245	16	*Ciconia ciconia*	Ciconiiformes	KM, SM, TET, ER
	358	25	*Ciconia ciconia*	Ciconiiformes	TET
G	300	26	*Aegypius monachus*	Falconiformes	KM, SM, TET, ER
	314	27	*Falco naumanni*	Falconiformes	-
	355	28	*Falco tinnunculus*	Falconiformes	SM, TET, ER, CLOR
OTHERS	360	17.5	*Otis tarda*	Gruiformes	-
	338	49.6	*Burhinus oedicnemus*	Charadriiformes	-
	339	31.4	*Ciconia ciconia*	Ciconiiformes	-
	346	50	*Falco tinnunculus*	Falconiformes	SM, TET, ER
	192	-	*Larus fuscus*	Charadriiformes	TET, ER
	350	-	*Ciconia ciconia*	Ciconiiformes	SM, TET, ER
	357	-	*Gyps fulvus*	Accipitriforme	TET, ER
	361	-	*Otis tarda*	Gruiformes	-

Abbreviations: CLOR, chloramphenicol; ER, erythromycin; GM, gentamicin; SM, streptomycin; TET, tetracycline, KM, kanamycin.

The isolates classified in four BAPS groups, namely BAPS 1 (19/56, 33.9%), BAPS 2 (26/56 isolates, 46.4%), BAPS 3 (8.9%) and BAPS 5 (10.7%). It is note the lack of association between BAPS groups and the hosts. While BAPS 1 are overrepresented by isolates from farm animal, BAPS 2 comprises isolates from disparate origins. Some STs belong to clonal complexes related to clinical, animal, and food isolates. Relations observed between STs are shown in the Phyloviz scheme (Figure 3).

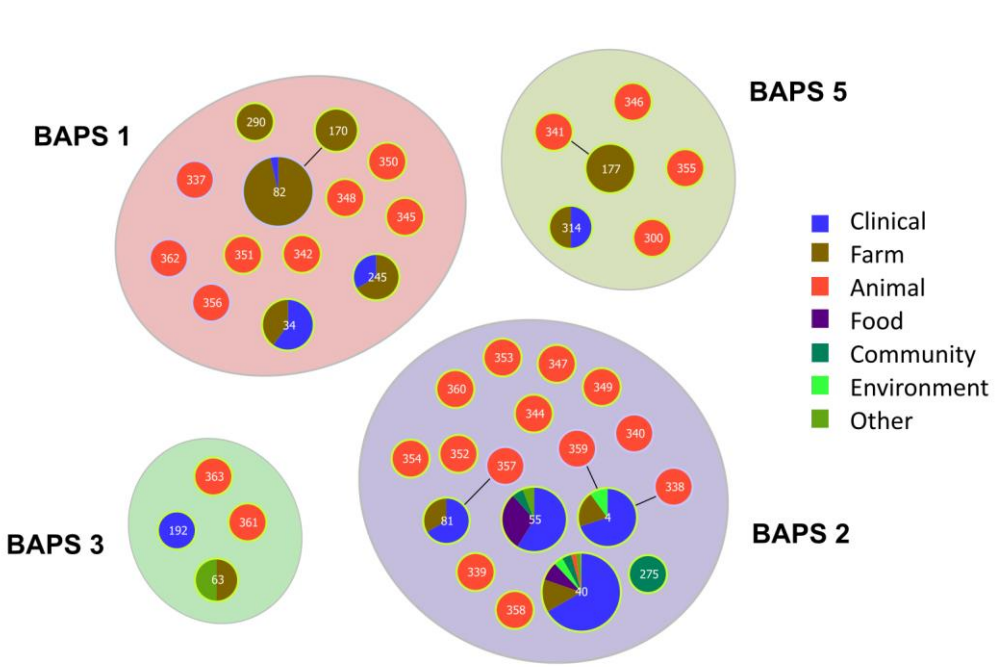


Figure 3. eBURST (PHYLOViZ) of the STs found in the collection of *E. faecalis* from wild birds. The size of the circles represents the abundance of each ST in the database. The pie-charts include the distribution by the source in the isolates deposited in the MLST database.

Mobile genetic elements (Figure 4)

The number of plasmids per strain ranged from 0 to 4 and the plasmid size ranging from 3 to 240kb. It is of note the apparent lack of plasmids in 35% of the isolates analysed.

The classification of enterococcal plasmids according to the content in replicases / relaxases / toxin-antitoxin systems is shown in Figure 4. The average number of rep genes per isolate is 2,37 (1- 5 rep genes per isolate) which correspond to i) tetha replicative plasmids of the families Rep_AN subgroups of pheromone responsive plasmids *rep9*, 73% (*rep9*A_{pAD1}, 23%; *rep9*B_{pBEE99}, 30%; *rep9*C_{pCF10}, 40%), and *rep17*$_{pRUM}$, 16%; Inc18 (*rep1*$_{pRE25}$, 12%; *rep2*$_{pVEF1}$, 5%), and

Rep3_small theta (*rep18*$_{pEF418}$, 5%; *rep*$_{pCIZ2}$, 5%; *rep6*$_{pAMα1/\ pS86}$, 33%) and ii) rolling circle plasmids of Rep_trans family (*rep7*$_{pS194}$-like, 44%; and *rep14*/ orf1$_{pEFNP1}$, 5%). Only relaxases of MOBC (Orf57$_{pAD1}$, 44%) and MOBP (PcfG$_{pCF10/pBEE99}$, 56%) families were detected; 44% of the strains containing both MOBC and MOBP. iii) TA systems of pheromone responsive plasmids (fst$_{pAD1}$, 23%) and Inc18 members (*ε-ζ*$_{Inc18}$, 7%) were identified in a minority of plasmids (71) (Figure 4).

We found the presence of different Tn*916*-like elements (Tn*916,* 35%; Tn*5398,* 19%; Tn*5801,* 2%; Tn*5397*; 2%) and Tn*917* (Tn3 family, 34%).

Pheromone plasmids are traditionally linked to virulence and antimicrobial resistance. The 87% of the positive in *asa1* were positive for rep$_9$, and only 7% of the samples analyzed for virulence genes were positive in *rep9* but negative *asa1* gene.

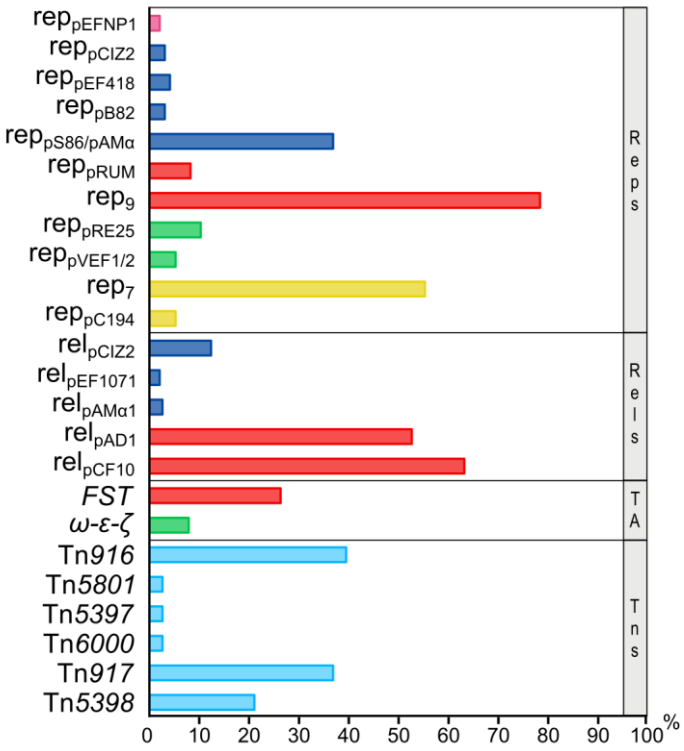


Figure 4. Plasmid content of *E. faecalis* from wild birds. Reps, replicases; Rels, relaxases; TA, toxin-antitoxin systems; Tns, transposons. Each colour represent a plasmid family: purple, RCR family; blue, small theta family; red, RepA_N family; green, Inc18; yellow, rolling circle plasmids of Rep_trans family; cyan, transposons.

Comparative analysis of core and accessory genomes

Given the wide diversity observed among the wild bird isolates according to the MLST results, we fully sequenced five strains from wild birds of different lifestyle (*Ciconia ciconia*, *Gyps fulvus*,

Strix aluco, *Ciconia ciconia* and *Falco tinnunculus* respectivelly),which showed a related *Sma*I-genomic digested DNA profile but different STs.

Core genome

We will focus our efforts in the isolates of animal origin, emphasizing the results from wild birds, which are scarcely represented in public databases. In fact, only the 7,5% (34/450) of the available genomes are from non-human origin and include pigs (n=21; 9 ST16, 4 ST40, 2 ST47, 1 ST32, 1 ST35, 1 ST58, 3 unknown ST), insects as Drosophila (n=2; 1 ST101, 1 ST102), non-flying birds (penguin, 2ST475; chicken, 1 ST59; duck, 1ST167; and turkey, 1 ST116), tunicates (1 ST92), mice (1 ST9), pets (cat, 1 ST202; dog, 1 ST108), and farm animals (sheep 1 ST16 and cattle, 1 ST40).

To analyze our isolates in the context of the population structure of *E. faecalis*, the SNP core of all the 450 available genomes in the databases were extracted and a phylogenetic tree comprising all the novel genomes from disparate hosts (including those from this study) was built. Most of the isolates sequenced and deposited in the database are from humans and clustered in two dominant branches, ST6 and ST40. In a previous study, Raven et al. sequenced a pool of clinical origin isolates from the UK and demonstrated the dominance of two more branches corresponding to CC87 [ST28-ST640 and CC388 (ST103) (31). In fact, all STs cluster together, with the exception of some new and scarcely represented STs that cluster between the dominant STs (Figure 5).

It is of note that the five isolates chosen for sequencing corresponded to known STs. Despite the similarity of the PFGE patterns, the wild bird genomes cluster in three different branches. That of AE51 (ST16) groups with isolates of animals (9 from pigs and 1 from turkey) and humans; that of AE63 (ST55) clusters in a large branch that includes ST58 (from pigs and environmental origin) and ST179 (hospitalized humans). Within this large branch, AE63 (ST55) within a small cluster with ST55 and it is closely related to ST40 from farm animals and clinical isolates). The AE8 (ST170), AE60 and AE65 (ST82) cluster in a small and distinct branch besides a ST82 isolate from human origin and two strains from penguins (ST256 and ST475).

Very recently, a 7.4 kb region of a pathogenicity island (PAI) of *E. faecalis* containing different traits associated with pathogenicity was considered as a discriminatory factor between pathogen or not-pathogen *E. faecalis* strains (81). This 7.4 kb region includes 4 genes, two of them encode for proteins related with type VI secretion systems, one with a membrane protein

of the MAEBL family, and a conserved lipoprotein of unknown function (81). We searched this specific region to determine the potential pathogenicity between our isolates and those phylogenetically related. This approach is not functionally validated but it has shown a highly discriminatory potential between two big phylogenetic clusters associated with commensals and pathogen enterococci. Two strains from birds, AE51 (ST16) and AE63 (ST55), were positive for this 74 kb region.

Accesory genome

We built an accessory genome network using the ACCNET tool with the aim to analyze the genomes related to their core and also to their accessory genome. We identified 11630 "representative" proteins in the accessory genome of all the available *E. faecalis* genomes, which were classified with the OrthoMCL tool (82). The genomes were clustered according the accessory genome profiles (95% of similarity between genomes and proteomes respectively) (Figure 6 and Figure 7).

Trees of core and accessory genomes were congruent for the isolates AE51 (ST16) and AE63 (ST55) but not for AE8 (ST170), AE60 and AE65 (ST82). According to the accessory genome, AE51 and AE63 cluster with other ST16 and ST40 isolates while the other three isolates cluster with two accessory genomes from penguins in the "root" of the dendrogram (Figure 4 and 5). The accessory genome of AE8, AE60 and AE65 isolates include a set of 102 proteins specifically associated with the ST82 branch. Using the COG database, a tool for genome-scale analysis of protein functions (82), only the function of 38/102 proteins could be catalogued. They include energy production and translation and ribosomal structure (5% each), membrane proteins (3%), carbohydrate transport and metabolism (16%), coenzyme transport and metabolism (8%), transcription (13%), replication (13%), recombination and repair, defense mechanisms (13%), and mobilome related proteins (10%). These CC82 isolates share a small cluster of 10 hypothetical proteins with ST256 and ST475 isolates.

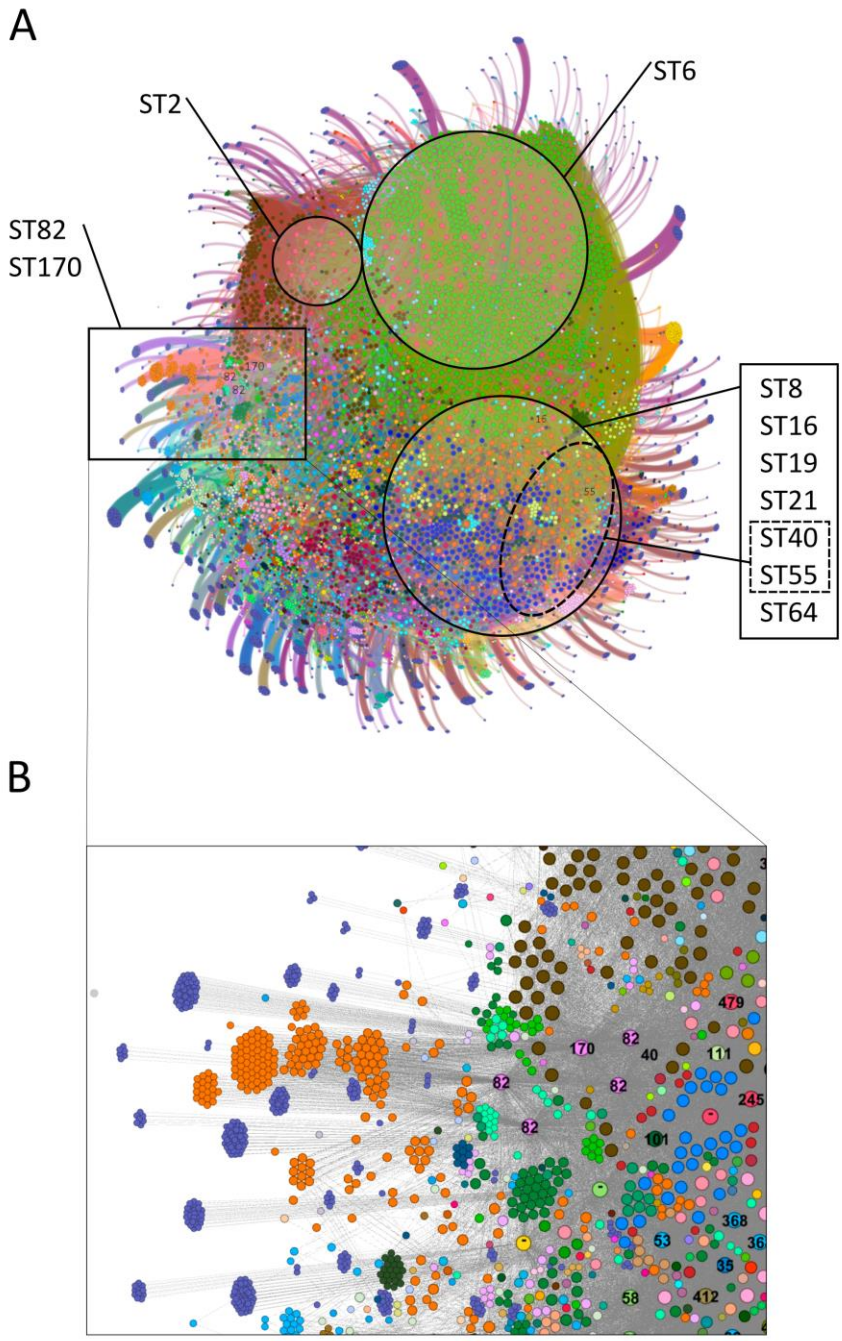


Figure 6. A. Representation of the accessory genome of available *E. faecalis* genomes (this study and deposited in public database) by using the ACCNET tool. The isolates (bigger nodes) as well as the proteins were subjected to a clusterization of 95% similarity. Such clusterizations can be visualized by the colours of the nodes. The network shows a major cluster with red (isolates) and green (proteins) nodes which correspond to ST2, ST6 and ST9 isolates. The green protein cluster comprises proteins common to isolates distributed along the network. This explain why the green cluster is integrated in the network despite of it appears as a specific cluster with red isolates. The orange isolate cluster includes a variety of STs comprising ST16 and ST55 of the wild bird isolates. B. The orange protein clusters shows specific proteins belonging to the wild bird isolates (ST82 and ST170).

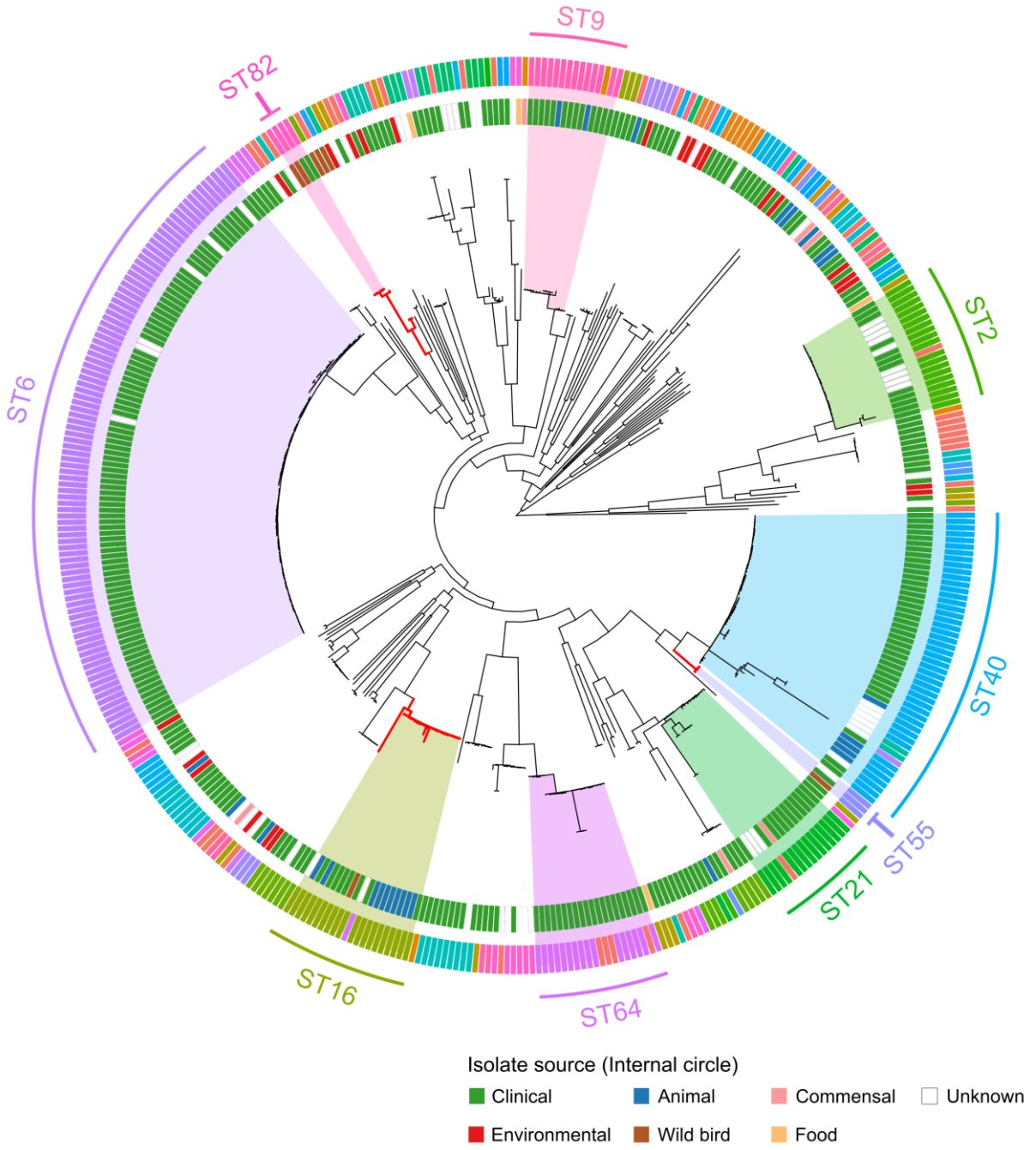


Figure 5. Core genome phylogenetic tree of all the *E. faecalis* genomes deposited in the databases. The outside circle indicate the STs: different shades of colours show different STs. The inner circle represents the source/host of the isolates. The red branches correspond to the wild bird isolates of our collection. The other coloured branches highlights the major STs presented in databases. The ST82 branch clusters with two isolates from wild birds (penguins) of the database.

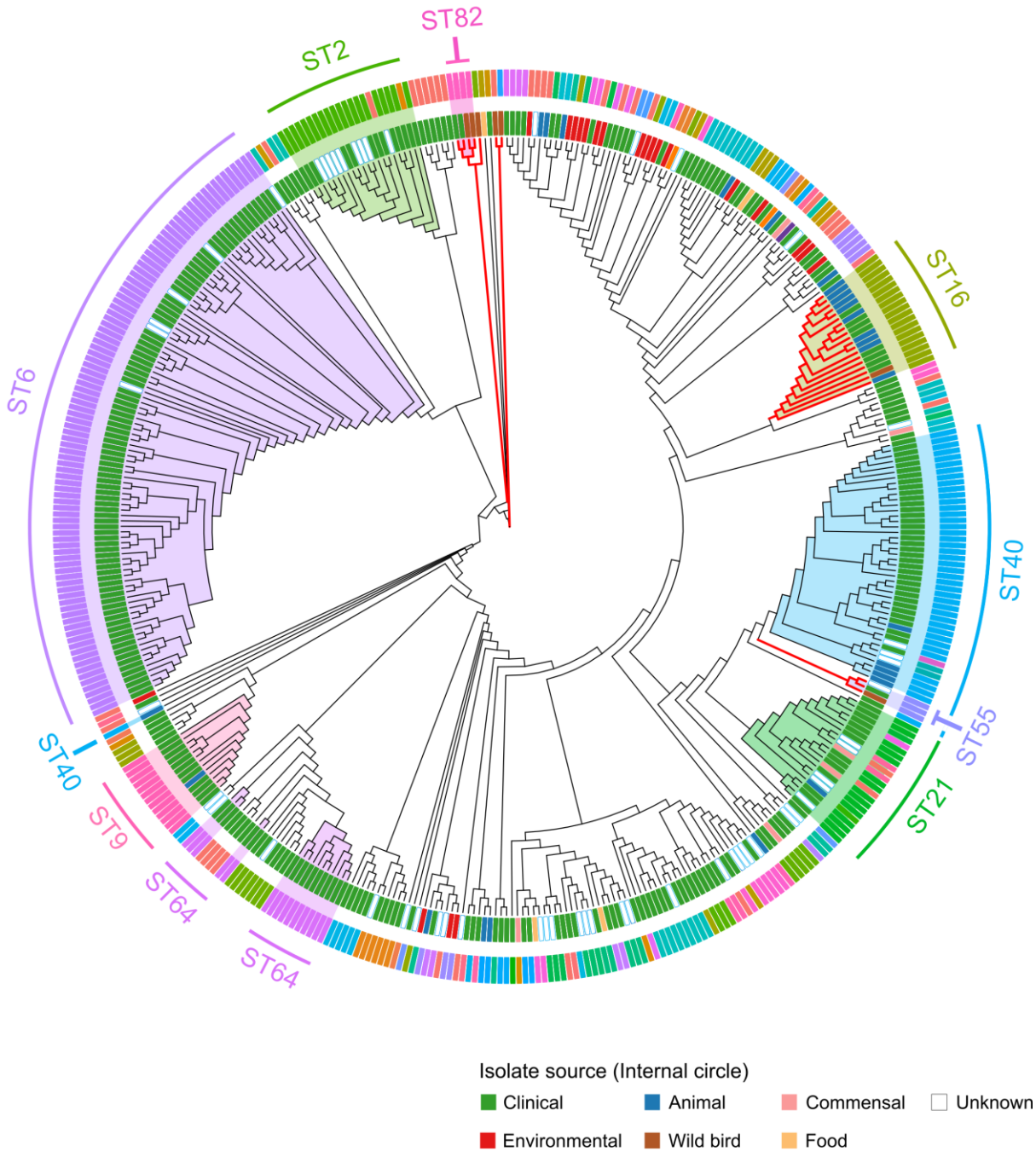


Figure 7. Accessory genome dendrogram of all the *E. faecalis* genomes of the databases. This dendrogram was obtained by constructing a binary comparison matrix with the presence/ absence of all the representative accessory proteins in all the isolates. The outside circle indicate the STs: different shades of colours show different STs. The inner circle represents the source of the isolates. The red branches include wild bird isolates from our collection. The other coloured branches highlight the major STs presented in databases. The ST82 branch clusters near the "root" or the origin of the dendrogram, and with a distinct accessory genome with the other groups and very similar to the two isolates from wild birds (penguins) of the database.

Plasmid content characterization of the *E. faecalis* from wild birds sequenced

The number of plasmids detected in the genome sequences using both PLACNET and SPADES was consistent with that observed in the S1-PFGE gels. The *E. faecalis* AE51 (ST16) strain does not contain any plasmid. All plasmids detected in the other four strains were pheromone responsive according of rep and MOB genes, and they never were described before but carrying some different modules and ICEs similar to other previously described (Figure 8).

The AE8 isolate (ST170) harbor a 77kb plasmid (pAE8) that is a chimera of pheromone (pCF10, pBEE99, pTEF2, pEF123) and Inc18 plasmids (pRE25). The pAE8 comprises a region identical to pRE25 (GenBank: X92945.2) (83) that include genes coding for the two replication initiator proteins of Inc18 plasmids and also genes conferring resistance to erythromycin (*erm(B)*) and chloramphenicol (*cat*). One scaffold was identical to other of pEF123, a pheromone related plasmid of an *E. faecalis* isolate from chicken (Genbank: KX579977.1; size: 79682 bp) that carries genes conferring ressitance to aminoglycosides (*str/aadE*).. The pAE8 relaxase is related to that of pheromone plasmids pBEE99, pEF123 and pCF10. Furthermore match with a truncated Tn*6248.* This transposon appear in database in *E. faecium* strain E506 (GenBank: KP834592.1), linked to a tetracycline and chloramphenicol resistance gene cluster [*tet*(M)-*tet*(L)-*cat*]. This fragment lacks the *cat* gene belonging to this transposon. Tn*6248* is widespread between streptococcus and staphylococcus. The last fragment of the plasmid does not match with any known plasmid but it contents a *mazF/pinE* region.

Plasmids pAE60 and pAE65 from the ST82 *E. faecalis* isolates AE60 and AE65 were identical (41.9 kb). They carry the replication gene of the pheromone related plasmids pCF10, pTW9 or pEF123 (with a 97% of similarity) and has a similar module with the bacteriocin related region of *E. faecalis* plasmid pYI17 (D78257.1). In addition, these plasmids carry the *mazE/pinE* region and different type-1 restriction modification system related genes. The entire plasmid match with different partial sequences of the *E. faecalis* EnGen0363 strain RMC5 (ST53) (Sequence IDs: ASDD01000025.1; ASDD01000025.1; ASDD01000029.1; ASDD01000030.1 and ASDD01000031.1). This plasmid has a pheromone origin, but its size is smaller than the average of most pheromone plasmids.

Finally, the pAE63 (58.6 kb) is another plasmid chimera that comprises sequences from pheromone plasmids pTEF1 and two plasmids previously identified in the *E. faecalis* CLB21560

clinical isolate, namely pB (Sequence ID: CP019514.1; Length: 65920) and pA (Sequence ID: CP019513.1; Length: 66602). It also contain *asa1*.

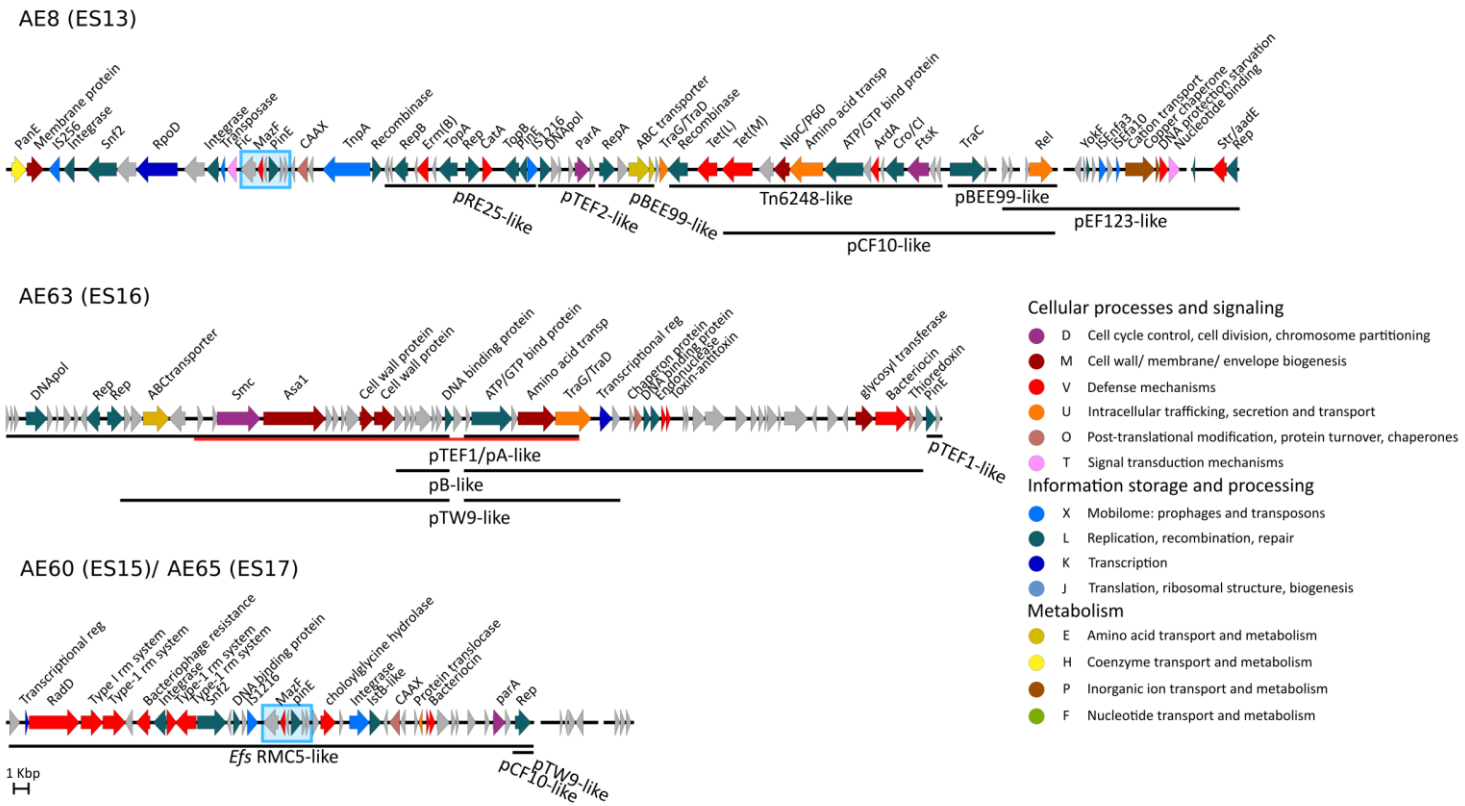


Figure 8. Novel *E. faecalis* plasmids identified in this study. The arrows represent the different genes coloured by the COG functional classification (indicated in the legend). The lines under the backbones represent the regions that exhibit high identity (97-99%) with other known plasmids. The GeneBank accession number of the sequences are: pRE25 (X92945.2), pTEF2 (AE016831.1), pBEE99 (GU046453.1), pCF10 (AY855841.2), pEF123 (KX579977.1), pTEF1 (AE016833.1), pTW9 (AB563188.1). The plasmids pA (CP019513.1) and pB (CP019514.1) were previously detected in the *E. faecalis* CLB21560 strain collected from a human clinical sample. The Tn*6248* was found in *E. faecium* strain E506 (KP834592.1). Finally, the plasmid detected in AE60 and AE65 were identical, and uniquely showed identity with some fragments of a plasmid identified in the partial genome database of *E. faecalis* EnGen0363 strain RMC5 (ST53) (ASDD01000025.1; ASDD01000025.1; ASDD01000029.1; ASDD01000030.1 and ASDD01000031.1).

Discussion

Different genetic studies based on MLST data analyzed by different tools (goEBURST, BAPS) (23, 26, 84, 85) or cgMLST (26, 30) have reflected the influence of recombination in the diversity of *E. faecalis* and evidenced the difficulties to establish the population structure of this species as well as to assign STs into evolutionary groups. An "endemic polyclonal structure" favoured by vertical and horizontal inter-host migration would allow colonizing disparate host with minimal impact. However, the scarcity of sequences from non-humans hosts and the lack of information about the accessory genome, the mechanisms of HGT preclude validating such conclusions.

The limited specialization of *E. faecalis* suggested in previous works, is also reflected by the results we obtained using BAPS and comparative genomics and correlates well with the observed high rates of recombination in this species. Isolates were found in different BAPS groups confirming either the low specificity of the method to establish ecotypes in this species (84) or the need to use other approaches for species with a small core and large pangenomes (Black Queen Hypothesis) (63, 86). Despite the apparent low specialization, some STs were predominant in different hosts (humans, farm animals and wild birds). The few works that characterize the isolates by MLST identified known STs as ST6 or ST16, although transmission could not be demonstrated in any of the cases. Our work revealed hallmarks that indicate evolution of some lineages in different hosts as ST40, ST16 or ST55. Moreover, CC82, the CC that only comprises isolates from birds (three isolates of different bird orders and two from pinguins) shows a similar core but specific accessory genome suggesting a niche-specific convergence. A preliminary analysis of habitat-specific proteins in the CC82 *E. faecalis* genomes revealed functions related to metabolism adaptation and mobile genetic elements although further studies are necessary to fully characterize these adaptive traits.

The presence of a hallmark of a pathogenicity island (PAI) that shows a discriminatory potential for distinguishing pathogenic and non-pathogenic strains of *E. faecalis* (four genes), in known ST (ST16 and ST55) but not in those of CC82, is of concern and would give these strains the ability to transfer their pathogenicity through wild bird migrations.

The accessory genome and the plasmidome of *E. faecalis* is firstly described in this work using novel tools ACCNET (66), and PLACNET (72). We described a large accessory genome and further highlight the relevance of HGT in the evolvability of the species. It is remarkable both the high number of strains lacking plasmids and the high mosaicism of plasmids carrying known

modules of different plasmid families like Inc18, pheromone plasmids, antibiotic resistance, restriction-modification systems, the *mazE/pinE* module, or bacteriocin related genes. In this regard, it is remarkable the high rate of unique STs found in both wild birds and humans. While all MLST alleles of wild birds were known alleles, a high percentage of MLST alleles of humans were new, probably due to recombination favored by an increasing population size under selection in hospitals. This finding, relevant for the evolution of the species would also explain the results obtained with plasmids. Frequent horizontal gene transfer events between phylogenetically distant or related populations constitute a new thrust in the adaptation of these strains to new hosts and environments.

In conclusion, wild birds seem to be reservoirs of *E. faecalis* resistant to antibiotics as reflects the recovery of different genes from some bird species. Comparative analysis of core and accessory genome confirm high levels of recombination that influence the host-specificity of some lineages. Further studies are necessary to explore the specific traits acquired by host adapted to specific lifestyles as that of some wild birds.

Chapter II

Diversity and Evolution of the Tn*5801-tet*(M)-like Integrative- Conjugative Elements Among *Enterococcus*, *Streptococcus* and *Staphylococcus*

The work related with this chapter was published in:

León-Sampedro R, Novais C, Peixe L, Baquero F, Coque TM. 2016. Diversity and Evolution of the Tn *5801-tet* (M)-Like Integrative and Conjugative Elements among *Enterococcus, Streptococcus*, and *Staphylococcus*. Antimicrob Agents Chemother **60**:1736–1746.

Chapter II: Diversity and Evolution of the Tn*5801-tet*(M)-like Integrative- Conjugative Elements Among *Enterococcus*, *Streptococcus* and *Staphylococcus*

Abstract

This work describes the diversity and evolution of Tn*5801* among enterococci, staphylococci and streptococci based on the analysis the 5,073 genomes of these bacterial groups available at gene databases. We also examined 610 isolates of *Enterococcus* (10 countries, 1987-2010) for the presence of this and other known CTn-*tet(*M*)* due to the scarcity of data about Tn*5801* among enterococci. Genome location (I*Ceu*-I-PFGE hybridization/integration-site identification), and transferability (mating filter method), Tn*5801* characterization (long-PCR mapping/sequencing) and clonality (PFGE/MLST) were studied. Twenty-three Tn*5801*-variants (17 unpublished) clustered in two groups designed as "A" (25kb; n=14; predominant in *Staphylococcus aureus*) or "B" (20kb; n=9, predominant in *Streptococcus agalactiae*). The GC % of the common backbone suggests a streptococcal origin of the Tn*5801* group-B, with further acquisition of a 5kb fragment that resulted in the group-A. Deep sequence analysis allowed identifying variants associated with clonal lineages of *S. aureus* (CC8, ST239), *S. agalactiae* (CC17), *E. faecium* (ST17/ST18) or *E. faecalis* (ST8), local variants, or variants located in different species and geographical areas. All Tn*5801* elements were chromosomally located upstream the *guaA* gene, which serves as integration hotspot. Transferability was only demonstrated for Tn*5801*–type B among *E. faecalis* clonal backgrounds, eventually harbouring another Tn*5801* copy. The study documents early acquisition of Tn*5801* by *Enterococcus*, *Staphylococcus* and *Streptococcus.* Clonal waves of these pathogens seem to have contributed to the geographical spread and local evolution of the transposon. Horizontal transfer, also demonstrated, would explain the variability observed, often containing sequences of different origins.

Introduction

The extended use of tetracycline since its introduction in the therapeutic arsenal in 1948 seems to have resulted in an evolutionary bottleneck in the population structure of some genera of Firmicutes of interest in human health. Recent studies have associated the acquisition and further fixation of tetracycline resistance (TETR) by pathogenic clones of group B *Streptococcus* (GBS) with the global increase of high mortality GBS neonatal infections in the last decades (87, 88). Similarly, some *Staphylococcus aureus* lineages are enriched in TETR elements (47, 89, 90).

Tetracycline resistance in major Gram Positive human opportunistic pathogens is mainly caused by the acquisition of integrative and conjugative elements (ICE) of the Tn*916* family carrying the *tet*(M) gene (47, 91, 92). These elements display a common synteny but differ in the integrase (*int*) and the excisionase (*xis*) sequences, the insertion site specificity, and the host range (55, 93). Variants of some Tn*916*-*tet*(M) members, namely Tn*916*, Tn*5397*, Tn*6000* or Tn*5801,* are widely spread among several genera of Firmicutes apparently suggesting a successful multi-host dissemination of these elements (92).

Tn*5801* has a site specific tyrosine recombinase that recognizes the 3´-end of guanosine monophosphate (GMP) synthase gene (*guaA*) (94, 95), an insertion hotspot of genomic islands coding for pathogenicity or antibiotic resistance in different Firmicutes (96–98). Tn*5801* was originally detected in the vancomycin-resistant *Staphylococcus aureus* clinical strain Mu50 (VRSA) recovered in Japan in 1997 (99), but several studies have documented the presence of platforms highly similar to Tn*5801* in early isolates of *Streptococcus agalactiae* (France, 1953) (87), *S. aureus* (Denmark, 1963; designed as Tn*6014*) (89), and *Clostridium perfringens* (USA, 1977; a truncated element designated as CW459*tet*(M)) isolates (100). The diversity of Tn*5801*-like backbones in a few isolates of different species (89, 101–104) collected throughout more than 50 years, suggests the evolutionary interplay of selective events, horizontal dissemination, and the genetic plasticity of this transposon, allowing its spread in different clonal backgrounds and environments. However, little is known about Tn*5801*, either at molecular or epidemiological level.

In this study, we created a dataset comprising all the genomes of Firmicutes with the presence of an ICE Tn*5801*-like. Due to the relevance of enterococci in the spread of TETR and the low number of the sequenced genomes in public databases at the time this study was started, we also analysed the presence and genetic background of Tn*5801* in a large collection of

enterococcal strains from different geographical areas as well as its transferability. This comprehensive phylogenetic and genomic analysis allowed documenting the genetic variability of Tn*5801*-like elements among genomes of enterococci, staphylococci, streptococci, and other bacterial species through years and continents. The results highlight the relevance of comprehensive analysis al local and global level to accurately establish transmission and evolvability of elements at different scales. It also proved the transferability of the element in different genetic contexts.

Materials and methods

Screening of Tn*5801* in Genbank databases and creation of datasets

DNA sequence from *S. aureus* Mu50 (GenBank accession number BA000017) was used as reference to screen the presence of Tn*5801* among 5,073 draft and complete genomes from Firmicutes (4,130 *S. aureus*, 345 *Enterococcus faecalis*, 301 *S. agalactiae*, 254 *Enterococcus faecium*, 23 *Streptococcus mitis*, 14 *Lactococcus garviae*, 4 *Staphylococcus pseudointermedius*, 2 *Enterococcus villorum* and 1 *Streptococcus thoraltensis*) obtained from NCBI archive (last updated January 2015). BLAST searches identified different sequences showing homology with Tn*5801-like* elements. They were further stored and analyzed by using MUSCLE algorithm (MUltiple Sequence Comparison by Log- Expectation) (http://www.ebi.ac.uk/Tools/msa/muscle/), and Artemis Comparison Tool (105) (http://www.sanger.ac.uk/resources/software/act/). ICE maps were created with Vector NTI (http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/Cloning/vector-nti-software.html) and Inkscape softwares (http://inkscape.org/).

Phylogenetic analyses were accomplished using MEGA version 6 (106). A common region of 13,895 bp was identified among all the sequences with homology with Tn*5801* (called here "Tn*5801* variant sequences"). The region comprises 16 *orfs* (*int*$_{Tn5801}$-*orf*24; *orf*22, *orf*18, *orf*16, *orf*15, *orf*12-10 and *orf*8-4) that were trimmed and concatenated to be further analyzed. A maximum likelihood phylogenetic tree was constructed with 1,000 bootstrap replications in order to infer the relationship among all Tn*5801* variants. Due to the identification of modules with different GC content and with recombination landmarks, the Tn*5801* variants were further split in five core fragments (*int*$_{Tn5801}$, *orf*26-19, *orf*18-14, *orf*12-8 and *orf*7-4), each fragment being separately analyzed as mentioned before. The *orf*26-19 fragment was excluded from this analysis because of the variability observed among variants due to frequent indels and recombinatorial events. An arbitrary number was attributed to different sequences of each core fragment, and Tn*5801* sequences with unique number profiles were used to calculate a consensus tree. The dataset created included epidemiological information of isolates (host/source, year of isolation, country, Sequence type, ST). To identify the ST of the isolates, we used the MLST 1.8 software of the Center of Genomic Epidemiology server (107).

Bacterial strains epidemiological background and identification of *tet*(M) transposons among enterococci

A collection of 610 enterococcal isolates from human (n=320; hospitalized patients, HP, n=195; and healthy human volunteers, HV, n=125); animals (n=236; poultry, P, n=164; and swine/piggeries, SP, n=72) and hospital/urban sewage (n=54, SW) collected in different countries along a 23 year period (1987-2010) was included (55, 56). All isolates were tested for susceptibility to tetracycline by disk diffusion and/or agar dilution method following CLSI guidelines (67). The presence of genes coding for tetracycline resistance [*tet*(M), *tet*(L), *tet*(S), *tet*(K), *tet*(O)] was assayed by PCR as previously described (108). The presence of main transposons of the Tn*916* family (Tn*5801*, Tn*916*, Tn*5397*, Tn*5386*, Tn*6000*) was investigated by a multiplex PCR assay designed here for detecting integrases, excisionases and other specific sequences of known Tns. The susceptibility and specificity of the assay was validated using appropriate controls. Primers and conditions are listed in Table S1. Clonal relationship among *E. faecium* and *E. faecalis* isolates harbouring the integrase of Tn*5801* (int_{Tn5801}) was established by pulsed field gel electrophoresis (PFGE) and multilocus sequence typing (MLST) (http://pubmlst.org/) (109, 110). Population genetic analysis was performed using BAPS software in *E. faecium* isolates as previously described (28, 34).

PCR- characterization of Tn*5801* backbones in *Enterococcus*

The Tn*5801* backbone was fully characterized by a PCR mapping assay based on the Tn*5801* sequence of *S. aureus* Mu50 (GenBank accession number BA000017). Amplified fragments obtained from positive int_{Tn5801} strains were further sequenced (Table 1, Figure S1). Genomic location of Tn*5801* was assessed by hybridization of I-*Ceu*I (TaKaRa biotechnology, Dalian, China) digested genomic DNA using specific int_{Tn5801} and 23S rRNA probes (111). Amplification of the int_{Tn5801}-*guaA* region was performed in order to characterize the integration site of the Tn*5801* (Table 1, Figure S1) (98).

Transferability of Tn*5801*

It was screened by filter mating using different recipient strains of *E. faecalis* (JH2-2, OG1RF, OG1SSp) and *E. faecium* (64/3, BM4105RF, BM4105SS), all being resistant to rifampicin and fusidic acid (designed as RF) or streptomycin and spectinomycin (designed as SS) and negative for the presence of *tet*(M) gene (112). The mating assays were performed at 37ºC for 36 h using aliquots of broth cultures of donors and recipient growth up to the exponential phase

(McFarland 0.5) and mixed at a ratio of 1:1 (112). Transconjugants were selected on Brain Heart Infusion (BHI) agar (Pronadisa, Laboratorios Conda, SA, Madrid, Spain) plates supplemented with tetracycline-10 mg/L plus rifampicin-30 mg/L and fusidic acid-25 mg/L or, streptomycin-250 mg/L and spectinomycin 250 mg/L (Sigma-Aldrich, Inc., St. Louis, MO). Unexpectedly, we identified a Tn*5801* Δ*tet*(M) variant in the laboratory recipient *E. faecalis* strain JH2-2, which is an emblematic receptor strain (see below). Secondary filter mating assays were performed using *E. faecalis* JH2-2 transconjugants as donors and *E. faecalis* OG1SS and *E. faecium* BM4105SS strains as recipients. Transconjugants were characterized according the susceptibility to tetracycline (Etest, bioMérieux SA, France), the presence of Tn*5801* [int_{Tn5801} and *tet*(M)], the transposon backbone (PCR overlapping) and comparison of *Sma*I-digested DNA PFGE profiles of transconjugants with those of recipients and wild type strains. The location of Tn*5801* in the recipient genomes was assessed by hybridization of *Sma*I-digested DNA with specific *guaA*, int_{Tn5801} and *tet*(M) probes. PCR of the fragment *guaA*-int_{5801} was performed in all transconjugants.

Nucleotide sequence GenBank accession number

The A4 variant of Tn*5801* detected in this study is available at the GenBank database with accession number KP001176.

Table 1. Characterization of Tn*5801*: PCR strategy and conditions

Primer nº	Sequence (5´-3´)	Position in Tn*5801** (bp)	PCR reaction	Sequence amplified	Amplicon size (bp)	Amplification conditions
gmp1	CCTGGTCTTGGTATTCGTGT	25763- 25783 [a]	gmp1-gmp2	*guaA*	170	95ºC- 10'- 1 cycle; 94ºC-30'', 54ºC-30'',72ºC-30''- 25 cycles; 72ºC-10'- 1 cycle
gmp2	GTAGTCTCCCATAACACC	25609- 25627 [a]				
Int1	CTGTTTCCGATATTGAGC	24490-24473 [a]	Int1- gmp1	*int*Tn*5801- guaA*	1311	35 cycles of 30'' at 96ºC, 1' at 45-59ºC (according to primer pair), 7' at 72ºC; 1 cycle 10' at 72ºC.
P1	GTTTCGCAAGTAGTCTACAG	25310-25329 [a]	P1- P2	*orf25- int*Tn*5801*	2085	
P2	GTAAAGGCGACAGATGG	23244-23260 [a]				
P3	CTTAGAGATGAGTTTCGTTTC	23391-23411 [a]	P3- P4	*orf22- orf25*	1455	
P4	GGATAGTTCTTTGTCTGTAAAG	21956-21977 [a]				
P5	CGATTTTAGAGCCGTTGGTTTAG	22210-22232 [a]	P5- P6	*tet*(M)- *orf22*	2448	
P6	GAACGTCAGAGAGGAATTAC	19784-19803 [a]				
P7	GGGAATCCCCATTTTCCTAA	19986-20005 [a]	P7- P8	*orf15*- tet(M)	6498	
P8	ACCCACTCGCTGTTTAATCG	13507-13526 [a]				
P9	CTTGCAAGGCTAGGTTGGAG	14162-14181 [a]	P9- P10	*orf7- orf15*	5993	
P10	ATGTCTGAAATGGGCTTTGG	8188-8207 [a]				
P11	CTTTCACTTCGTGCGGTACA	8382-8401 [a]	P11- P12	*orf3- orf7*	4631	
P12	GGGTGGGGACAATACATCAG	3770-3789 [a]				
P13	CAGCCATGTAGCGTCTTTGA	4315-4334 [a]	P13- P14	*orf1- orf3*	4228	
P14	ACGGAGTTAACGGCTTTCCT	106-125 [a]				
P15	CCGTATGTTCTTTCAACCACT	1030-1051 [b]	P15- P11	*Orf5- orf7*	2675	
P16	GCTGACAGTTCCAGTATCC	18096-18115 [c]	P16- P5	*orf23'- orf22*	1209	
P17	CAGAACCAGCCATTACC	18561-18578 [c]	P17- P18	*orf23'- orf19'*	3938	
P18	GAAGATACGAGAAACCAATAG	14602-14623 [c]				
P19	GGTGGCAATTCAAGTGTTCC	14859-14878 [c]	P19- P20	*orf19'- orf17*	1957	
P20	GGCGTATGACAAAGCTGG	12921-12939 [c]				

Results

Genetic diversity of Tn*5801* in the genomes of four major Gram-positive genera

Twenty-three Tn*5801* variants were identified among the 5,073 genomes screened, 17 of which had not been previously published. The 23 distinct Tn*5801* backbones corresponded to 225 strains (139 *S. aureus*, 43 *Streptococcus agalactiae*, 24 *E. faecium*, 15 *E. faecalis*, 1 *Streptococcus mitis*, 1 *Enterococcus villorum,* 1 *Streptococcus thoraltensis* and 1 *Lactococcus garviae*). Table 2 and Table S2 (Appendix 1) show the epidemiological data and the dataset of the analyzed genomes.

The Tn*5801* backbones were classified in two main groups arbitrarily designated by capital letters as "group A" (14 types, A1-A14) and "group B" (9 types B15-B23) that are shown in Figure 1. These two groups differ in length (20-25kb, due to the presence of an extra 5kb fragment at the left arm of the group A transposons), the nucleotide sequence, and the presence of indels and rearrangements in different regions. Variants within each group were designated by the corresponding capital letter followed by a number. The detailed analysis of the core sequences facilitated detection of differences and common features within a given type (Figure 2; Table S2 in Appendix 1). The GC percentage of the common Tn*5801* backbone (GC=36%) contrasted with that of the 5kb fragment found in type A sequences (GC=32%) (Figure S2).

Table 2. Presence of Tn*5801* types A and B in Firmicutes sequences deposited in the GenBank genome database as of January 2015

Species	**No. of genomes**	**Presence of**	**Presence of**
Streptococcus agalactiae	301	1	42
Streptococcus mitis	23	1	-
Streptococcus thoraltensis	1	-	1
Enterococcus faecalis	344	1	13
Enterococcus faecium	254	16	7
Enterococcus villorum	2	2	-
Staphylococcus aureus	4130	139	-
Staphylococcus	4	1	-
Lactococcus garvieae	14	1	-

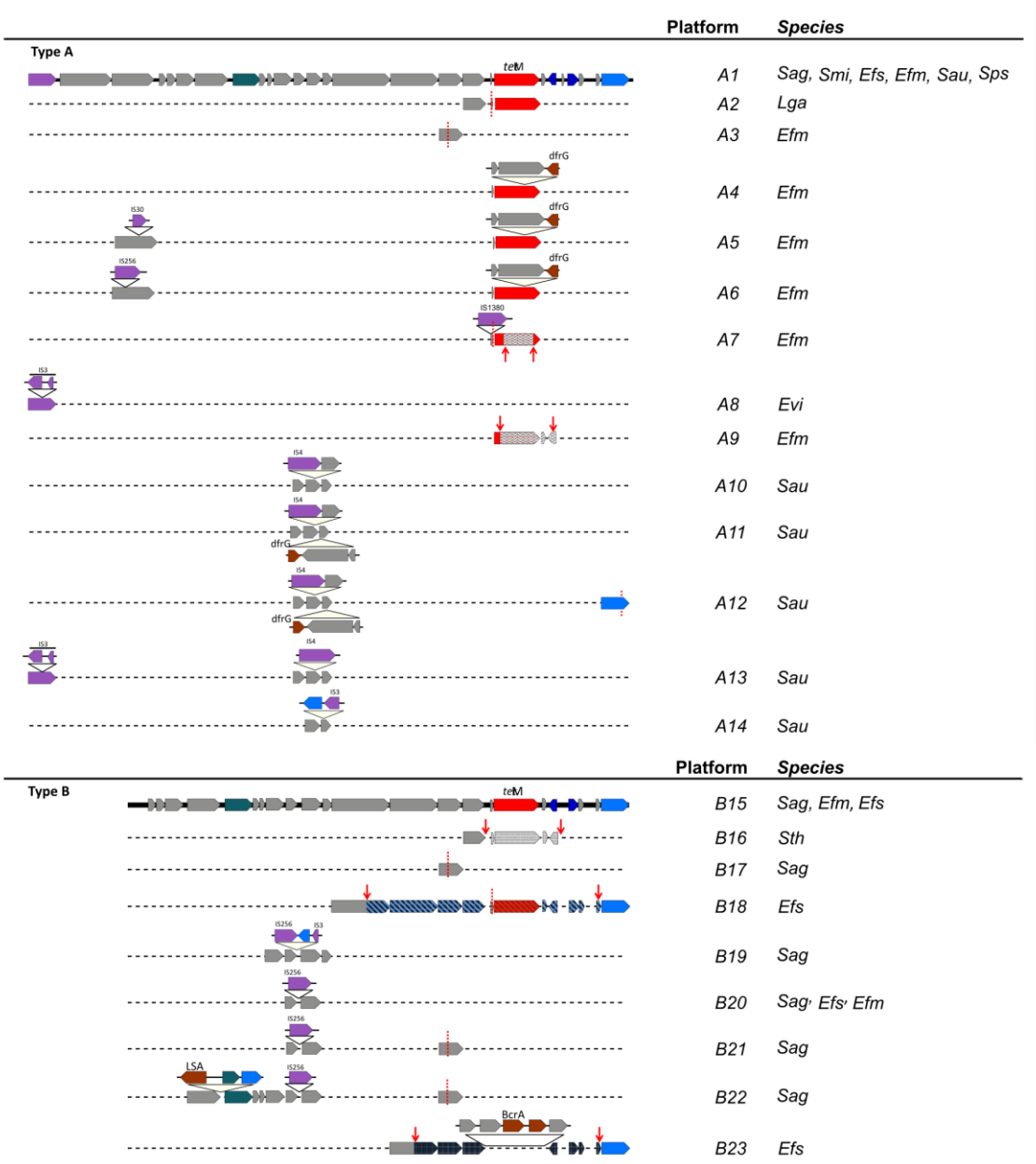


Figure 1. Diversity of **Tn*5801*-**like elements, types A and B, based on strains used in this study and sequences deposited in the GenBank genome database. In silico comparative analysis was made using BLAST and MUSCLE software available at the BLAST (http://www.ncbi.nlm.nih.gov/blast//) and MUSCLE (http://www .ebi.ac.uk/Tools/msa/muscle/) websites. Dashed red lines, deletion sequences; blue ORFs, ORFs encoding integrase-like proteins or mobilization-like proteins; dark blue ORFs, ORFs encoding replication proteins; grey ORFs, ORFs encoding hypothetical proteins; red ORFs, tetracycline resistance gene *tet*(M); brown ORFs, other resistance genes; purple ORFs, insertion sequences (IS). Wavy lines, recombination with Tn*5397*; straight lines, recombination with Tn916; diagonal lines, recombination with type A variant. ORFs with spots, B23 sequence differing from type B variants. dfrG, dihydrofolate reductase; LSA, lincosamides and streptogramins A; BcrA, bacitracin transport; Sag, *Streptococcus agalactiae*; Smi, *Streptococcus mitis;* Efs, *Enterococcus faecalis*; Efm, *Enterococcus faecium*; Sau, *Staphylococcus aureus*; Sps, *Staphylococcus pseudintermedius*. All IS corresponds to IS families: IS1062 (IS30 family, A5), IS1542 (IS256 family, A6), IS1678 (IS1380 family, A7), ISEfa8 (IS3 family, A8), IS1542 (IS256 family; B19, B20, B21, and B22), ISLgar5/ISEfm2 (IS256 family, orf1).

In the following paragraphs we analyze in detail the features of variants of these clusters.

a) Tn*5801* "group A". Fourteen variants (A1-A14) were identified among the available sequences (Figure 1, Table 3, and Table S2 in Appendix 1). Phylogenetic analysis allowed us to group them in different subgroups of highly related variants (Figure 2).

Tn*5801* type A1 was considered the paradigm of the group. It was mostly detected in *Staphylococcus* (96%; n=107 *S. aureus/S. pseudointermedius* out of 112 isolates with Tn*5801* type A1) and was also the predominant Tn*5801* variant within this genus (76%; n=107/140 of *Staphylococcus* carrying Tn*5801*). Variant A1 was also present in clonally unrelated *E. faecium* and *E. faecalis* isolates and species of *Streptococcus* (*S. agalactiae* and *S. mitis*) as reported (103). It is of note that most Tn*5801* type A1 showing particular mutations were associated with specific geographical locations and/or from specific clonal backgrounds (Figure 2, Table S2 in Appendix 1 and S3; see below). As example, most *S. aureus* isolates (83%, n=88/106) carrying Tn*5801* A1 belonged to CC8 (ST8, ST507, ST609; many of them recovered in the USA) and only a few isolates were associated with other clonal lineages (ST5 and ST664).

Variant A4 contained a 3285 bp insertion within the *tet*(M) gene that was flanked by direct and inverted repeats that comprised *dfrG* (coding a dihydrofolate reductase) and two hypothetical protein sequences. Such element confers resistance to trimethoprim and it was also detected within the *orf*14 of Tn*5801* types A11 and A12 (see below) and the Tn*916*-like transposon (Tn*6198*) of *Listeria monocytogenes*, the plasmid pMG1 from *E. faecium* and also in *Streptococcus pyogenes* isolates from India (113–115). The backbone of A4 was identical to that of Tn*5801* types A5 and A6, which also showed additional insertion sequences (IS) interrupting the *uvrD* (*orf*3) gene (IS*30* or IS*256*, respectively). All A4, A5 and A6 variants were only found among *E. faecium* ST18 isolates. Interestingly, the core sequences of A4, A5 and A6 variants clustered with type A1 sequences from human isolates of *E. faecium* and *S. aureus* of Japan and different European countries, and shared specific mutations (G2336A_*orf*17, A3918T_*orf*16; A6574T_*orf*15; T13303C_*orf*6) suggesting a common origin for them (Figure 2; Table S3).

Types A2, A7, A8, A13, A14 and some A1 sub-variants also clustered together (Figure 2). Types A7 and A8 from enterococci and A2 from *L. garviae* were highly similar according to the backbone consensus sequence (only 1 nt of difference, G2644A_*orf*17). Types A2 and A7 (and also B18, an "A+B" Tn*5801* hybrid, see below) shared a 125 bp deletion in *orf*19 (*orf*12$_{\text{Tn}916\text{-like}}$) that also appeared in the Tn*5801*-like element originally called CW459*tet*(M) of *C. perfringens*

(100). In A7, such deletion is upstream of *tet*(M) and next to an IS*1380*/IS*1678* insertion. A7 also showed a mosaic *tet*(M) gene with Tn*5397 tet*(M) (GeneBank NG_034213.1). Variant A8 identified in a genome of *E. villorum* had the *orf*1 (identified as IS*Lgar5*/ IS*Efm2* belonging to the IS*256* family) truncated by a fragment comprising two *orfs* identified as IS*Efa8* (IS*3* family). Variants A13 (ST5) and A14 (ST247) from *S. aureus* isolates exhibited insertions of partial and complete copies of IS*4* family, and IS*Efa8* (IS3 family) sequences in *orf*14 and *orf*1 (variant A13) or an insertion with an integrase core domain within *orf*14 (variant A14). A2, A7, A8, A13, A14 and A1 shared 5 mutations, namely G1457A and T1804C in *orf*18; G3043C_*orf*17; G70008T_*orf*15; C13758T_*orf*4. Some isolates collected in common geographical areas also had specific additional amino acid changes (Table S3).

Variants A10, A11, and A12 clustered together (Figure 2) and contained insertions within *orf*13 and *orf*14 at different positions, which includes an ATP binding protein and a hypothetical protein within *orf*13 (A10, A11 and A12) or the cassette comprising *dfrG* mentioned above in *orf*14 (A11 and A12). A12 also had a 107 bp deletion in the integrase. They were mostly observed in *S. aureus* belonging to ST239 (84%, n=26/31) and shared four nucleotide changes mutations (T7199C_*orf*15, G7884A_*orf*14, T9310A_*orf*11, C12423T_*orf*7) (Table S3). Specific mutations were detected among isolates from specific geographical areas (variant A10 from China had an A4479G_*orf*16 which was absent in A10 from the USA or Australia).

Variant A3, detected in a few isolates of *E. faecium* (ST50) collected in France during the 1990s, clustered separately and showed a 81 bp deletion in *orf*17 (Figure 2). Variant A9, previously described as Tn*6086* (GeneBank accession no. HM636636.1), was a chimera of Tn*5397* (region *tet*(M)-*orf*21-*orf*22) and Tn*5801* and also had an A/T mutation at the 985 nt of *uvr*D (*orf*3) gene that resulted in a stop codon, and lack of *orf*1. It is of note that the A9 platform detected in *E. faecium* TC6 clustered with other A1 variants from *S. mitis* and *E. faecalis* isolated in Germany and Denmark, respectively, all sharing three distinct mutations (T3046A, T3586A and C4861T). In fact, *E. faecium* TC 6 is an *E. faecium* D344SRF (ST25) transconjugant from C68 strain that contains the same A9 variant (116), thus reflecting transferability of the element. Independent recombination events between Tn*5801* and Tn*5397* seem to have occurred in A9 and A7.

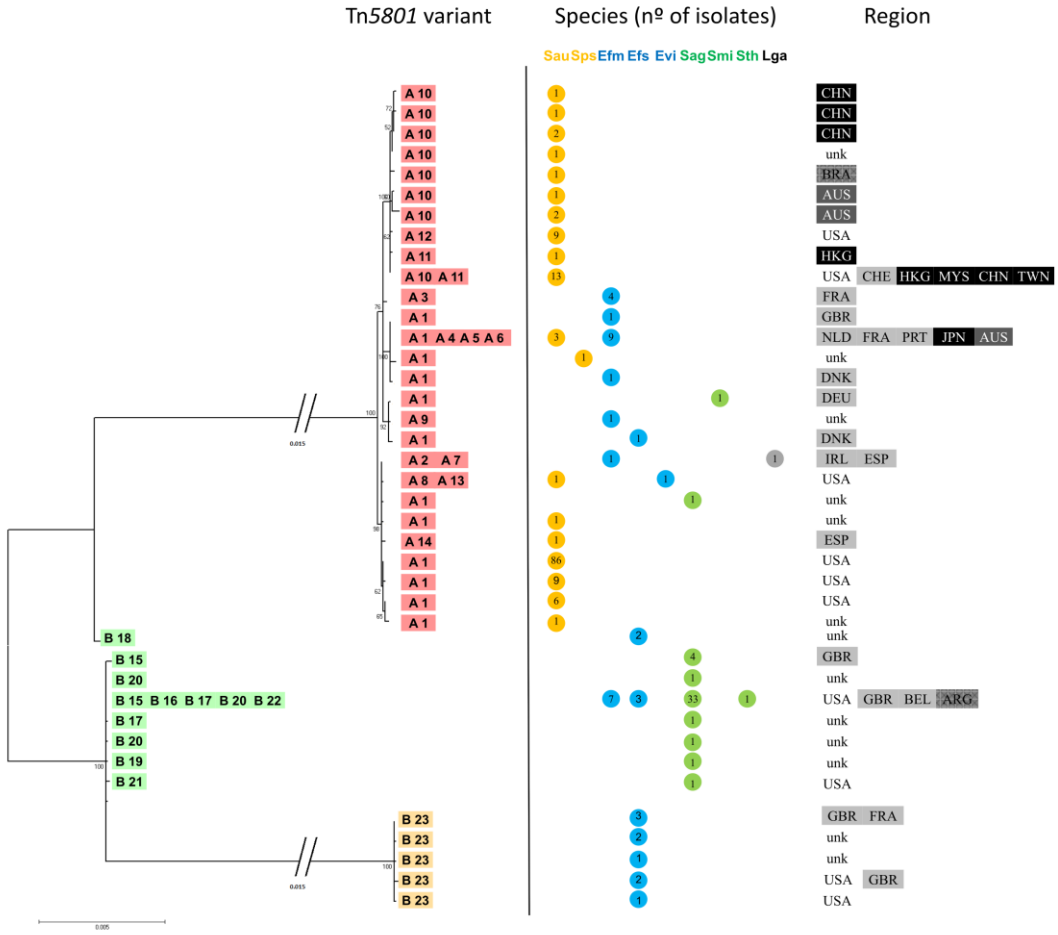


Figure 2. Phylogenetic analysis of Tn5801 variants in Firmicutes. A maximum-likelihood phylogenetic tree is represented, with 1,000 bootstrap replications. The orf26-19 fragment was excluded from this analysis because of its variability (resulting from both mutation and recombination events). The distribution of Tn*5801* variants by species and geographical regions is also represented. Countries within each continent are represented by different tones of gray, black and white. Sau, *Staphylococcus aureus*; Sps, *Staphylococcus pseudintermedius*; Efm, *Enterococcus faecium;* Efs, *Enterococcus faecalis;* Evi, *Enterococcus villorum*; Sag, *Streptococcus agalactiae*, Smi, *Streptococcus mitis*, Sth, *Streptococcus thoraltensis*; Lga, *Lactococcus garvieae*; ARG, Argentina; AUS, Australia; BEL, Belgium; BRA, Brazil; CHE, Switzerland; CHN, China; DEU, Germany; DNK, Denmark; FRA, France; GBR, Great Britain; HKG, Hong Kong; IRL, Ireland; JPN, Japan; MYS, Malaysia; NLD, The Netherlands; PRT, Portugal; ESP, Spain; TWN, Taiwan; USA, United States; unk, unknown.

b) Tn*5801* "group B". This group comprises nine Tn*5801* elements identified among available sequences of streptococci and enterococci (B15-B23) (Figure 1, Table 3, Table S2 in Appendix 1). They split in three main clusters represented by platforms B15-B22 (mainly in streptococci), B18 and B23 (both in enterococci).

The B15 variant (also designed as Tn*5801.Sag*) (104) was considered the paradigm of Tn*5801* type and was detected among *E. faecalis* (this study), *E. faecium*, and *S. agalactiae*. Some type B Tn*5801* elements shared a deletion of 81 bp in *orf*17 (B17, B21, and B22 from *S. agalactiae*; also

detected in A3 from *E. faecium*) or an IS*256*-like insertion upstream of *orf*13 (B20, B21 and B22). The B19 sequence had this IS*256*-insertion followed by an integrase core domain and an IS*3*-family element. The B22 variant had an insertion of a 5,227 bp between *orf*7 and *orf*8 which comprises a phage-specific recombinase, a replication protein and a *lsa*(C) gene encoding resistance to lincosamide, streptogramin A and pleuromutilins (117). Such insertion is identical to other found in *S. agalactiae* UCN70 (GenBank accession number HM990671.1) and similar to that of a Tn*916*-like element from *S. mitis* lacks although lacking the *lsa*(C) gene (102, 118).

Type B23 has a region of five *orfs* (a transcriptional regulator, an histidine kinase, a bacitracin ABC transporter, a bacitracin transport permease and an ABC transporter permease-like) instead of *orf*19-*tet*(M)-*orf*21. The sequence between *orf*26 and *orf*16 only shared a 92% similarity with the other type B sequences. This B23 variant was restricted to CC8 (ST8, ST64, ST90) *E. faecalis* isolates, one of the most ancient *E. faecalis* lineages (14). It included the oldest TETR strain described to date (an ST90 isolate of 1954, part of the historical collection of the Laboratory of Streptococcal Diseases at the National Institute of Allergy and Infectious Diseases, in the USA), and ST8 related isolates of an emblematic strain widely used as recipient (strains FA2 and JH2-2) which predate 1973 (14). The B18 variant is a mosaic of type A [region between *orf*26 and *orf*15] and type B platforms (integrase, *orf*4 to part of *orf*15) and exhibits the 125 bp deletion in *orf*19 also observed in type A variants (A2 and A7). The integrases of B18 and B23 share three unique nucleotide changes. Finally, variant B16, detected in a *S. thoraltensis* isolate (GeneBank accession no. ARCI01000001.1), revealed a recombination event that resulted in the replacement of the *orf*19-*orf*22 region of Tn*5801* for that of Tn*916*.

Epidemiological background of Tn*5801* among enterococci

The int_{Tn5801} was detected in 10% of the enterococcal isolates analyzed (n=61/610, 41 *E. faecium*, 13 *E. faecalis* and 7 *Enterococcus* spp) (Table 3). A high proportion of the int_{Tn5801} positive isolates (84%, n=51/61) were resistant to tetracycline and most of them carried *tet*(M), *tet*(L) and/or *tet*(S) (77%, n=47/61; 38%, n=23/61; 2%, n=1/61, respectively). Specific transposon sequences as int/xis_{Tn916} (16%; n=10/61); $resolvase_{Tn5397}$ (7%, n=4/61), or int/xis_{Tn6000} (2%; n=1/61) were also observed. The Tn*5801* platforms identified in isolates of this collection correspond to A1, A4 and B15 variants. Epidemiological features of int_{Tn5801} positive isolates of *E. faecium* and *E. faecalis* are shown in Table 3.

Table 3. Epidemiological features of the Tn*5801*-like ICEs carrying enterococci isolates

Species	Tn*5801* platform[1]	MLST (BAPS)[2]	Nº isolates	Date	Source[2]	Country[2]	TET^R phenotype[2]	TET^R genes[2]	Other Tns[2]
E. faecalis	B15 (n=2)	ST9/ CC9	3	2001	HP	ESP, ARG	+	*tet*(M) (n=3), *tet*(L) (n=3)	-
	A1	ST30/ CC30	1	1997	HP	AUS	+	*tet*(M)	-
	B15 (n=1)	ST318/ CC9	2	ND	HP	BRA	+	*tet*(M)	-
	B15	ST55/ CC55	1	2001	HV	PRT	+	*tet*(M)	Tn*916*, Tn*6000*
	A1	ST445	1	2001	SP	PRT	+	*tet*(M), *tet*(L), *tet*(S)	-
	B15 (n=3)	ND/ ND	5	2001/ 2002	HP (n=2); HV (n=2); SW (n=1)	BRA (n=2); PRT (n=3)	+	*tet*(M)	Tn*916* (n=3)
E. faecium	+	ST17/ CC17 (3.32)	7	1998-2005	HP (n=5); SW (n=2)	ESP (n=3); AUS (n=1); POL (n=1); PRT (n=2)	+ (n=6); - (n=1)	*tet*(M) (n=5); *tet*(L) (n=3)	-
	+	ST16/ CC17 (3.32)	2	1999	HP	AUS (n=2)	+	*tet*(M)	-
	A1 (n=1); A4 (n=2)	ST18/ CC17 (3.31)	9	1996-2005	HP (n=7); SW (n=2)	PRT (n=8); SER (n=1)	+	*tet*(M) (n=7), *tet*(L) (n=7)	Tn*916* (n=2); Tn5397 (n=2)

Species	Tn*5801* platform[1]	MLST (BAPS)[2]	Nº isolates	Date	Source[2]	Country[2]	TET^R phenotype[2]	TET^R genes[2]	Other Tns[2]
	+	ST64/ CC17 (3.31)	4	2003-2004	HP	CHI	+	*tet*(M) (n=3)	-
	+	ST173/ CC17 (3.31)	1	2000	HP	AUS	+	*tet*(M), *tet*(L)	-
	+	ST80/ CC17 (2.1a)	1	2004	HP	HUN	+	-	-
	+	ST132/CC17 (3.31)	1	2001	HP	PRT	+	*tet*(M), *tet*(L)	Tn*5397*
	A1	ST182/ CC17 (7)	1	1992	HP	USA	+	*tet*(M)	-
	+	ST202/ CC17 (3.32)	1	2005	HP	POL	+	*tet*(M), *tet*(L)	-
	A1	ST50/ CC9 (2.1a)	1	NA	HP	BRA	+	*tet*(M)	-
	+	ST366 (5)	1	2000	HP	PRT	-	*tet*(M)	Tn*916*
	+	ST368 (3.31)	1	2001	SW	PRT	-	-	
	+	ND/ ND	11	2001-2002	HP (n=4); HV (n=1); SW (n=3); P (n=2); SP (n=1)	PRT (n=9); ESP (n=1); TUR (n=1)	+ (n=8); - (n=3)	*tet*(M) (n=8), *tet*(L) (n=5)	Tn*916* (n=2); Tn*5397* (n=1)

Species	**Tn*5801* platform[1]**	**MLST (BAPS)[2]**	**Nº isolates**	**Date**	**Source[2]**	**Country[2]**	**TET^R phenotype[2]**	**TET^R genes[2]**	**Other Tns[2]**
Enterococcus spp	+	ND/ ND	7	2001-2006	HV (n=4); SW (n=2); SP (n=1)	PRT	+ (n=3); - (n=4)	*tet*(M) (n=4), *tet*(L) (n=1)	Tn*916* (n=1)

1 The "+" indicates a Tn5801-like platform which has not been characterized in the corresponding type. 2Abbreviations: ST, Sequence Type; BAPS, Bayesian Analysis of Population Structure; HV, healthy volunteer; HP, hospital patient; SW, sewage; P, poultry; SP, swine/piggeries; ARG, Argentina; AUS, Australia; BRA, Brasil; CHI, China; HUN, Hungary; POL, Poland; PRT, Portugal; ESP, Spain; SRB, Serbia; TUR, Turkey; USA, United States of America; ND, not determined; TETR-tetracycline resistance; Tn, transposons.

Location of Tn*5801*

All Tn*5801* platforms identified in enterococci in this study were chromosomally located at the 3´-end of the *guaA*, as previously observed for three emblematic Tn*5801* elements of staphylococci, *Clostridium* and *S. agalactiae* (13, 28). In the available sequences of the Tn*5801* variants were detected the 11-bp sequence described previously as direct repeats (DRs) situated within the 3'-end of the *guaA* (97). All *guaA* sequences within the same species had high similarity (99-100%), which was somewhat lower (70-80%) when comparing isolates from different species.

Tn*5801* transfer in *Enterococcus*

The ability of Tn*5801* to be transferred was studied in mating assays using enterococcal isolates carrying the Tn*5801* variants A1, A4 and B15 as donors. Only Tn*5801* variant B15 was successfully transferred by conjugation to *E. faecalis* strains JH2-2 and OG1RF. Conjugation frequency was similar for the two receptors despite of the presence of B23 variant in the genome of the JH2-2 recipient. Hybridization of *Sma*I digested DNA from transconjugants with specific probes for *guaA*, int_{Tn5801} and *tet(M)* allowed inferring the location and the size of the transferred region (Figure S3). A single band of 150kb that hybridized with the three probes tested was observed for all *E. faecalis* OG1RF transconjugants obtained in primary mating experiments. For *E. faecalis* JH2-2 transconjugants, the size of the bands that hybridized with the probes was 200 and 250kb (the estimated size of the transferred region was of 20kb for OG1RF and of 50 or 100kb for the two JH2-2 transconjugants, respectively). Mating assays using these *E. faecalis* JH2-2::Tn*5801*_1 (50kb) and *E. faecalis* JH2-2::Tn*5801*_2 (100kb) transconjugants as donors and *E. faecalis* OG1SS as recipient yield *E. faecalis* OG1SS::Tn*5801* transconjugants with one or two copies of a ca. 20kb transposable element. All transconjugants that contain two Tn*5801* copies showed positive hybridization of two bands with int_{Tn5801} and *tetM* genes, but only one band hybridized with the *guaA* probe (Figure S3). Moreover, a PCR-mapping assay demonstrated the presence of a B15 Tn*5801*-like element in all primary and secondary transconjugants while amplification of specific B23 fragments was only observed for one primary and one secondary transconjugants *Efs* JH2-2::Tn*5801*_2 and *Efs* OG1SS::Tn*5801*_2.3 respectively (Figure S3).

Discussion

This study represents the first description of Tn*5801*-like elements in enterococci and one of the first works addressing the evolutionary changes of antibiotic resistance transposons throughout decades. To date, only scarce Tn*5801*-like backbones of a few number of *S. aureus* and *Streptococcus* spp. strains have been analyzed in separate studies (89, 99, 102–104) which did not identifying the clonal context of the isolates, thus limiting the understanding of the evolvability of this element.

Sequence analysis of type A and B variants revealed a common Tn*5801* backbone compatible with a *S. agalactiae* origin on the basis of the GC content (36%) and suggested the further acquisition of a 5kb extra fragment by staphylococci (GC=32%). Such acquisition might have occurred through an IS*256*-family insertion event (the *orf*1 of the "group A" platforms) and would have resulted in the emergence of type A variants.

The apparent confinement of Tn*5801* types A or B in particular clonal backgrounds is remarkable. A recent study suggests that the introduction and massive use of tetracycline since 1950s could have contributed to a positive selection of neonatal hyper-virulent clonal lineages (CC17, CC23 or CC10) of *S. agalactiae* carrying *tet*(M)-CTn, as Tn*5801*, which we designated here as Tn*5801* B type platforms (87). We were also able to detect the Tn*5801* in predominant antibiotic resistant clonal lineages associated with human infections as *S. aureus* (CC8, CC5, ST239) and *E. faecium* (ST17, ST18) from the 90s, or early lineages of *E. faecalis* (CC8). However, our data and available knowledge imposes a carefully interpretation of the relationship between the emergence and persistence of antibiotic resistance and selective processes of particular clones due to antimicrobial treatments (119). In this respect, we should take into account therapeutic strategies and relevant bacterial features that could have facilitated the selective acquisition of certain elements by particular clones. For example, anti-Restriction Modification (RM) occurring in Tn*916*-like elements, often influences selective horizontal gene transfer between staphylococcal populations and other species of Firmicutes (120, 121). Putative anti-RM detected in Tn*5801* (*orf*12) acts against staphylococcal Type I RM systems which are present in *S. aureus* CC5, CC8 (the backgrounds in which Tn*5801* seems to have initially been detected in the late 1950s) (89, 99) and ST239 (an emergent hybrid lineage of ST8 and ST30) (94, 101, 122).

Deep sequence analysis is a powerful tool to accurately establish transmission of pathogens (123). In this study, the comprehensive analysis of Tn*5801*-variants identified some types

associated with sympatric populations of major clonal lineages (as reflected by the presence of specific nucleotide changes among isolates of particular geographical areas). This seems clear for *S. aureus* ST239 (that carries highly related A10, A11 and A12 variants) and ST8/ST609 human isolates from the USA recovered from 1994 to date (carrying A1) for which, differences in the genome sequences of regional endemic clones have also been reported (101, 121). Besides clonal spread of strains carrying Tn*5801,* the presence of some variants in isolates of different species recovered in different countries and, eventually, different hosts indicates that horizontal gene transfer has also contributed to the current occurrence of Tn*5801*. This study demonstrates the transference of Tn*5801* between different *E. faecalis* backgrounds, even if the strain already harbored a Tn*5801* element. Although the *guaA*-associated islands theoretically integrate site specifically, tyrosine recombinases can be functional at secondary integration sites (95, 96). This would explain the insertion of elements at different genomic locations (transconjugants *Efs* OG1SS::Tn*5801*_1.2, OG1SS::Tn*5801*_2.2, OG1SS::Tn*5801*_2.3). Moreover, integration at the same site cannot be discarded if a consensus hotspot remains available for new integration events after the first Tn*5801* acquisition. For example, two *E. faecalis* genomes (*E. faecalis* Com1 and *E. faecalis* Com2; GeneBank accession no. AJES01000002.1 and AJBL01000002.1, respectively) harbored a composite structure that comprises Tn*5801* B15 and a putative transposable element (GC% of 27,3%, compatible with a clostridial origin (124)), both located upstream the *guaA* gene. The clostridial element seems to have been inserted at the 11-bp hotspot of a *guaA-int*$_{\text{Tn}5801}$, region. Finally, recombination cannot be discarded for some cases as described for other elements in *E. faecalis* (125). Such events increase the number (if they include different genes) or the level of resistance (if the gene dosage is increased as we observed here).

Although the classical JH2-2 lab strain was serendipically useful for this work as allowed detecting the acquisition of more than one Tn*5801* element by a single strain, the presence of conjugative genetic elements that can interfere in the transferability of other elements under study open the discussion about the suitability of certain classic laboratory strains to be used in transfer experiments.

CTn-*tet*(M) elements with the same genetic backbone may carry different determinants associated with resistance (e.g. to bacitracin, lincosamides/streptogramines, and trimethoprim in our study) and/or ISs (IS*3*, IS*4*, IS*30*, IS*256*, IS*1380* families) which are frequently designated with different numbers in other studies (92, 104, 126). However, all these variants reflect the dynamics of a basic particular element circulating among bacterial communities. In this study,

ISs with intact IR were observed within different genes (*orf*1, *orf*3, *orf*13, *orf*19) of all Tn*5801* type A of *S. aureus* and *E. faecium*, suggesting their recent acquisition. Among type B, only the mutator family IS*256* was consistently detected upstream of the *orf*13 (types B20, B18, B19), stability that persisted even when some particular variants (e.g., B20) that were disseminated across genera. Some ISs from the IS*256* family have been implied as modulators of gene expression, suggesting that future studies might reveal the role of this IS in the function of type B Tn*5801*-like elements (127–130).

In summary, this study shows that Tn*5801*-like variants are present among Firmicutes since at least the 1950´s, being predominant in microorganisms obtained from human hosts (51). Different waves of expansion in human populations of certain clones in *Staphylococcus*, *Enterococcus*, and *Streptococcus* seems to play an important role in their occurrence and evolvability although transfer events between different clonal backgrounds would also contribute to the spread, demonstrated experimentally here for *E. faecalis*. The detection of genetic rearrangements in the four functional regions highlight once more the plasticity of CTns, potentially influencing micro-evolutionary events eliciting adaptive functions in microbial populations and communities in common or heterogeneous hosts and environments (131–134).

Supplementary material

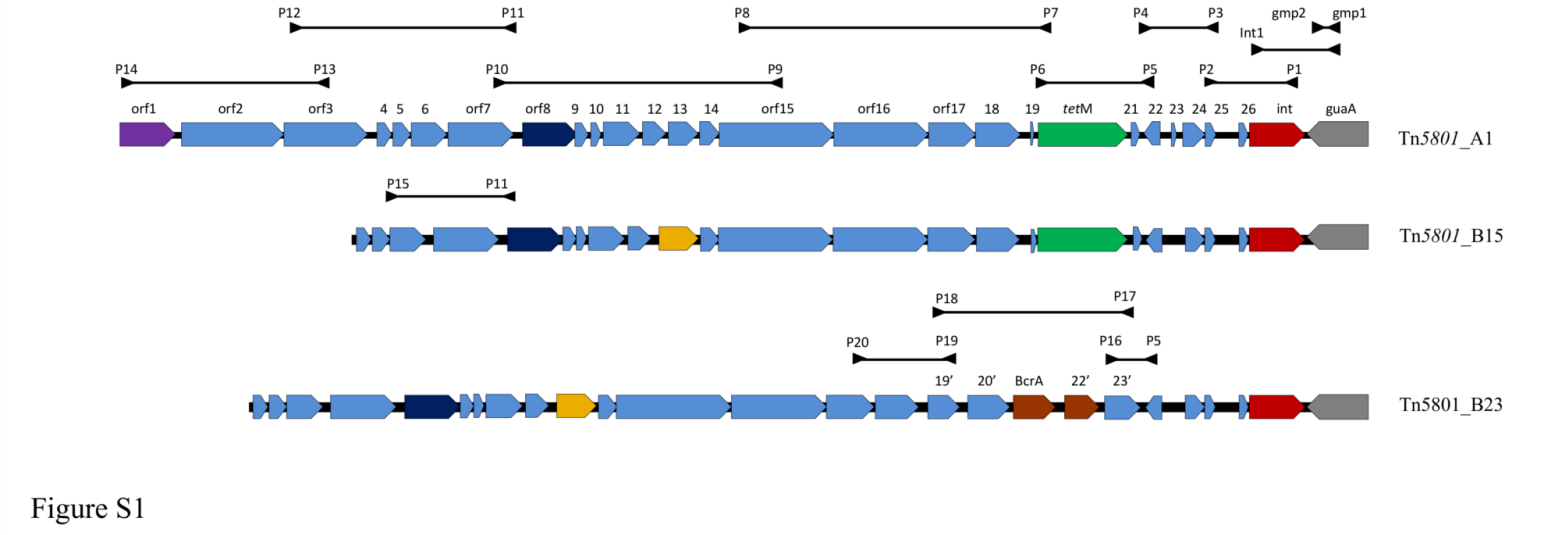


Figure S1

Figure S1. Characterization of Tn*5801*: PCR strategy.

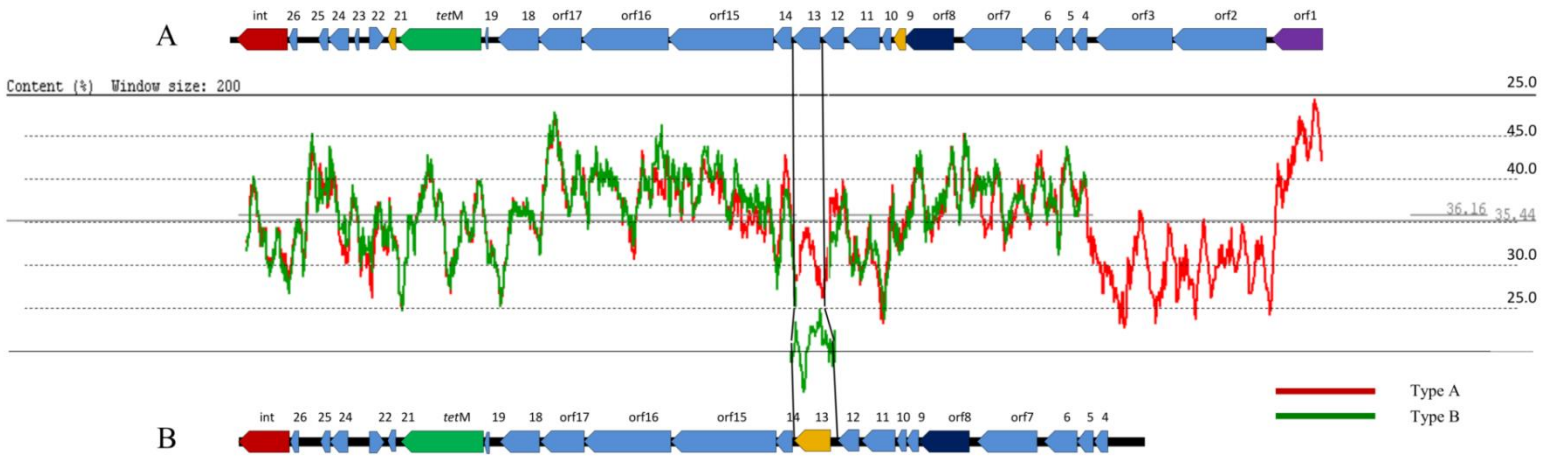


Figure S2

Figure S2. GC% content of Tn*5801*.A1 and B15.

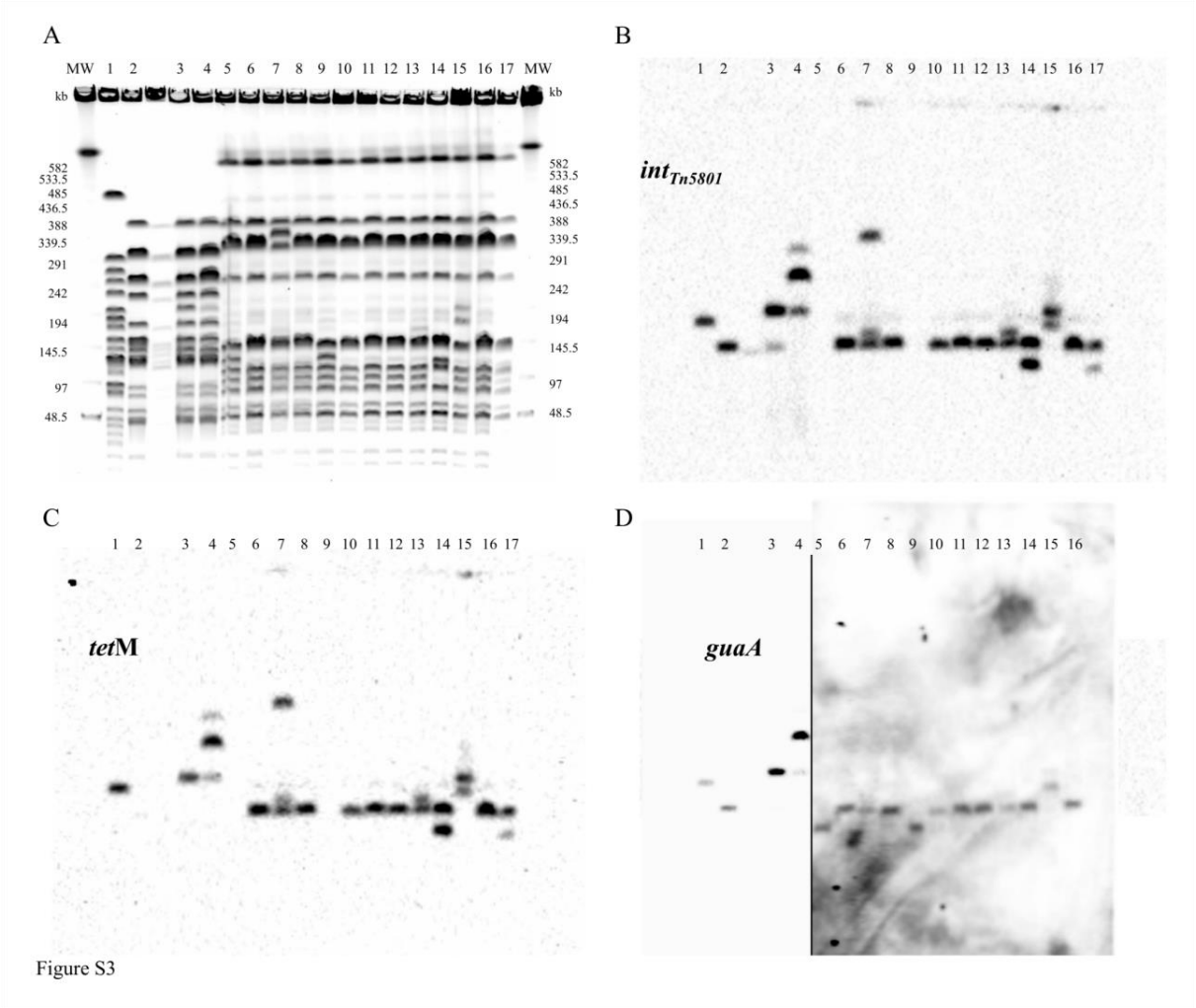
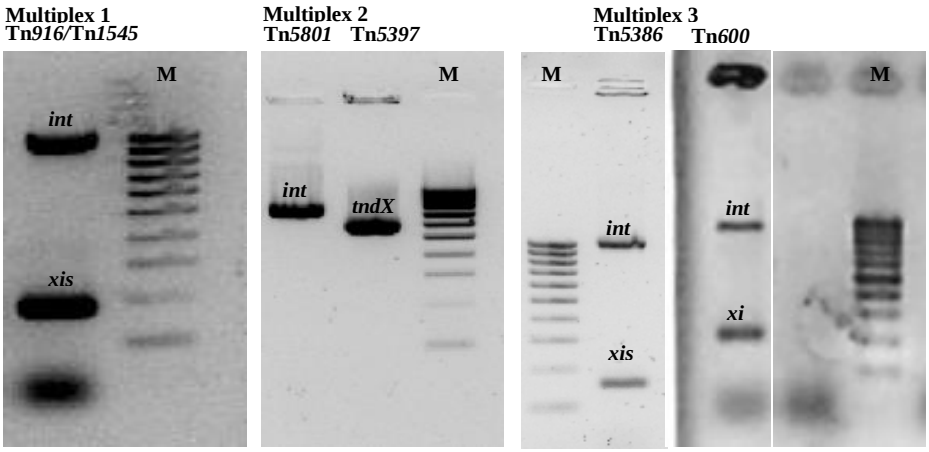


Figure S3

Figure S3. Location of Tn*5801* markers *int*Tn*5801*, *tet*(M) and *guaA* genes. *SmaI*-PFGE of *E. faecalis* donors, recipients and tranconjugant strains (A); hybridization of *SmaI* digested genomic *E. faecalis* DNA with a *int*Tn*5801*, *tet*(M) and *guaA* probes (panels B, C, D, respectively). Primary transconjugants were represented in lines 3, 4, 10, 11 and 12; and secondary transconjugants in lines 6, 7, 8, 13, 14, 15, 16 and 17. MW= molecular weight marker (PFGE-standard lambda ladder, 48.5–1,000 kb; New England Biolabs); lane 1, *Ef1*; lane 2, *Efs* strain JH2-2; lane 3, *Efs* strain JH2-2::Tn*5801*_1; lane 4, *Efs* strain JH2-2::Tn*5801*_2; lane 5, *Efs* strain OG1SS; lane 6, *Efs* strain OG1SS::Tn*5801*_1.1; lane 7, *Efs* strain OG1SS::Tn*5801*_1.2; lane 8, *Efs* strain OG1SS::Tn*5801*_1.3; lane 9, *Efs* strain OG1RF; lane 10, *Efs* strain OG1RF::Tn*5801*_1; lane 11, *Efs* strain OG1RF::Tn*5801*_2; lane 12, *Efs* strain OG1RF::Tn*5801*_3; lane 13, *Efs* strain OG1SS::Tn*5801*_2.1; lane 14, *Efs* strain OG1SS::Tn*5801*_2.2; lane 15, *Efs* strain OG1SS::Tn*5801*_2.3; lane 16, *Efs* strain OG1SS::Tn*5801*_2.4; lane 17, *Efs* strain OG1SS::Tn*5801*_2.5. Panel D comprises two different hybridization experiments. strain OG1SS::Tn*5801*_2.3; lane 16, *Efs* strain OG1SS::Tn*5801*_2.4; lane 17, *Efs* strain OG1SS::Tn*5801*_2.5. Panel D comprises two different hybridization experiments.

Table S1. PCR typing method for the detection of diverse Tn*916*-like conjugative transposons



PCR reaction	Primers	Sequence (5'-3')	Amplification conditions	Amplicon size (bp)	GenBank accession No.	Control isolates
Multiplex 1 Tn*916/1545*	Int_{916}-F	GCGTGATTGTATCTCACT	95ºC- 10'- 1 cycle; 94ºC-30'', 52ºC-30'',72ºC-30''-25 cycles; 72ºC-10'- 1 cycle	1046	U09422.1	*E. faecium* H638466 (this study)
	Int_{916}-R	GACGCTCCTGTTGCTTCT				
	Xis_{916}-F	AAGCAGACTGACATTCCTA		194		
	Xis_{916}-R	GCGTCCAATGTATCTATAA				
Multiplex 2 Tn*5397*	tndX-F	ATGATGGGTTGGACAAAGA	95ºC- 10'- 1 cycle; 94ºC-30'', 55ºC-30'',72ºC-45''-25 cycles;72ºC-10'- 1 cycle	611	AF333235	*E. faecium* H446511 (this study)
	tndX-R	CTTTGCTCGATAGGCTCTA				
Tn*5801*	Int_{5801}-F	CTGTTTCCGATATTGAGC		857	AF329848.1	*E. faecium* E240 (this study; KP001176.1)
	Int_{5801}-R	GTTTCGCAAGTAGTCTACAG				
Multiplex 3 Tn*6000*	Int_{6000}-F	CATCGAGTCTAACCGATTGT	95ºC- 10'- 1 cycle; 94ºC-30'', 55ºC-30'',72ºC-30''-25 cycles; 72ºC-10'- 1 cycle	869	FN555436.1	*E. faecalis* C386 (JN208881.1)
	Int_{6000}-R	GACCCAAACGAACTTGTACT				
	Xis_{6000}-F	CGAAGTATTAACCGAACAGA		203		
	Xis_{6000}-R	TATCATCGCGATCAAAACGA				
Tn*5386*	Int $_{5386}$ -F	CTTGTTCCTACGGACAGAGT		1021	DQ321786.1	*E. faecium* DR344R[25]
	Int_{5386}-R	AGCCGTGAGCGTAATAATTC				
	Xis_{5386}-F	ATACTGATGTGCCTGTATGG		155		
	Xis_{5386}-R	TCGCTTTATCTGAATACGG				

Abbreviations: M, DNA marker- 100bp (Hypperladder IV 100bp-BioRon); *Int*, integrase; *xis*, excisionase, *tndX*, recombinase site-specific resolvase.

Table S3. Identification of specific nucleotide changes in each sequences with unique pattern, following the order of the phylogenetic consensus tree (Figure 2).**

Variant	Species	Country (year)	Int_{Tn5801}	orf18	orf17	orf16	orf15	orf14	orf12	orf11	orf10	orf9	orf8	orf7	orf6	orf5	orf4
Type A																	
A 10	**Sau** (1)	CHN	T151C	-	-	A4479G	A5611G, T7199C	G7884A	-	T9310A	-	-	-	C12423T	-	-	-
A 10	**Sau** (1)	CHN	-	-	T2452G	A4479G	A5611G, T7199C	G7884A	-	T9310A	-	-	-	C12423T	-	-	-
A 10	**Sau** (2)	CHN	-	-	-	A4479G	A5611G, T7199C	G7884A	-	T9310A	-	-	-	C12423T	-	-	-
A 10	**Sau** (1)	unk	-	-	-	A4479G	T7199C	G7884A	-	D T8908, T9310A	-	-	-	C12423T	-	-	-
A 10	**Sau** (1)	BRA	-	-	-	-	T7199C	G7884A	-	T9310A	-	-	C10577A, T10723C	C12423T	-	-	-
A 10	**Sau** (1)	AUS	-	-	-	-	T7199C	G7884A	-	T9310A	-	-	C10032T	G11744A, C12423T	-	-	-
A 10	**Sau** (2)	AUS	T1085G	-	-	-	G6800A, T7199C	G7884A	-	T9310A	-	-	C10032T	C12066A, C12423T	G13254A	-	-
A 12	**Sau** (9)	USA	D381-478, A484C	-	-	-	D 5541-5549, C7199T	G7884A	-	T9310A	-	-	-	C12423T	-	-	-
A 11	**Sau** (1)	HKG	-	-	-	-	D 5541-5549, T7199C	G7884A	-	T9050C, T9310A	-	-	-	C12423T	-	-	-
A 10, A 11	**Sau** (13)	USA (A10), CHE (A10), HKG (A10), MYS (A11), CHN (A11), TWN (A11)	-	-	-	-	T7199C	G7884A	-	T9310A	-	-	-	C12423T	-	-	-
A 3	**Efm** (4)	FRA	-	-	D2237-2319	G3359A	-	-	-	-	-	-	T11131C	-	-	-	-
A 1	**Efm** (1)	GBR	-	-	G2636A	A3918T	A6574T	-	-	-	-	-	-	-	T13303C	-	-
A 1, A 4, A 5, A 6	**Sau** (3), **Efm** (9)	NLD (A4), FRA (A4, A6), PRT (A4), JPN (A1), AUS (A4, A5)	-	-	G2636A	A3918T	A6574T	-	-	-	-	-	-	-	T13303C	-	-
A 1	**Sps** (1)	unk	-	-	G2636A	A3918T	A6574T, CC6852TT	-	-	-	-	-	-	-	T13303C	-	A13764G
A 1	**Efm** (1)	DNK	-	-	T2529C, G2636A	A3918T	A6574T	-	-	-	-	-	-	-	T13303C	-	-
A 1	**Smi** (1)	DEU	-	-	G3046A	G3419A, T3586A, C4861T	-	-	-	-	-	-	-	-	T13303C	-	-
A 9	**Efm** (1)	unk	-	-	G3046A	T3586A, C4861T	-	-	-	-	-	-	G10586T	-	-	-	-
A 1	**Efs** (1)	DNK	-	-	G3046A	T3586A, C4861T	-	G7914A	-	-	-	-	-	G11506A	-	-	-
A 2, A 7	**Efm** (1), **Lga** (1)	IRL (A7), ESP (A2)	-	G1457A, T1804C	G3043C	-	G7008T	-	-	-	-	-	-	-	-	-	C13758T
A 8, A 13	**Sau** (1), **Evi** (1)	USA (A13), (A8) (unk)	-	G1457A, T1804C	G2644A, G3043C	-	G7008T	-	-	-	-	-	-	-	-	-	C13758T
A 1	**Sag** (1)	unk	-	G1457A, T1804C	A2928G, G3043C	IN: A4197	G7008T	-	-	-	-	-	-	-	-	-	C13758T

A 1	**Sau** (1)	unk	DT1110	G1457A, T1804C	G3043C, T3084A	-	G7008T	-	-	-	-	-	-	-	-	-	C13758T
A 14	**Sau** (1)	ESP	-	G1457A, T1804C	C2463T, G3043C	-	G7008T, IN: T 7374	-	-	-	-	-	-	-	C12895T	-	C13758T
A 1	**Sau** (86)	USA	-	G1457A, T1804C	G3043C	-	A5215T, G7008T	-	-	-	-	-	-	-	-	-	C13758T
A 1	**Sau** (9)	USA	-	G1457A, T1804C, G2085A	G3043C	-	A5215T, G7008T	-	-	-	-	-	-	-	-	-	C13758T
A 1	**Sau** (6)	USA	-	G1457A, T1804C	G3043C	-	A5215T, G7008T	-	G8177A	-	-	-	-	-	-	-	C13758T
A 1	**Sau** (1)	unk	-	G1457A, T1804C	G3043C	G3267T	A5215T, G7008T	-	G8177A	-	-	-	-	C11910T	-	-	C13758T
Type B																	
B 18	**Efs** (2)	unk	A127G, A184G, A192G; C285T*, T402C*	-	A3046G*	A3586T*	T5575C	-	-	-	-	-	-	-	-	-	-
B 15	**Sag** (4)	GBR	-	G1420T	-	-	T6270C	-	D8070-8077, T8510C	-	-	-	-	-	-	-	-
B 20	**Sag** (1)	unk	-	-	-	-	-	-	D8070-8077	-	-	-	-	-	-	-	-
B 15, B 16, B 17, B 20	**Efm** (7), **Efs** (3), **Sag** (33), **Sth** (1)	USA (B20), GBR (B16), BEL (B16), ARG (B15), unk (B17)	-	-	G3107T	-	-	-	D8070-8077	-	-	-	-	-	-	-	-
B 17	**Sag** (1)	unk	-	-	D2237-2319	-	-	-	D8070-8077	-	-	-	-	-	-	G13370A	-
B 20	**Sag** (1)	unk	-	-	-	-	-	-	-	-	-	-	-	DTC11143	-	-	A13905G
B 19	**Sag** (1)	unk	-	-	-	-	-	-	D8070-8077	T8953C	-	-	-	-	-	-	-
B 21	**Sag** (1)	USA	-	A1668C	D2237-2319	-	-	-	D8070-8077	-	-	-	A10355C	-	-	-	-
B 23	**Efs** (3)	GBR, FRA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B 23	**Efs** (2)	unk	-	-	-	-	-	-	-	-	-	-	-	A12056G	-	-	-
B 23	**Efs** (1)	unk	-	-	-	-	-	-	-	-	-	-	-	G12537C	-	-	-
B 23	**Efs** (2)	USA, GBR	-	-	-	-	C5514T	-	-	-	C9645T	-	-	C12525T	-	-	-
B 23	**Efs** (1)	USA	-	-	-	-	-	-	-	-	-	-	G10714T	-	-	-	-

** The position of each change is given from the alignment of the consensus sequences.
D, deletion; IN: insertion;
*These nucleotide changes are related to variants of type A (B18 is a recombinant variant).

Chapter III

A novel conjugative element composed of Tn*5801* and the new transposon ICE*Efs*Asa1: transfer dynamics between *Enterococcus faecalis* populations

The work related with this chapter was published in:

León-Sampedro R; San Millán A, Novais C; Peixe L; Baquero F; Coque TM.
A Novel Conjugative Element Composed of Tn*5801* and a new Transposon: Transfer Dynamics between *Enterococcus faecalis* Populations.
7th FEMS Microbiology Congress 2017, Valencia (Spain).
Abstract poster FEMS7-2531.
(manuscript submitted)

Chapter III: A novel conjugative element composed of Tn*5801* and the new transposon ICE*Efs*Asa1: transfer dynamics between *Enterococcus faecalis* populations

Abstract

Conjugative-transposon (CTn) Tn*5801-tet*(M) conferring resistance to tetracycline is commonly detected in disparate genus of Firmicutes, including Enterococcus. Tn*5801* harbour a site - specific integrase with apparent affinity for the 3´end of the gene encoding the GMP synthase (*guaA*). In the previous chapter, we demonstrated the unexpected transferability and integration of one or more copies of this CTn to different chromosomal sites. We now analyze the genetic context of the Tn*5801* transferred regions to further understand the mobilization events of Tn*5801* and the impact of acquiring this CTn in the bacterial fitness of *Enterococcus faecalis.*

The Tn*5801* elements harboured in donors (Ef1 clinical strain), the *E. faecalis* laboratory recipient strains JH2-2 (with a Tn*5801*-Δ*tet*M), and transconjugants of *E. faecalis* strains JH2-2, OG1RF and OG1SSp obtained in primary and secondary filter mating assays in Chapter 2 were comprehensivelu analyzed by whole genome sequencing (Illumina-HiSeq-2500) and further bioinformatics analysis using a variety of tools.. Growth rate was estimated in a Synergy HTX Plate Reader in the absence and presence of sub-inhibitory tetracycline concentrations (0,5-2mg/L).

We describe the transference of a novel 53,3kb composite element designed as ICE50CE comprising Tn*5801*.B25 and a new 30kb ICE called ICE*Efs*Asa1, responsible of the mobilization of Tn*5801* in *E. faecalis.* Moreover, we demonstrate the ability of Tn*5801* elements to integrate at site-specific sites present in 3´end of the *guaA* gene and other chromosomal locations of *E. faecalis* genomes (7-11 sites). Transconjugants harboring two 50CE copies showed higher values of $MIC_{Tetracycline}$ and V_{max} of the growth curves than those with a single 50CE copy (24-48 vs 96vs1mg/L and 4.28 vs 5.97, respectively). These findings suggest the effect of tetracycline in the transfer of Tn*916*-elements and also the enhanced bacterial fitness of these isolates in tetracycline selective environments. The present work enlarges the list of ICEs/CTns that transfer by conjugation and highlights the diversity of mechanisms enabling their conjugative mobilization, in particular, that resulting from tandem accretion after site-specific

recombination. The strategy could facilitate the persistence and amplification of some pathogenic clones besides the contribution to the evolvability of the species and the impact in the population structure.

Introduction

Tn*5801* is a conjugative transposon that has traditionally been classified within the Tn*916*-family. It differs from other Tn*916* members in the integrase (*int*) and the excisionase (*xis*) sequences, and the insertion site specificity. The Tn*5801*int is a site-specific tyrosine recombinase that recognizes a conserved core sequence at the 3' end of the glutamine amidotransferase (GMP) synthetase housekeeping gene (*guaA*) (REF), a known specific insertion site of genomic islands, transposons, and or bacteriophages in different bacterial species (57, 135, 136).

Tn*5801* variants clusters in two distinct groups, Tn*5801* group A, a long form (ca. 25 kb) predominant in *Staphylococcus*, and Tn*5801* group B (ca. 20 kb) predominant in *Streptococcus*. Both groups are commonly detected in *Enterococcus*, with a remarkable confinement of Tn*5801* group A or B in particular clonal backgrounds (57). In the previous chapter, the transferability of variants Tn*5801* A and B was analyzed by filter mating using different recipient strains of *E. faecalis* (JH2-2, OG1RF, and OG1SSp), all being negative for the presence of the *tet*M gene which is present in almost all Tn*5801* and many Tn*916*-like members (57). During such screening of Tn*5801*, we noticed the presence of a Tn*5801* variant (Tn*5801*. B23 lacking *tet*M gene) inserted in the *guaA* gene of *E. faecalis* JH2-2. When a clinical isolate of *E. faecalis* Ef1 (ST318) harbouring a Tn*5801*.B15 variant was used as a donor, *Efs*. JH2- transconjugants containing either different copies of the transposon or Tn*5801*-like elements of unusual size were recovered. The results were unexpected attending to the known site specificity of Tn*5801*. However, the lack of sequences precluded establishing conclusion about the regions transferred and the events involved in the generation of novel Tn*5801*-elements. Due to the role of enterococci as a hub of antimicrobial transmission (35, 37), the increasing reports about new conjugation strategies (137), and the lack of studies analysing the transfer dynamics of Tn*5801* and other site specific transposons, the aims of this study were to compare the genomes of both donors and transconjugants; and to analyse the impact of acquiring different Tn*5801* variants in the expression of antimicrobial resistance and in the fitness of *E. faecalis*.

Materials and methods

Bacterial strains

In the previous chapter, we suggested different transfer events based on the presence of some gene sequences in donors and transconjugants (Figure 1). The donor, recipient and transconjugants strains included in this study appear in Figure 1 and Table 1. Conjugation and antibiotic susceptibility testing was carried out as previously described (57).

Growth rate in presence of tetracycline assays

To evaluate the effect of tetracycline in the growth rate of all isolates harbouring different combinations of ICEs, a growth rate assay was performed testing different tetracycline concentrations (0, 0.5, 10 and 20 mcg/mL). Transconjugants and the corresponding recipient strain were cultured at 37°C in BHI broth overnight. Grown cultures were diluted 1:1000 into BHI broth to obtain an inoculum of approximately 10^5 CFU/ml. Growth curves were done using a BioTek Synergy HTX plate reader (BioTek Instruments, Potton, UK) using 96-well plates; OD_{600} was measured every 15 min for 22 h at 37 ºC. To guarantee culture optical homogeneity, the plates were shaken for 10 s before all measurements. Three biological and technical replicates were assayed for each strain.

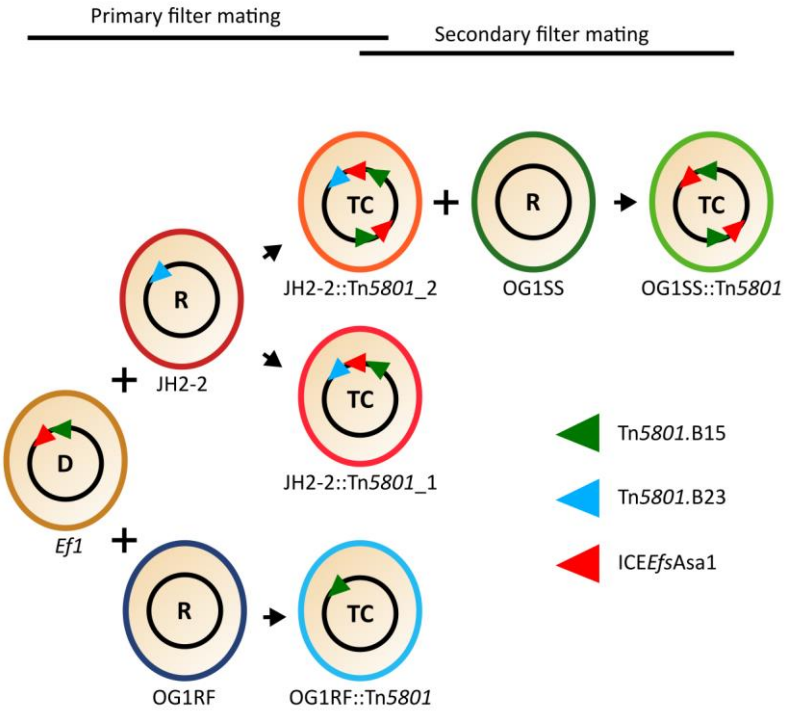


Figure 1. Scheme and strategy follow in the primary and secondary filter mating experiments. The "+" indicate the two members involve in each conjugation (donor and recipient) and the arrows show the result of each conjugation (the transconjugant). In the secondary filter mating, one of the transconjugants (*E. faecalis* JH2-2::Tn*5801*_2) is the donor.

Whole Genome Sequencing (WGS) and bioinformatics

The genomes of the donor, the recipient strains and the four transconjugants selected were sequenced and analyzed. Total DNA from these strains was extracted from 5mL overnight cultures using Wizard® Genomic DNA Purification Kit (Promega) and DNA concentration was measured with Qubit™ Fluorometer and Nanodrop 1000 (Thermo Scientific, Waltham, MA, USA). Genomic DNA was sequenced using the Illumina HiSeq 2500 platform. Sequencing was carried out using a standard 2 x 100 base protocol in a Genome Analyzer IIx (Illumina, San Diego, CA).

The paired-end reads were "de novo" assembled by SPAdes genome assembler (v.3.5.0); and QUAST for further evaluating the quality of the genome assembly (REF). Using an in-house comparison pipeline, the genomes of all transconjugants were compared with their respective recipients and donors to identify and describe the location site and the mobilized elements. Briefly, the contigs obtained by SPAdes were annotated with Prokka v1.12 tool (74). Easyfig v2.2.2 (75) was used for drawing the schemes and comparisons. The database of Clusters of Orthologous Groups of proteins (COGs) was used to classify the known gene functions

(http://www.ncbi.nlm.nih.gov/COG/) (138). The novel transposable elements were compared with available sequences at the NCBI database.

Table 1. Characterization of the Tn*5801* transconjugants with resistance against tetracycline (TETR) and growth rate with different concentrations of the antibiotic.

Tranconjugants	**Donor**	**Recipient**	**Number of Tn*5801* detected**	**ICEs detected**	**TETR µg/mL**	**Transfer rates per donor**	**Max V [Optical Density: 600] and standard deviation**							
							0 ug/mL		0,5 ug/mL		10 ug/mL		20 ug/mL	
JH2-2	**-**	**-**		Tn*5801*.B23	1		5,987	0,127	0,034	0,005	0,048	0,010	0,034	0,001
JH2-2::Tn*5801*_1	*Efs* Ef1*	JH2-2	1	Tn*5801*.B23, ICE*Efs*Asa1-Tn*5801*.	48	1.6x10-8	5,409	0,144	4,048	0,098	0,216	0,052	0,039	0,015
JH2-2::Tn*5801*_2	*Efs* Ef1*	JH2-2	2	Tn*5801*.B23, ICE*Efs*Asa1-Tn*5801* (x2)	96	1.6x10-8	5,316	0,148	3,907	0,174	1,169	0,099	0,932	0,154
OG1-RF	-	-		-	4		6,363	1,123	0,083	0,057	0,043	0,013	0,053	0,025
OG1-RF::Tn*5801*	*Efs* Ef1*	OG1RF	1	Tn*5801*.B15	48	ND	5,674	0,798	5,974	0,604	0,037	0,005	0,076	0,084
OG1-SS	-	-		-	4		6,852	0,654	0,075	0,067	0,069	0,019	0,066	0,019
OG1-SS::Tn*5801*	JH2-2::Tn5801_2	OG1-SS	2	ICE*Efs*Asa1-Tn*5801* (x2)	96	1.6x10-5	5,608	0,606	4,285	0,922	1,402	0,069	1,357	0,192

Results

Transfer dynamics

Comparative genomics and comprehensive bioinformatics analysis allowed characterizing the conjugative elements involved in the transfer events we suggested in chapter 2 (Figure 1). Under the light of these findings, we could confirm and extend previous suggestions.

Briefly, the size of the Tn*5801*.B15 from *E. faecium* Ef1 varied with the recipient strain in primary mating experiments. Using *E. faecalis* OG1RF as recipient, an acquired region of 20 kb corresponding to the entire Tn*5801*.B15 was detected. However, the additional bands of ca. 50 and 100 kb identified when using *E. faecalis* JH2-2 correspond to the insertion of one or two copies of a novel 53,3 kb ICE*Efs*Asa1-Tn*5801* composite element consisting in a tandem of a new ca. 30 kb ICE called here ICE*Efs*Asa1 and a copy of Tn5*801*.B15. One copy of this 53,3kb element was inserted at the 3´site of Tn*5801*.B23 (Δ*tetM*) of the JH2-2 recipient. The precise location of the second copy of the 53,3kb element could not be established due to the limitations imposed by short-length Illumina sequencing approach. Detailed analysis of the genome of the *E. faecalis* Ef1 revealed the presence of both Tn*5801*.B15 and ICE*Efs*asa1 in the donor (Figure 2).

Due to the unexpected presence of a Tn*5801* element in the JH2-2 recipient, we made a secondary filter mating using JH2-2::Tn*5801*-2 as a donor and a recipient lacking Tn*5801* elements in order to verify the transferability of the novel composite elements and to avoid stochastic or recombination events. Two copies of the 53,3kb element were also detected in the *E. faecalis* OG1SS::Tn*5801* transconjugant. In this case, one of the copies of the tandem ICE*Efs*Asa1 and Tn*5801*.B15 was inserted within the *guaA* gene of the donor and the other in a site-specific sequence identical to that of the *guaA* gene (Figure 2).

Detailed analysis of the boundaries of each of these ICEs in the different donor, receptor and transconjugants revealed common sequences in the 3´end of the integrases of the Tn*5801*.B23 and ICEEfs*asa1* and also common sequences in the attachment sites of the hosts which were the 3´end of the *guaA* gene and also other chromosomal locations of donor and recipients (the 11bp located in the *guaA* gene are also detected at 11 positions in the *E. faecalis* V583 genome (3 with 11/11 bp identical, and 8 with 10/11) and at 7 locations in the *E. faecalis* Ef1 11/11 bp identical and 9 sequences with 10/11 bp identical).

The analysis of these ICEs (ICE*Efs*Asa1-Tn*5801*, the tandem of Tn*5801*.B15 and ICE*Efs*Asa1) revealed a chimera generated by site-specific accretion. Each integrase (a tyrosine recombinase XerC protein) seems to catalyze the attachment site (*att*) of the corresponding ICE*Efs*Asa1 with the *att* (*attB*) present at the 3´end of the *guaA* and other chromosomal positions. These recombination events would integrate ICEs and would produce flanking attachment sites *attL* and *attR sites.* Excision would reverse this IntS-recombination toward formation of the *attP* and *attB* sites (Figure 2). Under this landscape, ICE*Efs*Asa1 would have been originally inserted in the *guaA* gene of *E. faecalis* Ef1 isolate and afterwards, the incoming Tn*5801*.B15 could integrate in any of the *att* of the resident ICE, leading to the tandem accretion of the two elements (ICE*Efs*Asa1-Tn*5801*). The process would be the same for the integration of the ICE*Efs*Asa1-Tn*5801* in JH2-2 recipient carrying Tn*5801*.B23 integrated at the 3´end of the *guaA* gene.

All the three elements (ICE*Efs*Asa1, Tn*5801*.B15 and Tn*5801*.B23) possess a FtsK-SpoIIIE conjugation module able to process double-stranded DNA and systematically associated with a replication module. The FtsK protein is associated with cell division and chromosome segregation and interacts with the integrase (XerC protein) activating the chromosome unlinking by recombination (131). The nick relaxase protein is probably involved in the transfer of these ICEs (139).

We demonstrate the transferability of the ICE*Efs*Asa1-Tn*5801* 53.3kb tandem avoiding the presence of the Tn*5801*.B23 variant, and the mobilization of the Tn*5801*.B15 with or without ICE*Efs*Asa1.

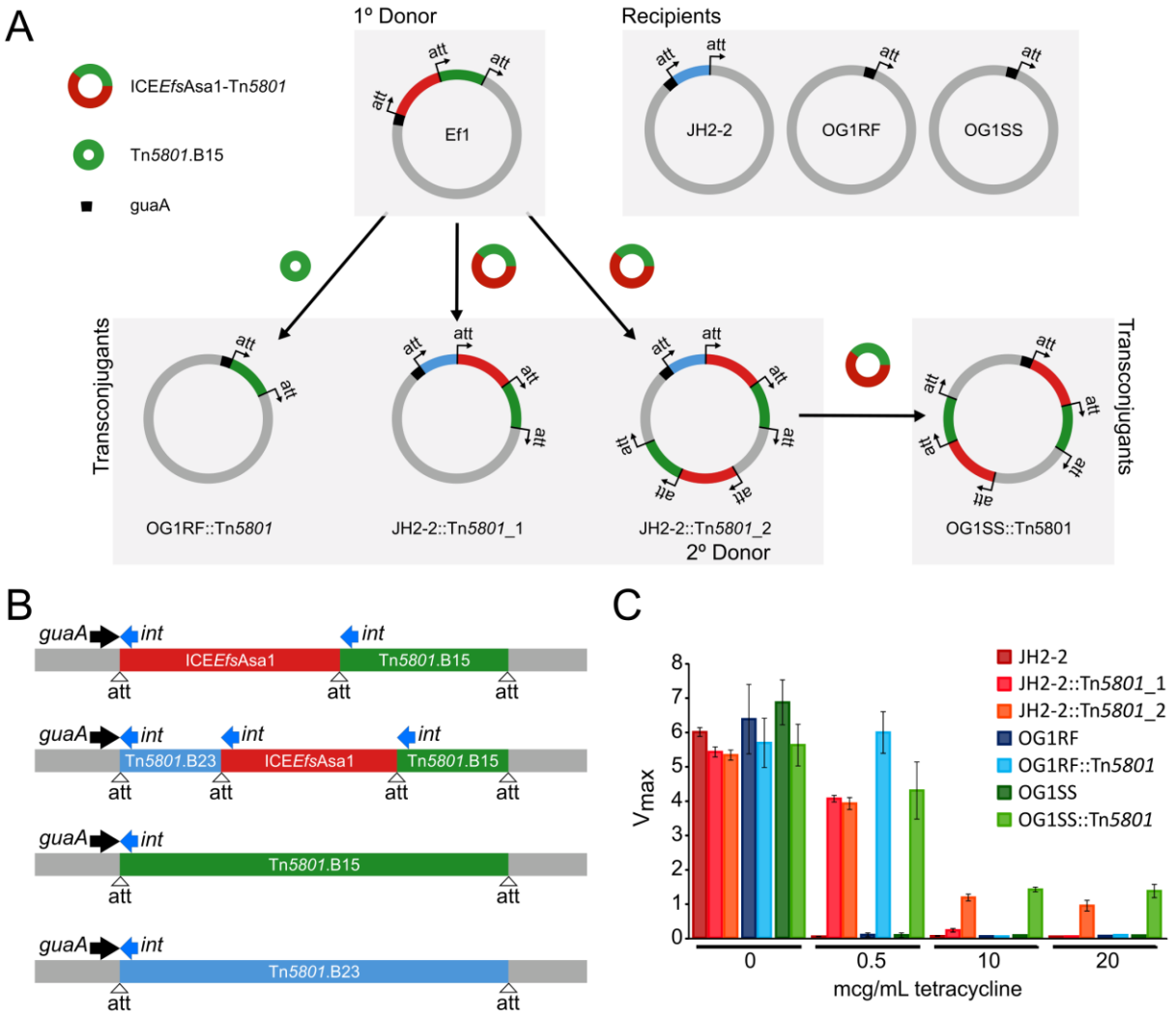


Figure 2. Transfer dynamics of Tn*5801*-like elements and behaviour in presence of tetracycline. (A and B) Tn*5801*.B15, Tn*5801*.B23 and ICE*Efs*Asa1 integrases are associated with the same attachment site (att). All the elements are flanked by the direct repeat (DR) of the att. The recombination event for integration produces the repeat sequence of the att, which allow the insertion of a new Tn*5801*-like element. In A, the primary donor (*E. faecalis* Ef1) harbours the ICE*Efs*Asa1-Tn*5801* tandem, which is circularized, transferred to JH2-2 recipient and two copies were inserted in *att* sites. The Tn*5801* from the primary donor shows the ability to excise and transfer to OG1RF by itself. In the secondary filter mating, JH2-2::Tn*5801*_2 acts as a secondary donor and the tandem in mobilized to OG1SS and two copies were inserted. (C) Vmax of the recipients and transconjugants growth rates in presence of different tetracycline concentrations (0, 0.5, 10 and 20 mcg/mL).

Characterization of ICE*Efs*Asa1, a novel conjugative element

A 32.7 kb element detected in the *E. faecalis* Ef1 strain, the original donor isolate in the filter mating experiments, is flanked by two 11-bp direct repeats (DRs, *attL* and *attR*) (GAGTGGGAATA) situated within the 3′ end of the *guaA gene*. The annotation of this putative transposon called ICE*Efs*Asa1 predicts 39 ORFs involved in disparate functions that include mobilization (site-specific integrase), replication, recombination and repair (ATP/GTP-binding

protein and Nick relaxase), cell wall biogenesis (periplasmic protein, Asa1), cell division (chromosome segregation protein and Ftsk) and transcription (*hipB*, a transcriptional regulator) (Figure 3). The sequence of ICE*Efs*Asa1int (1167 bp) shows a low coverage and identity with that of Tn*5801*int (68% each), although is the highest identity found in the database for Tn*5801*int. The *asa1* gene codes for the aggregation substance, a putative virulence factor usually found in pheromone responsive plasmids or in a chromosomal PAI. The aggregation substance allows donor and recipient strains to maintain physical contact to permit conjugative transfer of the conjugative element from donor to recipient strain (63).

Blast search revealed the presence of ICE*Efs*Asa1 in five genomes, that of *E. faecalis* V583 (GenBank AE016830.1; 100% of coverage and identity) and four draft genomes, with 100% of coverage and 99% of identity, *E. faecalis* EnGen0287 strain YI6-1 (ST28, GenBank: AJEO01000004.1), *E. faecalis* 12 from a copper fed pig (GenBank: JTKT01000001.1), *E. faecalis* Hp_74-d1 (GenBank: FREL01000004.1), and *E. faecium* Hp_74-d2 (GenBank: FREF01000004.1). In all cases, the ICE*Efs*Asa1 was inserted at the 3´end of the *guaA* gene.

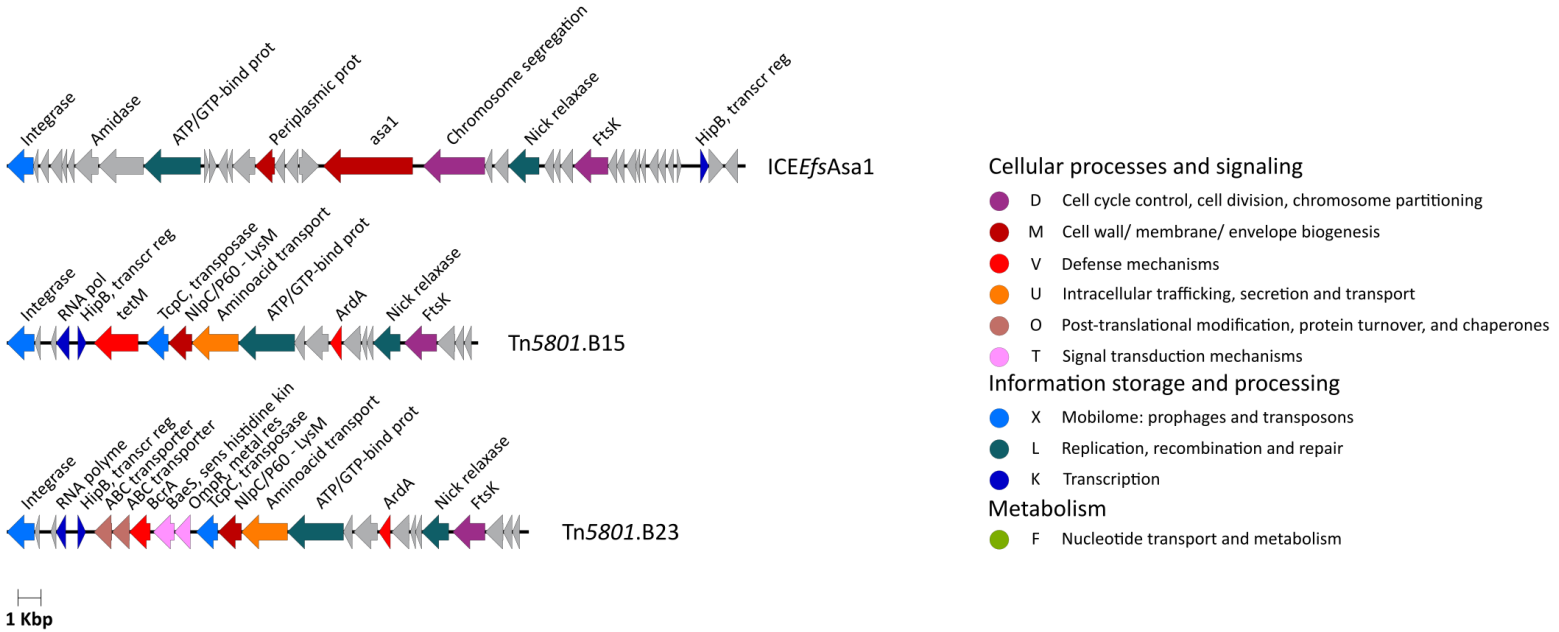


Figure 3. Backbones of the ICE*Efs*Asa1, Tn*5801*.B15 and Tn*5801*.B23 with the functional COG classification.

Effect of different tetracycline concentrations in growth rates of Tn5801 transconjugants

We measured the growth rates of all the isolates in the presence of different concentrations of tetracycline. The aim of these assays was to evaluate the benefit of harboring different ICEs or different copies of ICEs (e.g. Tn*5801*.B15). Note we previously had noted changes in the MIC values, but the effect of these elements in the growth rate of the host had not been addressed. All the Vmax values are shown in Table 1.

In the absence of tetracycline, we did not note significant differences in the Vmax of any of the isolates studied (Figure 2). The same occurred when using a concentration of tetracycline of 0.5 mcg/mL for the *E. faecalis* transconjugants JH2-2::Tn*5801*_2 containing two copies of the tandem ICE*Efs*Asa1-Tn*5801*.B15 or JH2-2::Tn*5801*_1 harbouring a single copy. However, using the same concentration, we observed that *E. faecalis* OG1-RF::Tn*5801* harboring one copy of Tn*5801*.B15 had a significantly higher Vmax (5.974 ± 0.604) than *E. faecalis* OG1-SS::Tn*5801* with two copies of the composite-tandem ICE (4.285 ± 0.922). In the presence of high tetracycline concentrations (10 and 20 mcg/mL) the transconjugants *E. faecalis* JH2-2::Tn*5801*_2 and OG1SS::Tn*5801* showed a significant difference compared to the corresponding receptor strains. This result highlights the loss of growth rate caused by the presence of the ICE*Efs*Asa1-Tn*5801*.B15 in comparison with the strains harbour a single Tn*5801*.B15 copy. At the same time, these results demonstrate that the second copy is functional.

Discussion

We fully characterize for the first time the genetic elements involved in the mobilization of Tn*5801* in *E. faecalis* and describe the ability of members of this family to integrate at site specific sites present not only in 3´end of the *guaA* gene but also in different chromosomal locations of *E. faecalis* genomes. The present work enlarges the list of ICEs/MGEs that transfer by conjugation and highlights the diversity of mechanisms enabling their conjugative mobilization, in particular, that resulting from tandem accretion after site-specific recombination.

A plethora of ICEs belonging to the Tn*916* family which differ in the integrase, excisionase and the insertion site specificity has been described (35). Although most Tn*916*int are tyrosine recombinases, their integration sites in the host vary from AT rich sites (e.g. Tn*916*) to specific integration sites at the 3´end of the gene encoding GMP (e.g. Tn*5801*, Tn*6000*, and many ICEs of *E. faecalis* as EfC2, EfC1, EfC2). The different affinity for the chromosomal integration site help to explain differences between the number of copies and host range of the Tn*916* paradigm and different members of the Tn*916* family, which have only been located as a single copy for many years.

Most ICEs have been found in a single copy per analyzed genome or per transconjugant, probably due to the low conjugation frequency of most elements (139). However, with the exponential increase of available sequences, identical or closely related ICEs that can integrate in several or numerous different sites are increasingly identified in different locations of some genomes, including ICEs encoding either tyrosine integrases or DDE transposases (131).

The detailed analysis of a transferable Tn*5801* element revealed a composite element in tandem with ICE*Efs*Asa1 in both the donor and the transconjugants, thus reflecting than they were generated by site-specific accretion (140). This finding enabled us to re-define the insertion site of these ICEs, restricted to the 3´end of the *guaA* gene to date. In this way, these ICEs (both Tn*5801*int and ICE*Efs*Asa1int are highly homologous) may be transferred by conjugation in a recipient cell carrying a related element already integrated in the *attB* site of the host and then, be integrated into one of the two *attL/attR* sites flanking that ICE leading to the tandem accretion of two (the 53,3kb conjugative element comprising Tn*5801* and ICE*Efs*Asa1), or three elements (the tandem of Tn*5801*.B23 of JH2-2, Tn*5801*.B23 and ICE*Efs*Asa1). A further illustration is provided by the example of *E. faecalis* genomes Com1 and Com2 (GenBank

accession no. AJES01000002.1 and AJBL01000002.1, respectively), which harbor a composite structure that comprises Tn*5801*.B15 and an ICE*Efs*Asa1-like element lacking *asa1* gene, both being located at the 3´end of the *guaA* gene. Similarly, these events are common among other Firmicutes as Clostrdium (141).

The presence of two complete copies of the 53,3kb composite ICE in the transconjugants of the different recipients JH2-2 and OG1SS remains to be fully elucidated. Although the two copies are identical, so conserved the same *attL* and *attR* sites, we were unable to determine if the acquired 100kb region correspond to two copies integrated in tandem or integrated at different chromosomal locations due to the limitations of the sequencing approach used (Figure XX). Anyway, the presence of two copies in the transconjugants obtained in primary and secondary mattings suggests the generation of different copies in the donor that further underwent different events involving integration and excision or homologous recombination.

Although the mechanism responsible for the acquisition of two copies remains unknown, some hypothesis could explain the process. A recent study demonstrated that Tn*916* undergoes autonomous rolling circle replication when excise the chromosome and become in the circular form, thus suggesting an autonomous replication of Tn*916* (139).

One intriguing question is the failure to transfer the tandem of three ICEs resulting from the insertion of the composite 53,3kb element (Tn*5801*.B15 and ICEefasa1) in the *attP* site of Tn*5801*.B25 of the JH2-2. A recent work confirms the influence of tetracycline in the integration and excision of Tn*916* by enhancing the transcription of integrase and the excisionase (139). The differences of the Tn*5801*.B23 with other Tn*5801* variants (including the lack of the region orf19—*tetM*-orf21 and the 92% similarity of the region between orf26 and orf16 with other Tn*5801* members) could influence the expression of the genes responsible for the integration and excision (57). Finally, it is of note how the increase of the gene dosage by acquiring more than one copy of Tn*5801* improves the level of resistance and the growth rate of the host (57).

In summary, the work provides evidence of the integration of site-specific transposons at different sites in a genome resulting in tandem accretion after site-specific recombination. The results modify the current understanding about the transferability of these elements (assumed to integrate in a single 3´end of a gene encoding a GMP-syntase and to be transferred vertically) and suggest a great contribution of these elements and the associated mobilization events to the dynamics of the *E. faecalis* genomes. The work enlarges the number of experimental transfer

leading to accretion, only demonstrated for very few elements. The findings are relevant to understand the population structure of the species and in general the dynamics of microorganisms with a recombinant population structure.

Integrative discussion

Integrative discussion

Enterococcus faecalis is a generalist bacterial species that inhabitant a variety of disparate hosts such as mammals, reptiles, insects and birds, being a potential cause of severe infections in all these hosts (10, 61, 62). This enterococcal species is also one of the major nosocomial pathogens and the main cause of hospital-acquired enterococcal infections (5, 142).

Some *E. faecalis* clones are predominant in the clinical settings and frequently carry genes conferring antibiotic resistance and virulence traits. These clones are CC2, CC9 o CC87 which are also recovered from other non-human reservoirs (15, 24, 26, 63, 143). The wide dissemination and adaptability of *E. faecalis* populations to a variety of environments suggest the relevance of horizontal gene transfer between populations of the same and different species which result in its recombinogenic structure. In this concern, the phylogenetic genome trees analysed in previous works reflected a lack of linkage between their topology and the ecological origins which also explains the recombinogenic nature of the species (23, 30, 84). Such population structure favours its behaviour as a collective evolutionary individual, and its high level of genetic connectivity in different niches. The low representation of *E. faecalis* isolates from non-human host and the limitation of tools to analyse species with large accessory genomes (86) do not allow confirming such hypothesis which would be based on fractionated knowledge. In this work we provide new knowledge about the structure of *E. faecalis* populations and their niche-specific adaptations. A detail revision of the accessory genome is given, focusing on the evolution of and transmission of specific emblematic mobile genetic elements as CTns known to have specific integration sites. CTns are considered for a long time, drivers of enterococcal evolution (92).

In **chapter I** we analysed a large collection of *E. faecalis* from Spanish migratory birds and revised the population structure of the species considering both the accessory and the accessory genome using novel bioinformatics tools (72, 66). To date, the population structure of *E. faecalis* is mainly based on MLST data analyzed by different tools (goEBURST, BAPS) (23, 26, 29, 84) that suggested an epidemic structure with a high recombination /mutation ratio and highlighted the difficulties to establish the population structure of this species. Wild animals are a good source of study because they are subjected to a variety of selective pressures (antibiotics, heavy metals, biocides) and have a high rate of network of contacts promoting the exchange and flow between different bacterial communities. We confirm the limited specialization in *E. faecalis* of previous

works by using BAPS and comparative genomics (ref). However, we also demonstrated the small core and large accessory genome that would support the Black Queen hypothesis according to which *E. faecalis* could rapidly adapt to different environments by acquisition of a large pangenome and the arsenal of mobile genetic elements that enable different horizontal gene transfer events. In this scenario, niche specific convergence between *E. faecalis* clones from different host species of wild birds emerges and resulted in clonal complexes restricted to some lifestyles like CC82 confined to wild birds. The MLST analysis and further genome sequencing allow questioning the validity of MLST classical scheme showing duplications of some of the housekeeping genes, promoted by recombinations and horizontal gene transfer events.

Both the small size of the core genome and the large size of the accessory genome reflects the phylogenetic distance of all clusters. Within the accessory genome, the plasmids showed a high degree of mosaicism further demonstrating their ability to recombine and construct chimeras. This high mosaicism and the lack of plasmids in some isolates suggest differences in the adaptation between hosts.

In **chapter II** we focused on the evolution of an integrative and conjugative element highly influenced by the use of tetracycline in therapeutics. Tetracycline resistance in major Gram positive human opportunistic pathogens is mainly caused by the acquisition of integrative and conjugative elements (ICE) of the Tn*916* family carrying the *tet*(M) gene (91, 100). These Tn*916*-like elements include a wide variety of ICE with a common synteny but differing in the integration and transfer mechanisms, the insertion site specificity and the host range. Tn*5801* was considered a site specific element that nonetheless, had been detected in different genus of clinical interest for a very long time (staphylococci, streptococci, enterococci). We comprehensively analyzed all the Tn*5801* sequences available in public databases and described different variants classified in two main groups designed as A (ca. 25kb) and B (ca. 20kb) overrepresented in genomes of Staphylococci and Streptococci, respectively. . The element apparently evolved from its short form (Tn*5801*.B15) in *S. agalactiae* further acquired 5kb by staphylococci. Moreover, the dispersion analysis of this element among the Firmicutes species suggests the intervention and the repository role of enterococcus, carrying both Tn*5801* types (A and B).

Besides the effects of the use of tetracycline, other bacterial features may contribute to the selection of virulent clones carrying Tn*5801* as anti-Restriction Modification (RM), which acts

against staphylococcal Type I RM systems and influences selective horizontal gene transfer between staphylococcal populations and other species of Firmicutes (120, 121).

We demonstrate here the transference of Tn*5801* between different *E. faecalis* backgrounds, even if the strain already harbored a Tn*5801* element. The finding is relevant as this transposon has a site specific tyrosine recombinase that recognizes the 3´-end of guanosine monophosphate (GMP) synthase gene (*guaA*) (94, 95), an insertion hotspot of genomic islands in Firmicutes and can be functional at secondary integration sites (95, 96). This study allows us the better understanding of the potential of transmission of these elements. Also, the high level of plasticity, demonstrated with the finding of the Tn*5801*.B23 variant in the classical *E. faecalis* JH2-2 laboratory strain.

In summary, this study shows that Tn*5801*-like variants are present among Firmicutes since at least the 1950´s, being predominant in microorganisms obtained from human hosts (51). Different waves of expansion in human populations of certain clones in *Staphylococcus*, *Enterococcus*, and *Streptococcus* seems to play an important role in their occurrence and evolvability although transfer events between different clonal backgrounds would also contribute to the spread, demonstrated experimentally here for *E. faecalis*.

In **chapter III**, we fully analyzed the genetic context of the Tn*5801* transconjugants obtained in chapter II in order to understand the mobilization events of Tn*5801* and the impact of acquiring this CTn in the bacterial fitness of *E. faecalis.* We demonstrated the unexpected transferability and integration of one or more than one copy of this CTn to different chromosomal sites

A novel 53,3kb composite element designed as ICE50CE comprising Tn*5801*.B25 and a new 30kb ICE called ICEEfsAsa1, was responsible of the mobilization of Tn*5801* in the *E. faecalis* analyzed in our work. Moreover, the ability of Tn*5801* elements to integrate at site-specific sites present in 3´end of the *guaA* gene and other chromosomal locations of *E. faecalis* genomes was also demonstrated. It is of note that the $MIC_{Tetracycline}$ values and V_{max} of the growth curves were higher for transconjugants harboring two 50CE copies than those with a single 50CE copy (24-48 vs 96vs1mg/L and 4.28 vs 5.97, respectively) in presence of tetracycline.

In summary, the work highlights the diversity of mechanisms enabling their conjugative mobilization, highlighting those of tandem accretion after site-specific recombination. Such strategy could facilitate the persistence and amplification of some pathogenic clones besides the contribution to the evolvability of the species and the impact in the population structure. Finally,

it suggests the effect of tetracycline in the transferability of Tn*916*-elements and the fitness of populations carrying tetracycline resistance elements in selective environments.

Conclusions

Conclusions

1. The analysis of *E. faecalis* genomes from wild birds and the comprehensive study both core and accessory genomes by using novel approaches firstly tested here for this species, revealed a large accessory genome and confirm the "endemic population structure" of the species, with a high rate of recombination as a result of introgression at different hierarchical levels .

2. The large *E. faecalis* accessory genome with a high degree of dispersion among the genomic clusters supports the genetic interchange and the recombinogenic nature of this species. Moreover, the presence of habitat-specific proteins in the *E. faecalis* CC82 cluster, related with metabolism adaptation and mobile genetic elements, suggests the contribution of a specific host adaptation for wild birds.

3. Data from this work revealed the lack of plasmids in a high percentage of *E. faecalis* isolates from wild bird hosts reinforcing the old (and barely explored idea) that the carriage of plasmids increases with the proximity to settings inhabited by humans. Moreover, the plasmids detected exhibited a high degree of mosaicism, with an unusual number of replicases, relaxases and sequences of plasmids from different origins, suggesting a long history of rearrangements and horizontal gene transfer events of *E. faecalis* plasmids between phylogenetically distant or related populations.

4. The diversity and evolutionary history of Tn*5801* was determined herein, showing two well-defined types (A and B) with a common backbone of streptococcal origin. These types show apparent confinement in particular clonal backgrounds. This confinement supports a positive selection of specific virulent clonal lineages since the introduction of tetracycline in the 1950s in *S. agalactiae*, *S. aureus, E. faecium* or *E. faecalis*. Some particular bacterial features like anti-restriction-modification systems could have facilitated the selective acquisition of certain elements by particular clones influencing selective horizontal gene transfer between genera.

5. The unusual number of copies and size of transconjugants containing Tn*5801*-like elements revealed the putative integration sites at different locations of the E. faecalis genomes which change the traditional belief that these elements were site specific transposon with a single target site at the 3´ end of the gene coding for GMP synthase. The possibility to target multiple sites resulted in tandem accretion after site-specific recombination and multiple possibilities for the evolvability. A composite element in tandem with ICEEfasa1 in both the donor and the

transconjugants can be transferred by conjugation in a recipient cell already carrying a related element leading to the tandem accretion of two or three conjugative elements.

6. The generation of composite transposon carrying the same antibiotic resistance gene or different copies of the same transposon leads to increase the gene dosage which improves the level of resistance and also improve the fitness by increasing the growth rate of the host. These elements and the HGT events involved in their maintenance contribute to understand the dynamics of the *E. faecalis* genomes and the selection and persistence of major clonal lineages.

Appendix 1. Supplementary dataset

Appendix 1. Supplementary dataset

Table S2 (Chapter II)

Tn*5801* Variant	Species	Isolate	MLST	Source/ Host	Year	Country	GenBank	Unique sequence pattern			
								orf7-4	orf12-8	orf18-14	Int_{Tn5801}
A 1	Sau	21209	507	unk	unk	unk	AGRP01000035.1	12	12	12	12
A 1	Sau	2395 USA500	8	Human	2000	USA	CP007499.1	12	14	12	12
A 1	Sau	A9765	8	unk	unk	unk	ACSN01000037.1	13	14	13	12
A 1	Efs	B16457	250	Human	2007	DNK	AIIL01000015.1	15	12	24	12
A 1	Smi	B6	-	Human	unk	DEU	NC_013853.1	14	12	25	12
A 1	Sag	BSU108	-	unk	unk	unk	ALRK01000022.1	12	12	17	12
A 1	Sau	CDC 73-57501 DP18	-	unk	unk	unk	JOVN01000005.1	12	12	14	12
A 1	Sau	CIG1114 PSABG_7	8	Human	1994 - 2003	USA	AHKD01000007.1	12	12	12	12
A 1	Sau	CIG1612 PSABG_10_	8	Human	1994 - 2003	USA	AHVW01000003.1	12	12	12	12
A 1	Sau	CIG1770 PSABG_2_	664	Human	1994 - 2003	USA	AHVN01000005.1	12	12	12	12
A 1	Sau	CIG2018 PSABG_29_	609	Human	1994 - 2003	USA	AHVV01000011.1	12	12	12	12
A 1	Efm	E1392	64	Human	2000	GBR	AHWV01000009.1	16	12	12	12
A 1	Sps	ED99	-	unk	unk	unk	CP002478.1	17	12	20	12
A 1	Sau	F26051	8	Human	2012	USA	JFUO01000021.1	12	14	12	12
A 1	Sau	H41336	8	Human	2013	unk	JEOX01000009.1	12	14	12	12
A 1	Sau	H79477	8	Human	2013	unk	JDNL01000026.1	12	14	12	12
A 1	Sau	M0002	609	Human	2003	USA	JAZD01000025.1	12	12	12	12
A 1	Sau	M0007	-	Human	2003	USA	JBVJ01000017.1	12	12	12	12
A 1	Sau	M0008	609	Human	2003	USA	JBVK01000009.1	12	12	12	12
A 1	Sau	M0047	609	Human	2003	USA	JAYZ01000016.1	12	12	12	12
A 1	Sau	M0059	609	Human	2003	USA	JAYY01000038.1	12	12	12	12
A 1	Sau	M0080	609	Human	2003	USA	JBXC01000028.1	12	12	12	12
A 1	Sau	M0088	609	Human	2003	USA	JAYW01000022.1	12	12	12	12
A 1	Sau	M0097	609	Human	2003	USA	JBXP01000026.1	12	12	12	12
A 1	Sau	M0100	609	Human	2003	USA	JBXS01000027.1	12	12	12	12
A 1	Sau	M0101	609	Human	2003	USA	JBXT01000023.1	12	12	12	12
A 1	Sau	M0107	609	Human	2003	USA	JBXU01000026.1	12	12	12	12
A 1	Sau	M0109	609	Human	2003	USA	JBXV01000027.1	12	12	12	12
A 1	Sau	M0111	609	Human	2003	USA	JBXX01000030.1	12	12	12	12
A 1	Sau	M0113	609	Human	2003	USA	JAYU01000027.1	12	12	12	12
A 1	Sau	M0123	609	Human	2003	USA	JBYG01000013.1	12	12	12	12
A 1	Sau	M0186	609	Human	2003	USA	JBZZ01000010.1	12	12	12	12
A 1	Sau	M0221	609	Human	2003	USA	AIWC01000007.1	12	12	12	12
A 1	Sau	M0284	609	Human	2003	USA	JAZU01000041.1	12	12	12	12
A 1	Sau	M0312	609	Human	2003	USA	ANZS01000016.1	12	12	12	12
A 1	Sau	M0337	609	Human	2003	USA	JCDL01000030.1	12	12	12	12
A 1	Sau	M0340	609	Human	2003	USA	AIWK01000006.1	12	12	12	12
A 1	Sau	M0352	609	Human	2003	USA	JAZP01000022.1	12	12	12	12
A 1	Sau	M0353	609	Human	2003	USA	JCDV01000007.1	12	12	12	12
A 1	Sau	M0424	609	Human	2003	USA	AJCW01000012.1	12	12	12	12

Appendix 1: Supplementary dataset

Tn*5801* Variant	Species	Isolate	MLST	Source/ Host	Year	Country	GenBank	Unique sequence pattern			
								orf7-4	orf12-8	orf18-14	Int_{Tn5801}
A 1	Sau	M0453	-	Human	2003	USA	JCGC01000026.1	12	12	12	12
A 1	Sau	M0454	609	Human	2003	USA	JCGD01000017.1	12	12	12	12
A 1	Sau	M0457	609	Human	2003	USA	JCGE01000011.1	12	12	12	12
A 1	Sau	M0468	609	Human	2003	USA	AIWS01000019.1	12	12	12	12
A 1	Sau	M0469	609	Human	2003	USA	JCGL01000025.1	12	12	12	12
A 1	Sau	M0511	609	Human	2003	USA	JBAQ01000019.1	12	12	12	12
A 1	Sau	M0524	609	Human	2003	USA	JBAP01000017.1	12	12	12	12
A 1	Sau	M0527	609	Human	2003	USA	JCHQ01000027.1	12	12	12	12
A 1	Sau	M0542	609	Human	2003	USA	JCHY01000009.1	12	12	12	12
A 1	Sau	M0563	609	Human	2004	USA	JCIN01000010.1	12	12	12	12
A 1	Sau	M0564	609	Human	2004	USA	JCIO01000007.1	12	12	12	12
A 1	Sau	M0567	609	Human	2004	USA	JCIQ01000020.1	12	12	12	12
A 1	Sau	M0577	-	Human	2004	USA	JCIV01000009.1	12	12	12	12
A 1	Sau	M0593	609	Human	2004	USA	JCJH01000009.1	12	12	12	12
A 1	Sau	M0615	609	Human	2004	USA	JCJV01000009.1	12	12	12	12
A 1	Sau	M0630	609	Human	2004	USA	JBEH01000015.1	12	12	12	12
A 1	Sau	M0633	609	Human	2004	USA	AJDF01000008.1	12	12	12	12
A 1	Sau	M0636	609	Human	2004	USA	JBED01000009.1	12	12	12	12
A 1	Sau	M0645	-	Human	2004	USA	JBDW01000007.1	12	12	12	12
A 1	Sau	M0658	609	Human	2004	USA	JBDM01000010.1	12	12	12	12
A 1	Sau	M0668	609	Human	2004	USA	JBDG01000010.1	12	12	12	12
A 1	Sau	M0671	609	Human	2004	USA	JBDD01000012.1	12	12	12	12
A 1	Sau	M0709	609	Human	2004	USA	JBCB01000016.1	12	12	12	12
A 1	Sau	M0729	609	Human	2004	USA	JBBM01000008.1	12	12	12	12
A 1	Sau	M0730	609	Human	2004	USA	JBBL01000010.1	12	12	12	12
A 1	Sau	M0760	609	Human	2004	USA	JCWN01000028.1	12	12	12	12
A 1	Sau	M0766	609	Human	2004	USA	JCWS01000016.1	12	12	12	12
A 1	Sau	M0771	609	Human	2004	USA	JCWV01000026.1	12	12	12	12
A 1	Sau	M0800	-	Human	2004	USA	JCPA01000006.1	12	12	12	12
A 1	Sau	M0801	609	Human	2004	USA	JCPB01000008.1	12	12	12	12
A 1	Sau	M0831	609	Human	2004	USA	AOAG01000025.1	12	12	11	12
A 1	Sau	M0833	609	Human	2004	USA	JCPU01000025.1	12	12	11	12
A 1	Sau	M0853	609	Human	2004	USA	JCQI01000013.1	12	12	12	12
A 1	Sau	M0858	609	Human	2004	USA	JCQK01000030.1	12	12	11	12
A 1	Sau	M0893	8	Human	2004	USA	JCRE01000030.1	12	12	12	12
A 1	Sau	M0894	8	Human	2004	USA	JCRF01000026.1	12	12	12	12
A 1	Sau	M0910	609	Human	2004	USA	JCRQ01000019.1	12	12	12	12
A 1	Sau	M1007	609	Human	2004	USA	AIYA01000005.1	12	12	12	12
A 1	Sau	M1017	609	Human	2004	USA	JCTZ01000007.1	12	12	12	12
A 1	Sau	M1056	609	Human	2004	USA	JCUU01000008.1	12	12	12	12
A 1	Sau	M1064	-	Human	2004	USA	AIYK01000007.1	12	12	12	12
A 1	Sau	M1094	609	Human	2004	USA	JCVO01000027.1	12	12	12	12
A 1	Sau	M1101	609	Human	2004	USA	JCVQ01000008.1	12	12	12	12

Tn*5801* Variant	Species	Isolate	MLST	Source/ Host	Year	Country	GenBank	Unique sequence pattern			
								orf7-4	orf12-8	orf18-14	Int_{Tn5801}
A 1	Sau	M1124	609	Human	2004	USA	JELG01000010.1	12	12	12	12
A 1	Sau	M1132	609	Human	2004	USA	JELB01000008.1	12	12	12	12
A 1	Sau	M1153	609	Human	2004	USA	JEKN01000017.1	12	12	12	12
A 1	Sau	M1183	507	Human	2004	USA	JAWX01000029.1	12	12	12	12
A 1	Sau	M1223	609	Human	2004	USA	AOBA01000009.1	12	12	11	12
A 1	Sau	M1262	507	Human	2004	USA	JEIB01000028.1	12	12	12	12
A 1	Sau	M1263	609	Human	2004	USA	JEIA01000024.1	12	12	11	12
A 1	Sau	M1270	609	Human	2004	USA	JEHV01000027.1	12	12	11	12
A 1	Sau	M1294	507	Human	2004	USA	JAWR01000015.1	12	12	12	12
A 1	Sau	M1297	609	Human	2004	USA	JEHC01000027.1	12	12	11	12
A 1	Sau	M1304	609	Human	2004	USA	JEGX01000023.1	12	12	12	12
A 1	Sau	M1319	609	Human	2004	USA	JEGN01000024.1	12	12	11	12
A 1	Sau	M1346	507	Human	2004	USA	JAWN01000027.1	12	12	12	12
A 1	Sau	M1348	609	Human	2004	USA	JCZZ01000022.1	12	12	11	12
A 1	Sau	M1351	609	Human	2004	USA	JDAC01000025.1	12	12	12	12
A 1	Sau	M1375	609	Human	2004	USA	JAXR01000023.1	12	12	12	12
A 1	Sau	M1400	609	Human	2004	USA	JCXP01000026.1	12	12	12	12
A 1	Sau	M1408	609	Human	2004	USA	JCXU01000028.1	12	12	12	12
A 1	Sau	M1427	609	Human	2004	USA	JAXL01000027.1	12	12	12	12
A 1	Sau	M1436	-	Human	2004	USA	JCYS01000033.1	12	12	12	12
A 1	Sau	M1447	609	Human	2004	USA	JCZA01000026.1	12	12	12	12
A 1	Sau	M1453	507	Human	2004	USA	JCZE01000028.1	12	12	12	12
A 1	Sau	M1462	609	Human	2004	USA	AIZO01000003.1	12	12	12	12
A 1	Sau	M42184	8	Human	2012	USA	JFTT01000023.1	12	14	12	12
A 1	Sau	M90388	8	Human	2012	USA	JFTI01000022.1	12	14	12	12
A 1	Sau	Mu3	5	Human	1996	JPN	NC_009782.1	16	12	22	12
A 1	Sau	Mu50	5	Human	2005	JPN	NC_002758.2	16	12	22	12
A 1	Sau	Mu50-omega	5	unk	-	-	BABM01000001.1	16	12	22	12
A 1	Efm	VRE 110	18	Human	2010	DNK	AIVD01000015.1	16	12	21	12
A 10	Sau	16K DNA	239	Human	unk	unk	BABZ01000026.1	18	15	28	12
A 10	Sau	21178	372	unk	unk	unk	AGRN01000068.1	18	15	26	12
A 10	Sau	ATCC BAA-39	239	Human	2005	USA	AEEK01000047.1	18	15	26	12
A 10	Sau	Bmb9393	239	Human	1993	BRA	NC_021670.1	18	18	26	12
A 10	Sau	CN79	239	Human	2006	CHN	ANCJ01000005.1	18	15	30	12
A 10	Sau	CUHK_BJ2002	239	Human	2002	CHN	AZMY01000017.1	18	15	29	11
A 10	Sau	CUHK_BJ2007	239	Human	2007	CHN	AZMX01000027.1	18	15	29	12
A 10	Sau	CUHK_HK1997	239	Human	1997	HKG	AZJQ01000029.1	18	15	26	12
A 10	Sau	IS-125	239	unk	unk	unk	AHVC01000080.1	18	15	26	12
A 10	Sau	IS-157	239	unk	unk	unk	AICH01000181.1	18	15	26	12
A 10	Sau	JKD6008	239	unk	unk	AUS	NC_017341.1	20	16	27	13
A 10	Sau	JKD6009	239	unk	unk	AUS	ABSA01000060.1	20	16	27	13
A 10	Sau	MRGR3	239	Human	1990	CHE	AHZL01000011.1	18	15	26	12
A 10	Sau	SK1585	2249	Human	1973	AUS	AYLT01000067.1	19	16	26	12

Appendix 1: Supplementary dataset

Tn*5801* Variant	Species	Isolate	MLST	Source/ Host	Year	Country	GenBank	Unique sequence pattern			
								orf7-4	orf12-8	orf18-14	Int_{Tn5801}
A 10	Sau	T0131	239	Human	2006	CHN	NC_017347.1	18	15	29	12
A 11	Sau	CUHK_HK2007	239	Human	2007	HKG	AZMZ01000003.1	18	17	26	12
A 11	Sau	IS-189	-	unk	unk	unk	AICJ01000068.1	18	15	26	12
A 11	Sau	PPUKM-261-2009	239	Human	2009	MYS	AMRB01000010.1	18	15	26	12
A 11	Sau	PPUKM-332-2009	239	Human	2009	MYS	AMRC01000013.1	18	15	26	12
A 11	Sau	PPUKM-775-2009	239	Human	2009	MYS	AMRE01000008.1	18	15	26	12
A 11	Sau	TW20	239	Human	unk	unk	NC_017331.1	18	15	26	12
A 11	Sau	XN108	239	Human	2004	CHN	CP007447.1	18	15	26	12
A 11	Sau	Z172	239	Human	2010	TWN	NC_022604.1	18	15	26	12
A 12	Sau	FVRH6002	-	Human	unk	USA	JBSL01000005.1	18	15	26	14
A 12	Sau	KINW6048	-	Human	unk	USA	JBEK01000003.1	18	15	26	14
A 12	Sau	LAMC0011	239	Human	unk	USA	JBRH01000025.1	18	15	26	14
A 12	Sau	SJOS6053	239	Human	unk	USA	JBOE01000006.1	18	15	26	14
A 12	Sau	SJOS6072	239	Human	unk	USA	JBNV01000011.1	18	15	26	14
A 12	Sau	UCIM6015	239	Human	unk	USA	JBMO01000003.1	18	15	26	14
A 12	Sau	UCIM6042	239	Human	unk	USA	JBPU01000008.1	18	15	26	14
A 12	Sau	WAMC6102	239	Human	unk	USA	JBGY01000011.1	18	15	26	14
A 13	Sau	WAMC6030	5	Human	unk	USA	JBIQ01000012.1	12	12	16	12
A 14	Sau	PSP1996	247	Human	1996	ESP	ANHU01000028.1	11	12	15	12
A 2	Lga	IPLA 31405	-	Spanish cheese	2008	ESP	AKFO01000017.1	12	12	18	12
A 3	Efm	UAA1023	50	unk	1996	FRA	AIZZ01000002.1	14	11	19	12
A 3	Efm	UAA1024	50	unk	1996	FRA	AJAA01000002.1	14	11	19	12
A 3	Efm	UAA723	50	unk	1993	FRA	AIUN01000002.1	14	11	19	12
A 3	Efm	UAA724	-	unk	1993	FRA	AIUO01000001.1	14	11	19	12
A 4	Efm	510	18	Human	unk	unk	AMBI01000178.1	16	12	22	12
A 4	Efm	BM4538	18	Human	2004	AUS	AXOJ01000037.1	16	12	22	12
A 4	Efm	E2883	18	Human	2002	NLD	AHXX01000026.1	16	12	22	12
A 4	Efm	NEF1	18	unk	2009	FRA	AXOK01000043.1	16	12	22	12
A 4	Efm	UAA1019	18	unk	1996	unk	AIZX01000001.1	16	12	22	12
A 4	Efm	UAA1433	18	unk	2000	FRA	AJAE01000001.1	16	12	22	12
A 5	Efm	Aus0004	18	Human	1998	AUS	NC_017022.1	16	12	22	12
A 6	Efm	UAA1025	18	unk	1996	FRA	AJAB01000001.1	16	12	22	12
A 7	Efm	E1590	163	Human	2001	IRL	AHXC01000005.1	12	12	18	12
A 8	Evi	ATCC 700913	-	unk	unk	unk	ASWG01000003.1	12	12	16	12
A 9	Efm	TC6/ D344SRF*	25	unk	unk	unk	HM636636.1	14	13	23	12
B 15	Sag	CCUG 37736	-	Human	unk	unk	ANDE01000016.1	2	2	2	2
B 15	Sag	CCUG 49086H	-	Human	unk	unk	ANDQ01000019.1	2	2	2	2
B 15	Sag	COH1	-	Human	unk	unk	AAJR01000021.1	2	2	2	2
B 15	Sag	FSL S3-102	31	Human	unk	unk	ANCS01000033.1	2	2	2	2
B 15	Efm	HM1074	79	unk	1994	unk	AIUA01000003.1	2	2	2	2
B 15	Efs	62	66	Human	2006	NOR	CP002491.1	2	2	2	2
B 15	Sag	LMG 15085	17	Human	unk	unk	ALQW01000035.1	2	2	2	2
B 15	Sag	LMG 15094	17	Human	unk	unk	ALRA01000021.1	2	2	2	2

Tn*5801* Variant	Species	Isolate	MLST	Source/ Host	Year	Country	GenBank	Unique sequence pattern			
								orf7-4	orf12-8	orf18-14	Int_{Tn5801}
B 15	Sag	LMG 15095	17	Human	unk	unk	ALRB01000014.1	2	2	2	2
B 15	Sag	MRI Z1-199	23	Grey seal	2002	GBR	ANEH01000024.1	2	1	3	2
B 15	Sag	MRI Z1-200	23	Grey seal	2002	GBR	ANQM01000011.1	2	1	3	2
B 15	Sag	MRI Z1-201	23	Grey seal	2003	GBR	ANQL01000005.1	2	1	3	2
B 15	Sag	MRI Z1-203	23	Grey seal	2006	GBR	ANQJ01000013.1	2	1	3	2
B 16	Sth	DSM 12221	-	Pig	2005	BEL	ARCI01000001.1	2	2	2	2
B 17	Sag	BSU133	6	unk	unk	unk	ANEE01000005.1	2	2	2	2
B 17	Sag	BSU92	196	unk	unk	unk	ALRJ01000017.1	2	2	2	2
B 17	Sag	BSU96	17	unk	unk	unk	ALRH01000031.1	2	2	2	2
B 17	Sag	CCUG 49087	17	unk	unk	unk	ALQV01000035.1	2	2	2	2
B 17	Sag	GB00097	17	unk	unk	unk	ALSV01000013.1	2	2	2	2
B 17	Sag	GB00891	17	Human	unk	unk	ALUG01000030.1	1	2	2	2
B 18	Efs	RMC1	90	Human	1954	unk	ASDN01000007.1	8	6	1	1
B 18	Efs	TX0411	90	Human	unk	unk	AECA01000066.1	8	6	1	1
B 19	Sag	BSU178	7	Human	unk	unk	ALRG01000009.1	2	5	2	2
B 20	Efm	509	18	Human	unk	unk	AMBJ01000074.1	2	2	2	2
B 20	Efm	511	18	Human	unk	unk	AMBH01000082.1	2	2	2	2
B 20	Efm	514	18	Human	unk	unk	AMBF01000103.1	2	2	2	2
B 20	Sag	BSU253	23	Human	unk	unk	ANEB01000015.1	2	2	2	2
B 20	Efs	Com 2	34	Human	2006	USA	AJBL01000002.1	2	2	2	2
B 20	Efs	Com1	34	Human	2006	USA	AJES01000002.1	2	2	2	2
B 20	Sag	FSL S3-090	23	Human	unk	unk	ANCQ01000024.1	2	4	2	2
B 20	Sag	GB00002	23	Human	unk	unk	ALSL01000010.1	2	2	2	2
B 20	Sag	GB00247	24	Human	unk	unk	ALTG01000014.1	2	2	2	2
B 20	Sag	GB00601	24	unk	unk	unk	ALTR01000009.1	2	2	2	2
B 20	Sag	GB00867	23	Human	unk	unk	ALUB01000017.1	2	2	2	2
B 20	Sag	GB00887	23	Human	unk	unk	ALUE01000017.1	2	2	2	2
B 20	Sag	GB00932	23	unk	unk	unk	ALUT01000017.1	2	2	2	2
B 20	Sag	GB00986	23	unk	unk	unk	ALVA01000104.1	2	2	2	2
B 20	Sag	GB01003	-	Human	unk	unk	ANQO01000012.1	2	2	2	2
B 20	Sag	Gottschalk 1002A	23	Human	unk	unk	ANEY01000024.1	2	2	2	2
B 20	Sag	LMG 15091	-	unk	unk	unk	ALQX01000029.1	2	2	2	2
B 20	Sag	MRI Z1-024	23	unk	unk	unk	ALRY02000028.1	3	2	2	2
B 20	Efm	TX0133a01	18	Human	unk	unk	AECJ01000012.1	2	2	2	2
B 20	Efm	TX0133B	18	Human	unk	unk	AECI01000003.1	2	2	2	2
B 20	Efm	TX0133C	18	Human	unk	unk	AEBG01000099.1	2	2	2	2
B 21	Sag	BSU188	23	Human	unk	unk	ALRF01000127.1	2	2	2	2
B 21	Sag	CCUG 28551	-	Human	unk	unk	ANDA01000105.1	2	2	2	2
B 21	Sag	GB00018	444	Human	unk	unk	ALSP01000019.1	2	2	2	2
B 21	Sag	GB00190	23	Human	unk	unk	ALSZ01000007.1	2	2	2	2
B 21	Sag	GB00245	23	Human	unk	unk	ALTF01000018.1	2	2	2	2
B 21	Sag	GB00992	-	Human	unk	unk	ALVB01000064.1	2	2	2	2
B 21	Sag	Gottschalk 31825	23	Human	unk	unk	ANFB01000024.1	2	2	2	2

Appendix 1: Supplementary dataset

Tn*5801* Variant	Species	Isolate	MLST	Source/ Host	Year	Country	GenBank	Unique sequence pattern			
								orf7-4	orf12-8	orf18-14	Int_{Tn5801}
B 21	Sag	MC623	23	Human	unk	USA	ATOX01000036.1	2	2	2	2
B 21	Sag	MC626	23	Human	unk	USA	ATOU01000049.1	2	3	2	2
B 21	Sag	MRI Z1-204	23	Dog	unk	GBR	ANQI01000010.1	2	2	2	2
B 22	Sag	GB00957	23	Human	unk	unk	ANQS01000008.1	2	2	2	2
B 23	Efs	B327	-	Human	1985	USA	JAHV01000006.1	4	8	4	1
B 23	Efs	FA2-2	8	Human	<1973	GBR	AJBA01000004.1	4	6	4	1
B 23	Efs	JH2-2	8	unk	1974	GBR	AXOI01000011.1	6	7	4	1
B 23	Efs	SF350	64	Human	1986	unk	ASEN01000008.1	7	6	4	1
B 23	Efs	SF5039	64	Human	1991	unk	ASEO01000007.1	5	6	4	1
B 23	Efs	SF6375	64	Human	1991	unk	ASCU01000005.1	5	6	4	1
B 23	Efs	T8	-	unk	unk	unk	ACOC01000103.1	4	6	4	1
B 23	Efs	TX4000	8	Human	unk	USA	AEBB01000060.1	6	7	4	1
B 23	Efs	UAA409	8	unk	1988	FRA	AJDH01000011.1	4	6	4	1

References

References

1. **Stuart CH**, **Schwartz SA**, **Beeson TJ**, **Owatz CB**. 2006. Enterococcus faecalis: its role in root canal treatment failure and current concepts in retreatment. J Endod **32**:93–8.

2. **Facklam R**, **Carvalho M**, **Teixeira L**. 2002. History, Taxonomy, Biochemical Characteristics and Antibiotic Susceptibility Testing of Enterococci, p. 1–54. *In* Gilmore, M, Clewell, D, Courvalin, P, Dunny, G, Murray, B, Rice, L (eds.), The Enterococci: Pathogenesis, Molecular Biology and Antibiotic Resistance. American Society for Microbiology, Washington, DC, USA.

3. **Nallapareddy SR**, **Weinstock GM**, **Murray BE**. 2003. Clinical isolates of Enterococcus faecium exhibit strain-specific collagen binding mediated by Acm, a new member of the MSCRAMM family. Mol Microbiol **47**:1733–47.

4. 2014. Enterococcus Diversity, Origins in Nature, and Gut Colonization. Massachusetts Eye and Ear Infirmary.

5. **Ramsey M**, **Hartke A**, **Huycke M**. 2014. The Physiology and Metabolism of Enterococci, p. . *In* Gilmore, MS, Clewell, DB, Ike, Y, Shankar, N (eds.), Enterococci: From Commensals to Leading Causes of Drug Resistant Infection. Massachusetts Eye and Ear Infirmary, Boston.

6. **Gilmore MS**, **Clewell DB**, **Ike Y**, **Shankar N**. 2014. Enterococci. Massachusetts Eye and Ear Infirmary.

7. **Devriese LA**, **Laurier L**, **De Herdt P**, **Haesebrouck F**. 1992. Enterococcal and streptococcal species isolated from faeces of calves, young cattle and dairy cows. J Appl Bacteriol **72**:29–31.

8. **Devriese LA**, **Hommez J**, **Pot B**, **Haesebrouck F**. 1994. Identification and composition of the streptococcal and enterococcal flora of tonsils, intestines and faeces of pigs. J Appl Bacteriol **77**:31–6.

9. **Tedim AP**, **Ruíz-Garbajosa P**, **Rodríguez MC**, **Rodríguez-Baños M**, **Lanza VF**, **Derdoy L**, **Cárdenas Zurita G**, **Loza E**, **Cantón R**, **Baquero F**, **Coque TM**. 2017. Long-term clonal dynamics of *Enterococcus faecium* strains causing bloodstream infections (1995–2015) in Spain. J Antimicrob Chemother **72**:48–55.

10. **Ruiz-Garbajosa P**, **del Campo R**, **Coque TM**, **Asensio A**, **Bonten M**, **Willems R**, **Baquero F**, **Canton R**. 2009. Longer Intestinal Persistence of Enterococcus faecalis Compared to Enterococcus faecium Clones in Intensive-Care-Unit Patients. J Clin Microbiol **47**:345–351.

11. **Nes IF**, **Diep DB**, **Holo H**. 2007. Bacteriocin diversity in Streptococcus and Enterococcus. J Bacteriol **189**:1189–98.

12. **Coque TM**, **Patterson JE**, **Steckelberg JM**, **Murray BE**. 1995. Incidence of hemolysin, gelatinase, and aggregation substance among enterococci isolated from patients with endocarditis and other infections and from feces of hospitalized and community-based persons. J Infect Dis **171**:1223–9.

13. **Creti R**, **Koch S**, **Fabretti F**, **Baldassarri L**, **Huebner J**. 2006. Enterococcal colonization of the gastro-intestinal tract: role of biofilm and environmental oligosaccharides. BMC Microbiol **6**:60.

14. **McBride SM**, **Fischetti VA**, **Leblanc DJ**, **Moellering RC**, **Gilmore MS**. 2007. Genetic diversity among Enterococcus faecalis. PLoS One **2**:e582.

15. **Freitas AR**, **Novais C**, **Ruiz-Garbajosa P**, **Coque TM**, **Peixe L**. 2009. Clonal expansion within clonal complex 2 and spread of vancomycin-resistant plasmids among different genetic lineages of Enterococcus faecalis from Portugal. J Antimicrob Chemother **63**:1104–11.

16. **Bakshi U**, **Sarkar M**, **Paul S**, **Dutta C**. 2016. Assessment of virulence potential of uncharacterized Enterococcus faecalis strains using pan genomic approach – Identification of pathogen–specific and habitat-specific genes. Sci Rep **6**:38648.

17. **Arias CA**, **Murray BE**. 2012. The rise of the Enterococcus: beyond vancomycin resistance. Nat Rev Microbiol **10**:266–78.

18. **de Kraker MEA**, **Jarlier V**, **Monen JCM**, **Heuer OE**, **van de Sande N**, **Grundmann H**. 2013. The changing epidemiology of bacteraemias in Europe: trends from the European Antimicrobial Resistance Surveillance System. Clin Microbiol Infect **19**:860–8.

19. **Sadowy E**, **Sieńko A**, **Hryniewicz W**. 2011. Comparison of multilocus variable-number tandem-repeat analysis with multilocus sequence typing and pulsed-field gel electrophoresis for Enterococcus faecalis. Polish J Microbiol **60**:335–9.

20. **Coque TM**, **Seetulsingh P**, **Singh K V**, **Murray BE**. 1998. Application of molecular techniques to the study of nosocomial infections caused by enterococci. Methods Mol Med **15**:469–93.

21. **Nallapareddy SR**, **Duh R-W**, **Singh K V**, **Murray BE**. 2002. Molecular typing of selected Enterococcus faecalis isolates: pilot study using multilocus sequence typing and pulsed-field gel electrophoresis. J Clin Microbiol **40**:868–76.

22. **Chowdhury SA**, **Arias CA**, **Nallapareddy SR**, **Reyes J**, **Willems RJL**, **Murray BE**. 2009. A trilocus sequence typing scheme for hospital epidemiology and subspecies differentiation of an important

nosocomial pathogen, Enterococcus faecalis. J Clin Microbiol **47**:2713–9.

23. **Ruiz-Garbajosa P**, **Bonten MJM**, **Robinson DA**, **Top J**, **Nallapareddy SR**, **Torres C**, **Coque TM**, **Canton R**, **Baquero F**, **Murray BE**, **del Campo R**, **Willems RJL**. 2006. Multilocus Sequence Typing Scheme for Enterococcus faecalis Reveals Hospital-Adapted Genetic Complexes in a Background of High Rates of Recombination. J Clin Microbiol **44**:2220–2228.

24. **McBride SM**, **Fischetti VA**, **LeBlanc DJ**, **Moellering RC**, **Gilmore MS**. 2007. Genetic Diversity among Enterococcus faecalis. PLoS One **2**:e582.

25. **Freitas AR**, **Coque TM**, **Novais C**, **Hammerum AM**, **Lester CH**, **Zervos MJ**, **Donabedian S**, **Jensen LB**, **Francia MV**, **Baquero F**, **Peixe L**. 2011. Human and swine hosts share vancomycin-resistant Enterococcus faecium CC17 and CC5 and Enterococcus faecalis CC2 clonal clusters harboring Tn1546 on indistinguishable plasmids. J Clin Microbiol **49**:925–31.

26. **Kuch A**, **Willems RJL**, **Werner G**, **Coque TM**, **Hammerum AM**, **Sundsfjord A**, **Klare I**, **Ruiz-Garbajosa P**, **Simonsen GS**, **van Luit-Asbroek M**, **Hryniewicz W**, **Sadowy E**. 2012. Insight into antimicrobial susceptibility and population structure of contemporary human Enterococcus faecalis isolates from Europe. J Antimicrob Chemother **67**:551–8.

27. **Novais C**, **Freitas AR**, **Silveira E**, **Antunes P**, **Silva R**, **Coque TM**, **Peixe L**. 2013. Spread of multidrug-resistant Enterococcus to animals and humans: an underestimated role for the pig farm environment. J Antimicrob Chemother **68**:2746–54.

28. **Tedim AP**, **Ruiz-Garbajosa P**, **Corander J**, **Rodríguez CM**, **Cantón R**, **Willems RJ**, **Baquero F**, **Coque TM**. 2015. Population Biology of Intestinal Enterococcus Isolates from Hospitalized and Nonhospitalized Individuals in Different Age Groups. Appl Environ Microbiol **81**:1820–1831.

29. **Willems RJL**, **Hanage WP**, **Bessen DE**, **Feil EJ**. 2011. Population biology of Gram-positive pathogens: high-risk clones for dissemination of antibiotic resistance. FEMS Microbiol Rev **35**:872–900.

30. **Palmer KL**, **Godfrey P**, **Griggs A**, **Kos VN**, **Zucker J**, **Desjardins C**, **Cerqueira G**, **Gevers D**, **Walker S**, **Wortman J**, **Feldgarden M**, **Haas B**, **Birren B**, **Gilmore MS**. 2012. Comparative Genomics of Enterococci: Variation in Enterococcus faecalis, Clade Structure in E. faecium, and Defining Characteristics of E. gallinarum and E. casseliflavus. MBio **3**:e00318–11–e00318–11.

31. **Raven KE**, **Reuter S**, **Gouliouris T**, **Reynolds R**, **Russell JE**, **Brown NM**, **Török ME**, **Parkhill J**, **Peacock SJ**. 2016. Genome-based characterization of hospital-adapted Enterococcus faecalis lineages. Nat Microbiol **1**:15033.

32. **Corander J**, **Marttinen P**, **Sirén J**, **Tang J**. 2008. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. BMC Bioinformatics **9**:539.

33. **Freitas AR**, **Tedim AP**, **Francia M V.**, **Jensen LB**, **Novais C**, **Peixe L**, **Sánchez-Valenzuela A**, **Sundsfjord A**, **Hegstad K**, **Werner G**, **Sadowy E**, **Hammerum AM**, **Garcia-Migura L**, **Willems RJ**, **Baquero F**, **Coque TM**. 2016. Multilevel population genetic analysis of *vanA* and *vanB* *Enterococcus faecium* causing nosocomial outbreaks in 27 countries (1986–2012). J Antimicrob Chemother **71**:3351–3366.

34. **Willems RJL**, **Top J**, **van Schaik W**, **Leavis H**, **Bonten M**, **Sirén J**, **Hanage WP**, **Corander J**. 2012. Restricted gene flow among hospital subpopulations of Enterococcus faecium. MBio **3**:e00151–12.

35. **Clewell D**, **Weaver K**, **Dunny G**, **Coque T**, **Francia M**, **Hayes F**. 2014. Extrachromosomal and Mobile Elements in Enterococci: Transmission, Maintenance, and Epidemiology., p. . *In* Gilmore, M, Clewell, D, Ike, Y, Shankar, N (eds.), Enterococci: From Commensals to Leading Causes of Drug Resistant Infection. Massachusetts Eye and Ear Infirmary, Boston.

36. **Lanza VF**, **Tedim AP**, **Martínez JL**, **Baquero F**, **Coque TM**. 2015. The Plasmidome of Firmicutes: Impact on the Emergence and the Spread of Resistance to Antimicrobials. Microbiol Spectr. asm Pub2Web.

37. **Werner G**, **Coque TM**, **Franz CMAP**, **Grohmann E**, **Hegstad K**, **Jensen L**, **van Schaik W**, **Weaver K**. 2013. Antibiotic resistant enterococci-tales of a drug resistance gene trafficker. Int J Med Microbiol **303**:360–79.

38. **Hegstad K**, **Mikalsen T**, **Coque TM**, **Werner G**, **Sundsfjord A**. 2010. Mobile genetic elements and their contribution to the emergence of antimicrobial resistant Enterococcus faecalis and Enterococcus faecium. Clin Microbiol Infect **16**:541–554.

39. **Jensen LB**, **Garcia-Migura L**, **Valenzuela AJS**, **Løhr M**, **Hasman H**, **Aarestrup FM**. 2010. A classification system for plasmids from enterococci and other Gram-positive bacteria. J Microbiol Methods **80**:25–43.

40. **Garcillán-Barcia MP**, **Francia MV**, **de la Cruz F**. 2009. The diversity of conjugative relaxases and its application in plasmid classification. FEMS Microbiol Rev **33**:657–87.

41. **Ehrenfeld EE**, **Clewell DB**. 1987. Transfer functions of the Streptococcus faecalis plasmid pAD1: organization of plasmid DNA encoding response to sex pheromone. J Bacteriol **169**:3473–81.

42. **Christie PJ**, **Korman RZ**, **Zahler SA**, **Adsit JC**, **Dunny GM**. 1987. Two conjugation systems associated with Streptococcus faecalis plasmid pCF10: identification of a conjugative transposon

that transfers between S. faecalis and Bacillus subtilis. J Bacteriol **169**:2529–36.

43. **Manson JM**, **Hancock LE**, **Gilmore MS**. 2010. Mechanism of chromosomal transfer of Enterococcus faecalis pathogenicity island, capsule, antimicrobial resistance, and other traits. Proc Natl Acad Sci **107**:12269–12274.

44. **Hong H-J**, **Hutchings MI**, **Buttner MJ**, **Biotechnology and Biological Sciences Research Council, UK**. 2008. Vancomycin resistance VanS/VanR two-component systems. Adv Exp Med Biol **631**:200–13.

45. **Shaw JH**, **Clewell DB**. 1985. Complete nucleotide sequence of macrolide-lincosamide-streptogramin B-resistance transposon Tn917 in Streptococcus faecalis. J Bacteriol **164**:782–96.

46. **Hall AD**, **Steed ME**, **Arias CA**, **Murray BE**, **Rybak MJ**. 2012. Evaluation of standard- and high-dose daptomycin versus linezolid against vancomycin-resistant Enterococcus isolates in an in vitro pharmacokinetic/pharmacodynamic model with simulated endocardial vegetations. Antimicrob Agents Chemother **56**:3174–80.

47. **Roberts AP**, **Mullany P**. 2011. Tn916-like genetic elements: a diverse group of modular mobile elements conferring antibiotic resistance. FEMS Microbiol Rev **35**:856–71.

48. **Galloway-Peña JR**, **Bourgogne A**, **Qin X**, **Murray BE**. 2011. Diversity of the fsr-gelE region of the Enterococcus faecalis genome but conservation in strains with partial deletions of the fsr operon. Appl Environ Microbiol **77**:442–51.

49. **Solheim M**, **Brekke MC**, **Snipen LG**, **Willems RJL**, **Nes IF**, **Brede DA**. 2011. Comparative genomic analysis reveals significant enrichment of mobile genetic elements and genes encoding surface structure-proteins in hospital-associated clonal complex 2 Enterococcus faecalis. BMC Microbiol **11**:3.

50. **Novais C**, **Freitas AR**, **Silveira E**, **Baquero F**, **Peixe L**, **Roberts AP**, **Coque TM**. 2012. Different genetic supports for the tet(S) gene in Enterococci. Antimicrob Agents Chemother **56**:6014–8.

51. **Agersø Y**, **Pedersen AG**, **Aarestrup FM**. 2006. Identification of Tn5397-like and Tn916-like transposons and diversity of the tetracycline resistance gene tet(M) in enterococci from humans, pigs and poultry. J Antimicrob Chemother **57**:832–9.

52. **Brouwer MSM**, **Mullany P**, **Roberts AP**. 2010. Characterization of the conjugative transposon Tn6000 from Enterococcus casseliflavus 664.1H1 (formerly Enterococcus faecium 664.1H1). FEMS Microbiol Lett **309**:71–6.

53. **Novais C**, **Freitas AR**, **Silveira E**, **Baquero F**, **Peixe L**, **Roberts AP**, **Coque TM**. 2012. A tet(S/M) hybrid from CTn6000 and CTn916 recombination. Microbiology **158**:2710–2711.

54. **Rice LB**, **Carias LL**, **Marshall S**, **Rudin SD**, **Hutton-Thomas R**. 2005. Tn5386, a novel Tn916-like mobile element in Enterococcus faecium D344R that interacts with Tn916 to yield a large genomic deletion. J Bacteriol **187**:6668–77.

55. **Novais C**, **Freitas AR**, **Silveira E**, **Baquero F**, **Peixe L**, **Roberts AP**, **Coque TM**. 2012. Different genetic supports for the tet(S) gene in Enterococci. Antimicrob Agents Chemother **56**:6014–8.

56. **Silveira E**, **Freitas AR**, **Antunes P**, **Barros M**, **Campos J**, **Coque TM**, **Peixe L**, **Novais C**. 2014. Co-transfer of resistance to high concentrations of copper and first-line antibiotics among Enterococcus from different origins (humans, animals, the environment and foods) and clonal lineages. J Antimicrob Chemother **69**:899–906.

57. **León-Sampedro R**, **Novais C**, **Peixe L**, **Baquero F**, **Coque TM**. 2016. Diversity and Evolution of the Tn *5801-tet* (M)-Like Integrative and Conjugative Elements among Enterococcus, Streptococcus, and Staphylococcus. Antimicrob Agents Chemother **60**:1736–1746.

58. **Singhal N**, **Kumar M**, **Kanaujia PK**, **Virdi JS**. 2015. MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. Front Microbiol **6**:791.

59. **Southern EM**. 1992. Detection of specific sequences among DNA fragments separated by gel electrophoresis. 1975. Biotechnology **24**:122–39.

60. **Weaver KE**. 2012. The par toxin-antitoxin system from Enterococcus faecalis plasmid pAD1 and its chromosomal homologs. RNA Biol **9**:1498–503.

61. **Murray BE**. 1990. The life and times of the Enterococcus. Clin Microbiol Rev **3**:46–65.

62. **MUNDT JO**. 1963. Occurrence of enterococci in animals in a wild environment. Appl Microbiol **11**:136–40.

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

64. **Raven KE**, **Reuter S**, **Gouliouris T**, **Reynolds R**, **Russell JE**, **Brown NM**, **Török ME**, **Parkhill J**, **Peacock SJ**. 2016. Genome-based characterization of hospital-adapted Enterococcus faecalis lineages. Nat Microbiol **1**:15033.

65. **Numminen E**, **Gutmann M**, **Shubin M**, **Marttinen P**, **Méric G**, **van Schaik W**, **Coque TM**, **Baquero**

F, **Willems RJL**, **Sheppard SK**, **Feil EJ**, **Hanage WP**, **Corander J**. 2016. The impact of host metapopulation structure on the population genetics of colonizing bacteria. J Theor Biol **396**:53–62.

66. **Lanza VF**, **Baquero F**, **de la Cruz F**, **Coque TM**. 2017. AcCNET (Accessory Genome Constellation Network): comparative genomics software for accessory genome analysis using bipartite networks. Bioinformatics **33**:283–285.

67. **CLSI**. 2014. Performance Standards for Antimicrobial Susceptibility Testing; Twenty-Second Informational Supplement. Clin Lab Stand Inst **34**:1–219.

68. **Barton BM**, **Harding GP**, **Zuccarelli AJ**. 1995. A general method for detecting and sizing large plasmids. Anal Biochem **226**:235–40.

69. **Freitas AR**, **Novais C**, **Tedim AP**, **Francia MV**, **Baquero F**, **Peixe L**, **Coque TM**. 2013. Microevolutionary Events Involving Narrow Host Plasmids Influences Local Fixation of Vancomycin-Resistance in Enterococcus Populations. PLoS One **8**:e60589.

70. **Wardal E**, **Gawryszewska I**, **Hryniewicz W**, **Sadowy E**. 2013. Abundance and diversity of plasmid-associated genes among clinical isolates of Enterococcus faecalis. Plasmid.

71. **Moritz EM**, **Hergenrother PJ**. 2007. Toxin-antitoxin systems are ubiquitous and plasmid-encoded in vancomycin-resistant enterococci. Proc Natl Acad Sci U S A **104**:311–6.

72. **Lanza VF**, **de Toro M**, **Garcillán-Barcia MP**, **Mora A**, **Blanco J**, **Coque TM**, **de la Cruz F**. 2014. Plasmid Flux in Escherichia coli ST131 Sublineages, Analyzed by Plasmid Constellation Network (PLACNET), a New Method for Plasmid Reconstruction from Whole Genome Sequences. PLoS Genet **10**:e1004766.

73. **Antipov D**, **Hartwick N**, **Shen M**, **Raiko M**, **Lapidus A**, **Pevzner PA**. 2016. plasmidSPAdes: assembling plasmids from whole genome sequencing data. Bioinformatics **32**:3380–3387.

74. **Seemann T**. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics **30**:2068–2069.

75. **Sullivan MJ**, **Petty NK**, **Beatson SA**. 2011. Easyfig: a genome comparison visualizer. Bioinformatics **27**:1009–1010.

76. **Novais C**, **Freitas A.**, **Francia MV**, **Peixe LV**, **Baquero F**, **Coque TM**. 2009. Characterization of Gram-positive conjugative transposons among international epidemic vancomycin-resistant isolates of Enterococcus faecalis and Enterococcus faecium, Abstract P949, p. S247. *In* 19th European Congress Clinical Microbiology Infections Diseases, Helsinki, Finland.

77. **Bankevich A**, **Nurk S**, **Antipov D**, **Gurevich AA**, **Dvorkin M**, **Kulikov AS**, **Lesin VM**, **Nikolenko SI**, **Pham S**, **Prjibelski AD**, **Pyshkin A V.**, **Sirotkin A V.**, **Vyahhi N**, **Tesler G**, **Alekseyev MA**, **Pevzner PA**. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol **19**:455–477.

78. **Gurevich A**, **Saveliev V**, **Vyahhi N**, **Tesler G**. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics **29**:1072–1075.

79. **Price MN**, **Dehal PS**, **Arkin AP**. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS One **5**:e9490.

80. **Yu G**, **Smith DK**, **Zhu H**, **Guan Y**, **Lam TT-Y**. 2017. ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol **8**:28–36.

81. **Bakshi U**, **Sarkar M**, **Paul S**, **Dutta C**. 2016. Assessment of virulence potential of uncharacterized Enterococcus faecalis strains using pan genomic approach – Identification of pathogen–specific and habitat-specific genes. Sci Rep **6**:38648.

82. **Fischer S**, **Brunk BP**, **Chen F**, **Gao X**, **Harb OS**, **Iodice JB**, **Shanmugam D**, **Roos DS**, **Stoeckert CJ**, **Jr.** 2011. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. Curr Protoc Bioinforma **Chapter 6**:Unit 6.12.1–19.

83. **Schwarz F V.**, **Perreten V**, **Teuber M**. 2001. Sequence of the 50-kb Conjugative Multiresistance Plasmid pRE25 from Enterococcus faecalis RE25. Plasmid **46**:170–187.

84. **Tedim AP**, **Ruiz-Garbajosa P**, **Corander J**, **Rodríguez CM**, **Cantón R**, **Willems RJ**, **Baquero F**, **Coque TM**. 2015. Population Biology of Intestinal Enterococcus Isolates from Hospitalized and Nonhospitalized Individuals in Different Age Groups. Appl Environ Microbiol **81**:1820–1831.

85. **Lebreton F**, **van Schaik W**, **Manson McGuire A**, **Godfrey P**, **Griggs A**, **Mazumdar V**, **Corander J**, **Cheng L**, **Saif S**, **Young S**, **Zeng Q**, **Wortman J**, **Birren B**, **Willems RJL**, **Earl AM**, **Gilmore MS**. 2013. Emergence of Epidemic Multidrug-Resistant Enterococcus faecium from Animal and Commensal Strains. MBio **4**:e00534–13–e00534–13.

86. **Morris JJ**, **Lenski RE**, **Zinser ER**. 2012. The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss. MBio **3**:e00036–12–e00036–12.

87. **Da Cunha V**, **Davies MR**, **Douarre P-E**, **Rosinski-Chupin I**, **Margarit I**, **Spinali S**, **Perkins T**, **Lechat P**, **Dmytruk N**, **Sauvage E**, **Ma L**, **Romi B**, **Tichit M**, **Lopez-Sanchez M-J**, **Descorps-Declere S**, **Souche E**, **Buchrieser C**, **Trieu-Cuot P**, **Moszer I**, **Clermont D**, **Maione D**, **Bouchier C**,

McMillan DJ, **Parkhill J**, **Telford JL**, **Dougan G**, **Walker MJ**, **Holden MTG**, **Poyart C**, **Glaser P**. 2014. Streptococcus agalactiae clones infecting humans were selected and fixed through the extensive use of tetracycline. Nat Commun **5**:4544.

88. **Edmond KM**, **Kortsalioudaki C**, **Scott S**, **Schrag SJ**, **Zaidi AKM**, **Cousens S**, **Heath PT**. 2012. Group B streptococcal disease in infants aged younger than 3 months: systematic review and meta-analysis. Lancet (London, England) **379**:547–56.

89. **de Vries LE**, **Christensen H**, **Skov RL**, **Aarestrup FM**, **Agersø Y**. 2009. Diversity of the tetracycline resistance gene tet(M) and identification of Tn916- and Tn5801-like (Tn6014) transposons in Staphylococcus aureus from humans and animals. J Antimicrob Chemother **64**:490–500.

90. **Soge OO**, **Beck NK**, **White TM**, **No DB**, **Roberts MC**. 2008. A novel transposon, Tn6009, composed of a Tn916 element linked with a Staphylococcus aureus mer operon. J Antimicrob Chemother **62**:674–80.

91. **Rice LB**. 1998. Tn916 family conjugative transposons and dissemination of antimicrobial resistance determinants. Antimicrob Agents Chemother **42**:1871–7.

92. **Clewell DB**, **Weaver KE**, **Dunny GM**, **Coque TM**, **Francia MV**, **Hayes F**. 2014. Extrachromosomal and Mobile Elements in Enterococci: Transmission, Maintenance, and EpidemiologyEnterococci: From Commensals to Leading Causes of Drug Resistant Infection. Massachusetts Eye and Ear Infirmary, Boston.

93. **Hegstad K**, **Mikalsen T**, **Coque TM**, **Werner G**, **Sundsfjord A**. 2010. Mobile genetic elements and their contribution to the emergence of antimicrobial resistant Enterococcus faecalis and Enterococcus faecium. Clin Microbiol Infect **16**:541–54.

94. **Smyth DS**, **McDougal LK**, **Gran FW**, **Manoharan A**, **Enright MC**, **Song J-H**, **de Lencastre H**, **Robinson DA**. 2010. Population structure of a hybrid clonal group of methicillin-resistant Staphylococcus aureus, ST239-MRSA-III. PLoS One **5**:e8582.

95. **Smyth DS**, **Robinson DA**. 2009. Integrative and sequence characteristics of a novel genetic element, ICE6013, in Staphylococcus aureus. J Bacteriol **191**:5964–75.

96. **Song L**, **Pan Y**, **Chen S**, **Zhang X**. 2012. Structural characteristics of genomic islands associated with GMP synthases as integration hotspot among sequenced microbial genomes. Comput Biol Chem **36**:62–70.

97. **Wipf JRK**, **Schwendener S**, **Nielsen JB**, **Westh H**, **Perreten V**. 2015. The new macrolide-lincosamide-streptogramin B resistance gene erm(45) is located within a genomic island in

Staphylococcus fleurettii. Antimicrob Agents Chemother **59**:3578–81.

98. **Boyd DA**, **Cabral T**, **Van Caeseele P**, **Wylie J**, **Mulvey MR**. 2002. Molecular characterization of the vanE gene cluster in vancomycin-resistant Enterococcus faecalis N00-410 isolated in Canada. Antimicrob Agents Chemother **46**:1977–9.

99. **Kuroda M**, **Ohta T**, **Uchiyama I**, **Baba T**, **Yuzawa H**, **Kobayashi I**, **Cui L**, **Oguchi A**, **Aoki K**, **Nagai Y**, **Lian J**, **Ito T**, **Kanamori M**, **Matsumaru H**, **Maruyama A**, **Murakami H**, **Hosoyama A**, **Mizutani-Ui Y**, **Takahashi NK**, **Sawano T**, **Inoue R**, **Kaito C**, **Sekimizu K**, **Hirakawa H**, **Kuhara S**, **Goto S**, **Yabuzaki J**, **Kanehisa M**, **Yamashita A**, **Oshima K**, **Furuya K**, **Yoshino C**, **Shiba T**, **Hattori M**, **Ogasawara N**, **Hayashi H**, **Hiramatsu K**. 2001. Whole genome sequencing of meticillin-resistant Staphylococcus aureus. Lancet **357**:1225–40.

100. **Roberts AP**, **Johanesen PA**, **Lyras D**, **Mullany P**, **Rood JI**. 2001. Comparison of Tn5397 from Clostridium difficile, Tn916 from Enterococcus faecalis and the CW459tet(M) element from Clostridium perfringens shows that they have similar conjugation regions but different insertion and excision modules. Microbiology **147**:1243–51.

101. **Yamamoto T**, **Takano T**, **Higuchi W**, **Iwao Y**, **Singur O**, **Reva I**, **Otsuka Y**, **Nakayashiki T**, **Mori H**, **Reva G**, **Kuznetsov V**, **Potapov V**. 2012. Comparative Genomics and Drug Resistance of a Geographic Variant of ST239 Methicillin-Resistant Staphylococcus aureus Emerged in Russia. PLoS One **7**:e29187.

102. **Brenciani A**, **Tiberi E**, **Tili E**, **Mingoia M**, **Palmieri C**, **Varaldo PE**, **Giovanetti E**. 2014. Genetic determinants and elements associated with antibiotic resistance in viridans group streptococci. J Antimicrob Chemother **69**:1197–204.

103. **Denapaite D**, **Brückner R**, **Nuhn M**, **Reichmann P**, **Henrich B**, **Maurer P**, **Schähle Y**, **Selbmann P**, **Zimmermann W**, **Wambutt R**, **Hakenbeck R**. 2010. The genome of Streptococcus mitis B6--what is a commensal? PLoS One **5**:e9426.

104. **Mingoia M**, **Morici E**, **Tili E**, **Giovanetti E**, **Montanari MP**, **Varaldo PE**. 2013. Characterization of Tn5801.Sag, a Variant of Staphylococcus aureus Tn916 Family Transposon Tn5801 That Is Widespread in Clinical Isolates of Streptococcus agalactiae. Antimicrob Agents Chemother **57**:4570–4.

105. **Carver TJ**, **Rutherford KM**, **Berriman M**, **Rajandream M-A**, **Barrell BG**, **Parkhill J**. 2005. ACT: the Artemis Comparison Tool. Bioinformatics **21**:3422–3.

106. **Tamura K**, **Stecher G**, **Peterson D**, **Filipski A**, **Kumar S**. 2013. MEGA6: Molecular Evolutionary

Genetics Analysis version 6.0. Mol Biol Evol **30**:2725–9.

107. **Larsen M V**, **Cosentino S**, **Rasmussen S**, **Friis C**, **Hasman H**, **Marvig RL**, **Jelsbak L**, **Sicheritz-Pontén T**, **Ussery DW**, **Aarestrup FM**, **Lund O**. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol **50**:1355–61.

108. **Aarestrup FM**, **Agerso Y**, **Gerner-Smidt P**, **Madsen M**, **Jensen LB**. 2000. Comparison of antimicrobial resistance phenotypes and resistance genes in Enterococcus faecalis and Enterococcus faecium from humans in the community, broilers, and pigs in Denmark. Diagn Microbiol Infect Dis **37**:127–37.

109. **Tenover FC**, **Arbeit RD**, **Goering R V**, **Mickelsen PA**, **Murray BE**, **Persing DH**, **Swaminathan B**. 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J Clin Microbiol **33**:2233–9.

110. **Coque TM**, **Willems RJL**, **Fortún J**, **Top J**, **Diz S**, **Loza E**, **Cantón R**, **Baquero F**. 2005. Population structure of Enterococcus faecium causing bacteremia in a Spanish university hospital: setting the scene for a future increase in vancomycin resistance? Antimicrob Agents Chemother **49**:2693–700.

111. **Freitas AR**, **Tedim AP**, **Novais C**, **Ruiz-Garbajosa P**, **Werner G**, **Laverde-Gomez JA**, **Cantón R**, **Peixe L**, **Baquero F**, **Coque TM**. 2010. Global spread of the hyl(Efm) colonization-virulence gene in megaplasmids of the Enterococcus faecium CC17 polyclonal subcluster. Antimicrob Agents Chemother **54**:2660–5.

112. **Werner G**, **Freitas AR**, **Coque TM**, **Sollid JE**, **Lester C**, **Hammerum AM**, **Garcia-Migura L**, **Jensen LB**, **Francia M V**, **Witte W**, **Willems RJ**, **Sundsfjord A**. 2011. Host range of enterococcal vanA plasmids among Gram-positive intestinal bacteria. J Antimicrob Chemother **66**:273–82.

113. **Bergmann R**, **Sagar V**, **Nitsche-Schmitz DP**, **Chhatwal GS**. 2012. first detection of trimethoprim resistance determinant dfrG in Streptococcus pyogenes clinical isolates in India. Antimicrob Agents Chemother **56**:5424–5.

114. **Sekiguchi J**, **Tharavichitkul P**, **Miyoshi-Akiyama T**, **Chupia V**, **Fujino T**, **Araake M**, **Irie A**, **Morita K**, **Kuratsuji T**, **Kirikae T**. 2005. Cloning and characterization of a novel trimethoprim-resistant dihydrofolate reductase from a nosocomial isolate of Staphylococcus aureus CM.S2 (IMCJ1454). Antimicrob Agents Chemother **49**:3948–51.

115. **Bergmann R**, **van der Linden M**, **Chhatwal GS**, **Nitsche-Schmitz DP**. 2014. Factors that cause trimethoprim resistance in Streptococcus pyogenes. Antimicrob Agents Chemother **58**:2281–8.

116. **Rice LB**, **Lakticová V**, **Carias LL**, **Rudin S**, **Hutton R**, **Marshall SH**. 2009. Transferable capacity for

gastrointestinal colonization in Enterococcus faecium in a mouse model. J Infect Dis **199**:342–9.

117. **Douarre P-E**, **Sauvage E**, **Poyart C**, **Glaser P**. 2015. Host specificity in the diversity and transfer of lsa resistance genes in group B Streptococcus. J Antimicrob Chemother **70**:dkv277.

118. **Malbruny B**, **Werno AM**, **Murdoch DR**, **Leclercq R**, **Cattoir V**. 2011. Cross-resistance to lincosamides, streptogramins A, and pleuromutilins due to the lsa(C) gene in Streptococcus agalactiae UCN70. Antimicrob Agents Chemother **55**:1470–4.

119. **Day T**, **Huijben S**, **Read AF**. 2015. Is selection relevant in the evolutionary emergence of drug resistance? Trends Microbiol **23**:126–133.

120. **Roberts GA**, **Chen K**, **Bower EKM**, **Madrzak J**, **Woods A**, **Barker AM**, **Cooper LP**, **White JH**, **Blakely GW**, **Manfield I**, **Dryden DTF**. 2013. Mutations of the domain forming the dimeric interface of the ArdA protein affect dimerization and antimodification activity but not antirestriction activity. FEBS J **280**:4903–14.

121. **Roberts GA**, **Houston PJ**, **White JH**, **Chen K**, **Stephanou AS**, **Cooper LP**, **Dryden DTF**, **Lindsay JA**. 2013. Impact of target site distribution for Type I restriction enzymes on the evolution of methicillin-resistant Staphylococcus aureus (MRSA) populations. Nucleic Acids Res **41**:7472–84.

122. **Sadykov MR**. 2014. Restriction–Modification Systems as a Barrier for Genetic Manipulation of Staphylococcus aureus, p. 9–23. *In* Bose, JL (ed.), The Genetic Manipulation of Staphylococci. Springer New York, New York.

123. **Worby CJ**, **Lipsitch M**, **Hanage WP**. 2015. Shared genomic variants: identification of transmission routes using pathogen deep sequence data. bioRxiv 032458.

124. **Lassalle F**, **Périan S**, **Bataillon T**, **Nesme X**, **Duret L**, **Daubin V**. 2015. GC-Content evolution in bacterial genomes: the biased gene conversion hypothesis expands. PLoS Genet **11**:e1004941.

125. **Rice LB**, **Carias LL**, **Marshall SH**, **Hutton-Thomas R**, **Rudin S**. 2007. Characterization of Tn5386, a Tn916-related mobile element. Plasmid **58**:61–7.

126. **Palmieri C**, **Mingoia M**, **Massidda O**, **Giovanetti E**, **Varaldo PE**. 2012. Streptococcus pneumoniae transposon Tn1545/Tn6003 changes to Tn6002 due to spontaneous excision in circular form of the erm(B)- and aphA3-containing macrolide-aminoglycoside-streptothricin (MAS) element. Antimicrob Agents Chemother **56**:5994–7.

127. **Maki H**, **Murakami K**. 1997. Formation of potent hybrid promoters of the mutant llm gene by IS256 transposition in methicillin-resistant Staphylococcus aureus. J Bacteriol **179**:6944–8.

128. **Benson MA**, **Ohneck EA**, **Ryan C**, **Alonzo F**, **Smith H**, **Narechania A**, **Kolokotronis S-O**, **Satola SW**, **Uhlemann A-C**, **Sebra R**, **Deikus G**, **Shopsin B**, **Planet PJ**, **Torres VJ**. 2014. Evolution of hypervirulence by a MRSA clone through acquisition of a transposable element. Mol Microbiol **93**:664–81.

129. **McEvoy CRE**, **Tsuji B**, **Gao W**, **Seemann T**, **Porter JL**, **Doig K**, **Ngo D**, **Howden BP**, **Stinear TP**. 2013. Decreased vancomycin susceptibility in Staphylococcus aureus caused by IS256 tempering of WalKR expression. Antimicrob Agents Chemother **57**:3240–9.

130. **Fléchard M**, **Gilot P**. 2014. Physiological impact of transposable elements encoding DDE transposases in the environmental adaptation of Streptococcus agalactiae. Microbiology **160**:1298–315.

131. **Bellanger X**, **Payot S**, **Leblond-Bourget N**, **Guédon G**. 2014. Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. FEMS Microbiol Rev **38**:720–60.

132. **Bellanger X**, **Roberts AP**, **Morel C**, **Choulet F**, **Pavlovic G**, **Mullany P**, **Decaris B**, **Guédon G**. 2009. Conjugative transfer of the integrative conjugative elements ICESt1 and ICESt3 from Streptococcus thermophilus. J Bacteriol **191**:2764–75.

133. **Burrus V**, **Pavlovic G**, **Decaris B**, **Guédon G**. 2002. Conjugative transposons: the tip of the iceberg. Mol Microbiol **46**:601–10.

134. **Roberts AP**, **Mullany P**. 2009. A modular master on the move: the Tn916 family of mobile genetic elements. Trends Microbiol **17**:251–8.

135. **de Vries LE**, **Hasman H**, **Jurado Rabadán S**, **Agersø Y**. 2016. Sequence-Based Characterization of Tn5801-Like Genomic Islands in Tetracycline-Resistant Staphylococcus pseudintermedius and Other Gram-positive Bacteria from Humans and Animals. Front Microbiol **7**:576.

136. **Clewell DB**, **Flannagan SE**, **Jaworski DD**. 1995. Unconstrained bacterial promiscuity: the Tn916-Tn1545 family of conjugative transposons. Trends Microbiol **3**:229–36.

137. **Ramsay JP**, **Firth N**. 2017. Diverse mobilization strategies facilitate transfer of non-conjugative mobile genetic elements. Curr Opin Microbiol **38**:1–9.

138. **Tatusov RL**, **Galperin MY**, **Natale DA**, **Koonin E V**. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res **28**:33–6.

139. **Wright LD**, **Grossman AD**. 2016. Autonomous Replication of the Conjugative Transposon Tn *916*. J Bacteriol **198**:3355–3366.

140. **Burrus V**, **Pavlovic G**, **Decaris B**, **Guédon G**. 2002. The ICESt1 element of Streptococcus thermophilus belongs to a large family of integrative and conjugative elements that exchange modules and change their specificity of integration. Plasmid **48**:77–97.

141. **Brouwer MSM**, **Warburton PJ**, **Roberts AP**, **Mullany P**, **Allan E**. 2011. Genetic organisation, mobility and predicted functions of genes on integrated, mobile genetic elements in sequenced strains of Clostridium difficile. PLoS One **6**:e23014.

142. **de Kraker MEA**, **Jarlier V**, **Monen JCM**, **Heuer OE**, **van de Sande N**, **Grundmann H**. 2013. The changing epidemiology of bacteraemias in Europe: trends from the European Antimicrobial Resistance Surveillance System. Clin Microbiol Infect **19**:860–868.

143. **Ruiz-Garbajosa P**, **Cantón R**, **Pintado V**, **Coque TM**, **Willems R**, **Baquero F**, **del Campo R**. 2006. Genetic and phenotypic differences among Enterococcus faecalis clones from intestinal colonisation and invasive disease. Clin Microbiol Infect **12**:1193–8.

Appendix 2. Diversity and Evolution of the Tn*5801-tet*(M)-like Integrative-Conjugative Elements Among *Enterococcus*, *Streptococcus* and *Staphylococcus*

Appendix 2. Diversity and Evolution of the Tn*5801-tet*(M)-like Integrative- Conjugative Elements Among *Enterococcus*, *Streptococcus* and *Staphylococcus*. León Sampedro *et al*. AAC 2016

AMERICAN SOCIETY FOR MICROBIOLOGY | Antimicrobial Agents and Chemotherapy

Diversity and Evolution of the Tn*5801*-*tet*(M)-Like Integrative and Conjugative Elements among *Enterococcus*, *Streptococcus*, and *Staphylococcus*

Ricardo León-Sampedro,[a,d] Carla Novais,[b] Luísa Peixe,[b] Fernando Baquero,[a,c,d] Teresa M. Coque[a,c,d]

Servicio de Microbiología, Hospital Universitario Ramón y Cajal, Madrid, Spain[a]; UCIBIO/REQUIMTE, Departamento de Ciências Biológicas, Faculdade de Farmácia, Universidade do Porto, Porto, Portugal[b]; Unidad de Resistencia a Antibióticos y Virulencia Bacteriana asociada al Consejo Superior de Investigaciones Científicas, Madrid, Spain[c]; CIBER Epidemiología y Salud Pública, Madrid, Spain[d]

This work describes the diversity and evolution of Tn*5801* among enterococci, staphylococci, and streptococci based on analysis of the 5,073 genomes of these bacterial groups available in gene databases. We also examined 610 isolates of *Enterococcus* (from 10 countries, 1987 to 2010) for the presence of this and other known CTn-*tet*(M) elements due to the scarcity of data about Tn*5801* among enterococci. Genome location (by I*Ceu*-I–pulsed-field gel electrophoresis [PFGE] hybridization/integration site identification), conjugation and fitness (by standard methods), Tn*5801* characterization (by long-PCR mapping/sequencing), and clonality (by PFGE/multilocus sequence typing [MLST]) were studied. Twenty-three Tn*5801* variants (17 unpublished) clustered in two groups, designated "A" (25 kb; *n* = 14; predominant in *Staphylococcus aureus*) and "B" (20 kb; *n* = 9; predominant in *Streptococcus agalactiae*). The percent GC content of the common backbone suggests a streptococcal origin of Tn*5801* group B, with further acquisition of a 5-kb fragment that resulted in group A. Deep sequence analysis allowed identification of variants associated with clonal lineages of *S. aureus* (clonal complex 8 [CC8], sequence type 239 [ST239]), *S. agalactiae* (CC17), *Enterococcus faecium* (ST17/ST18), or *Enterococcus faecalis* (ST8), local variants, or variants located in different species and geographical areas. All Tn*5801* elements were chromosomally located upstream of the *guaA* gene, which serves as an integration hot spot. Transferability was demonstrated only for Tn*5801* type B among *E. faecalis* clonal backgrounds, which eventually harbored another Tn*5801* copy. The study documents early acquisition of Tn*5801* by *Enterococcus*, *Staphylococcus*, and *Streptococcus*. Clonal waves of these pathogens seem to have contributed to the geographical spread and local evolution of the transposon. Horizontal transfer, also demonstrated, could explain the variability observed, with the isolates often containing sequences of different origins.

The extended use of tetracycline since its introduction into the therapeutic arsenal in 1948 seems to have resulted in an evolutionary bottleneck in the population structure of some genera of *Firmicutes* that are of interest in human health. Recent studies have associated the acquisition and further fixation of tetracycline resistance (Tet[r]) by pathogenic clones of group B *Streptococcus* (GBS) with the global increase of high-mortality GBS neonatal infections in the last decades (1, 2). Similarly, some *Staphylococcus aureus* lineages are enriched in Tet[r] elements (3–5).

Tetracycline resistance in major Gram-positive human opportunistic pathogens is caused mainly by the acquisition of integrative and conjugative elements (ICE) of the Tn*916* family carrying the *tet*(M) gene (5–7). These elements display a common synteny but differ in the integrase (*int*) and the excisionase (*xis*) sequences, insertion site specificity, and host range (8, 9). Variants of some Tn*916*-*tet*(M) members, namely Tn*916*, Tn*5397*, Tn*6000*, or Tn*5801*, are widely spread among several genera of *Firmicutes*, apparently suggesting a successful multihost dissemination of these elements (7).

Tn*5801* has a site-specific tyrosine recombinase that recognizes the 3′ end of the GMP synthase gene (*guaA*) (10, 11), an insertion hot spot of genomic islands coding for pathogenicity or antibiotic resistance in different *Firmicutes* (12–14). Tn*5801* was originally detected in the vancomycin-resistant *Staphylococcus aureus* (VRSA) clinical strain Mu50 recovered in Japan in 1997 (15), but several studies have documented the presence of platforms highly similar to Tn*5801* in early isolates of *Streptococcus agalactiae* (France, 1953) (1), *S. aureus* (Denmark, 1963; designated Tn*6014*) (3), and *Clostridium perfringens* (United States, 1977; a truncated element designated CW459*tet*(M)] (16). The diversity of Tn*5801*-like backbones in a few isolates of different species (3, 17–20) collected throughout more than 50 years suggests the evolutionary interplay of selective events, horizontal dissemination, and the genetic plasticity of this transposon, allowing its spread in different clonal backgrounds and environments. However, little is known about Tn*5801*, at either the molecular or the epidemiological level.

In this study, we created a data set comprising all the genomes of *Firmicutes* with the presence of a Tn*5801*-like ICE. Due to the relevance of enterococci in the spread of Tet[r] and the low number

Received 6 August 2015 **Returned for modification** 7 September 2015
Accepted 27 December 2015

Accepted manuscript posted online 4 January 2016

Citation León-Sampedro R, Novais C, Peixe L, Baquero F, Coque TM. 2016. Diversity and evolution of the Tn*5801*-*tet*(M)-like integrative and conjugative elements among *Enterococcus*, *Streptococcus*, and *Staphylococcus*. Antimicrob Agents Chemother 60:1736–1746. doi:10.1128/AAC.01864-15.

Address correspondence to Teresa M. Coque, mariateresa.coque@salud.madrid.org.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/AAC.01864-15.

of sequenced genomes in public databases at the time this study was started, we also analyzed the presence and genetic background of Tn*5801* in a large collection of enterococcal strains from different geographical areas as well as its transferability. This comprehensive phylogenetic and genomic analysis allowed documentation of the genetic variability of Tn*5801*-like elements among genomes of enterococci, staphylococci, streptococci, and other bacterial species through years and across continents. The results highlight the relevance of comprehensive analysis at local and global levels to accurately establish the transmission and evolvability of elements at different scales. It also proved the transferability of the element in different genetic contexts.

MATERIALS AND METHODS

Screening of Tn*5801* in GenBank databases and creation of data sets. The DNA sequence from *S. aureus* Mu50 (GenBank accession number BA000017) was used as a reference to screen for the presence of Tn*5801* among 5,073 draft and complete genomes from *Firmicutes* (4,130 *S. aureus*, 345 *Enterococcus faecalis*, 301 *S. agalactiae*, 254 *Enterococcus faecium*, 23 *Streptococcus mitis*, 14 *Lactococcus garvieae*, 4 *Staphylococcus pseudintermedius*, 2 *Enterococcus villorum*, and 1 *Streptococcus thoraltensis*) obtained from the NCBI archive (last updated January 2015). BLAST searches identified different sequences showing homology with Tn*5801*-like elements. They were further stored and analyzed by using the MUSCLE (multiple-sequence comparison by log expectation) algorithm (http://www.ebi.ac.uk/Tools/msa/muscle/) and Artemis comparison tool (21) (http://www.sanger.ac.uk/resources/software/act/). ICE maps were created with Vector NTI (ThermoFisher Scientific) and Inkscape (http://inkscape.org/) software.

Phylogenetic analyses were accomplished using MEGA version 6 (22). A common region of 13,895 bp was identified among all the sequences with homology with Tn*5801* (here called "Tn*5801* variant sequences"). The region comprises 16 open reading frames (ORFs) (int_{Tn5801}-*orf*24; *orf*22, *orf*18, *orf*16, *orf*15, *orf*12-10, and *orf*8-4), which were trimmed and concatenated to be further analyzed. A maximum-likelihood phylogenetic tree was constructed with 1,000 bootstrap replications in order to infer the relationship among all Tn*5801* variants. Due to the identification of modules with different GC contents and with recombination landmarks, the Tn*5801* variants were further split into five core fragments (int_{Tn5801}, *orf*26-19, *orf*18-14, *orf*12-8, and *orf*7-4), with each fragment being separately analyzed as mentioned before. The *orf*26-19 fragment was excluded from this analysis because of the variability observed among variants due to frequent indels and recombinatorial events. An arbitrary number was attributed to different sequences of each core fragment, and Tn*5801* sequences with unique number profiles were used to calculate a consensus tree. The data set created included epidemiological information for isolates (host/source, year of isolation, country, and sequence type [ST]). To identify the STs of the isolates, we used the MLST 1.8 software of the Center of Genomic Epidemiology server (23).

Epidemiological background of bacterial strains and identification of *tet*(M) transposons among enterococci. A collection of 610 enterococcal isolates from humans ($n = 320$; hospitalized patients [HP], $n = 195$; healthy human volunteers [HV], $n = 125$); animals ($n = 236$; poultry [P], $n = 164$; swine/piggeries [SP], $n = 72$), and hospital/urban sewage (SW) ($n = 54$) collected in different countries over a 23-year period (1987 to 2010) was included (9, 24). All isolates were tested for susceptibility to tetracycline by the disk diffusion and/or agar dilution method following CLSI guidelines (25). The presence of genes coding for tetracycline resistance [*tet*(M), *tet*(L), *tet*(S), *tet*(K), and *tet*(O)] was assayed by PCR as previously described (26). The presence of the main transposons of the Tn*916* family (Tn*5801*, Tn*916*, Tn*5397*, Tn*5386*, and Tn*6000*) was investigated with a multiplex PCR assay designed here for detecting integrases, excisionases, and other specific sequences of known transposons. The susceptibility and specificity of the assay were validated using appropriate controls. Primers and conditions are listed in Table S1 in the supplemental material. The clonal relationship among *E. faecium* and *E. faecalis* isolates harboring the integrase of Tn*5801* (int_{Tn5801}) was established by pulsed-field gel electrophoresis (PFGE) and multilocus sequence typing (MLST) (http://pubmlst.org/) (27, 28). Population genetic analysis was performed using BAPS software with *E. faecium* isolates as previously described (29, 30).

PCR characterization of Tn*5801* backbones in *Enterococcus*. The Tn*5801* backbone was fully characterized by a PCR mapping assay based on the Tn*5801* sequence of *S. aureus* Mu50 (GenBank accession number BA000017). Amplified fragments obtained from positive int_{Tn5801} strains were further sequenced (Table 1; see Fig. S1 in the supplemental material). Genomic location of Tn*5801* was assessed by hybridization of I-*Ceu*I (TaKaRa Biotechnology, Dalian, China) digested genomic DNA using specific int_{Tn5801} and 23S rRNA probes (31). Amplification of the int_{Tn5801}-*guaA* region was performed in order to characterize the integration site of Tn*5801* (Table 1; see Fig. S1 in the supplemental material) (14).

Transferability of Tn*5801*. Transferability of Tn*5801* was screened by filter mating using different recipient strains of *E. faecalis* (JH2-2, OG1RF, and OG1SSp) and *E. faecium* (64/3, BM4105RF, and BM4105SS), all being resistant to rifampin and fusidic acid (designated RF) or streptomycin and spectinomycin (designated SS) and negative for the presence of the *tet*(M) gene (32). The mating assays were performed at 37°C for 36 h using aliquots of broth cultures of donors and recipients grown up to the exponential phase (McFarland 0.5) and mixed at a ratio of 1:1 (32). Transconjugants were selected on brain heart infusion (BHI) agar (Pronadisa, Laboratorios Conda, SA, Madrid, Spain) plates supplemented with tetracycline (10 mg/liter) plus rifampin (30 mg/liter) and fusidic acid (25 mg/liter) or streptomycin (250 mg/liter) and spectinomycin (250 mg/liter) (Sigma-Aldrich, Inc., St. Louis, MO). Unexpectedly, we identified a Tn*5801* Δ*tet*(M) variant in the laboratory recipient *E. faecalis* strain JH2-2, which is an emblematic receptor strain (see below). Secondary filter mating assays were performed using *E. faecalis* JH2-2 transconjugants as donors and the *E. faecalis* OG1SS and *E. faecium* BM4105SS strains as recipients. Transconjugants were characterized according to the susceptibility to tetracycline (Etest; bioMérieux SA, France), the presence of Tn*5801* [int_{Tn5801} and *tet*(M)], the transposon backbone (PCR overlapping), and comparison of SmaI-digested DNA PFGE profiles of transconjugants with those of recipients and wild-type strains. The location of Tn*5801* in the recipient genomes was assessed by hybridization of SmaI-digested DNA with specific *guaA*, int_{Tn5801}, and *tet*(M) probes. PCR of the *guaA*-int_{5801} fragment was performed in all transconjugants.

Fitness cost assays. Transconjugants and the corresponding recipient strain were cultured at 37°C in BHI broth overnight. Grown cultures were diluted 1:1,000 into BHI broth to obtain an inoculum of approximately 10^5 CFU/ml. An aliquot (300 μl) of this dilution was transferred into a multiwell plate and incubated at 37°C, and the growth rate was estimated from the interval of optical density at 600 nm estimated to be exponential (0.01 to 0.1) by using the GrowthRates 2.1 program (33). The measurement was performed every 15 min for 22 h in a Microbiology Workstation Bioscreen C (ThermoLabSystems, Helsinki, Finland). To guarantee culture optical homogeneity, the plates were shaken for 10 s before all measurements. Three biological and technical replicates were assayed for each strain. Relative values for each parameter were calculated by dividing the average corresponding growth rate of a given transconjugant by the average corresponding parameter of the donor in the same experiment (34).

Nucleotide sequence accession number. The A4 variant of Tn*5801* detected in this study is available in the GenBank database with accession number KP001176.

RESULTS

Genetic diversity of Tn*5801* in the genomes of four major Gram-positive genera. Twenty-three Tn*5801* variants were identified among the 5,073 genomes screened, 17 of which had not been previously published. The 23 distinct Tn*5801* backbones corre-

TABLE 1 Characterization of Tn*5801*: PCR strategy and conditions

Primer	Sequence (5′→3′)	Position in Tn*5801* (bp)	PCR	Sequence amplified	Amplicon size (bp)	Amplification conditions
gmp1	CCTGGTCTTGGTATTCGTGT	25763–25783[a]	gmp1-gmp2	*guaA*	170	1 cycle of 95°C for 10 min; 25 cycles of 94°C for 30 s, 54°C for 30 s, 72°C for 30 s; 1 cycle of 72°C for 10 min
gmp2	GTAGTCTCCCATAACACC	25609–25627[a]				
Int1	CTGTTTCCGATATTGAGC	24490–24473[a]	Int1-gmp1	*int*Tn*5801-guaA*	1,311	35 cycles of 96°C for 30 s, 45–59°C (according to primer pair) for 1 min, 72°C for 7 min; 1 cycle of 72°C for 10 min
P1	GTTTCGCAAGTAGTCTACAG	25310–25329[a]	P1-P2	*orf25-int*Tn*5801*	2,085	
P2	GTAAAGGCGACAGATGG	23244–23260[a]				
P3	CTTAGAGATGAGTTTCGTTTC	23391–23411[a]	P3-P4	*orf22-orf25*	1,455	
P4	GGATAGTTCTTTGTCTGTAAAG	21956–21977[a]				
P5	CGATTTTAGAGCCGTTGGTTTAG	22210–22232[a]	P5-P6	*tet*(M)-*orf22*	2,448	
P6	GAACGTCAGAGAGGAATTAC	19784–19803[a]				
P7	GGGAATCCCCATTTTCCTAA	19986–20005[a]	P7-P8	*orf15-tet*(M)	6,498	
P8	ACCCACTCGCTGTTTAATCG	13507–13526[a]				
P9	CTTGCAAGGCTAGGTTGGAG	14162–14181[a]	P9-P10	*orf7-orf15*	5,993	
P10	ATGTCTGAAATGGGCTTTGG	8188–8207[a]				
P11	CTTTCACTTCGTGCGGTACA	8382–8401[a]	P11-P12	*orf3-orf7*	4,631	
P12	GGGTGGGGACAATACATCAG	3770–3789[a]				
P13	CAGCCATGTAGCGTCTTTGA	4315–4334[a]	P13-P14	*orf1-orf3*	4,228	
P14	ACGGAGTTAACGGCTTTCCT	106–125[a]				
P15	CCGTATGTTCTTTCAACCACT	1030–1051[b]	P15-P11	*orf5-orf7*	2,675	
P16	GCTGACAGTTCCAGTATCC	18096–18115[c]	P16-P5	*orf23′-orf22*	1,209	
P17	CAGAACCAGCCATTACC	18561–18578[c]	P17-P18	*orf23′-orf19′*	3,938	
P18	GAAGATACGAGAAACCAATAG	14602–14623[c]				
P19	GGTGGCAATTCAAGTGTTCC	14859–14878[c]	P19-P20	*orf19′-orf17*	1,957	
P20	GGCGTATGACAAAGCTGG	12921–12939[c]				

[a] Reference sequence, *Staphylococcus aureus* Mu50 (GenBank accession number BA000017).
[b] Reference sequence, *Streptococcus agalactiae* COH1 (GenBank accession number AAJR01000021.1).
[c] Reference sequence, *Enterococcus faecalis* JH2-2 (GenBank accession number AXOI01000011.1).

sponded to 225 strains (139 *S. aureus*, 43 *Streptococcus agalactiae*, 24 *E. faecium*, 15 *E. faecalis*, 1 *Streptococcus mitis*, 1 *Enterococcus villorum*, 1 *Streptococcus thoraltensis*, and 1 *Lactococcus garvieae*). Table 2 and Table S2 in the supplemental material show the epidemiological data and the data set of the analyzed genomes.

The Tn*5801* backbones were classified in two main groups arbitrarily designated by capital letters as "group A" (14 types, A1 to A14) and "group B" (9 types, B15 to B23), which are shown in Fig. 1. These two groups differ in length (20 to 25 kb, due to the presence of an extra 5-kb fragment at the left arm of the group A transposons), nucleotide sequence, and the presence of indels and rearrangements in different regions. Variants within each group were designated by the corresponding capital letter followed by a number. The detailed analysis of the core sequences facilitated detection of differences and common features within a given type (Fig. 2; see Table S2 in the supplemental material). The GC percentage of the common Tn*5801* backbone (36%) contrasted with that of the 5-kb fragment found in type A sequences (32%) (see Fig. S2 in the supplemental material). An alignment of selected variants is shown in Fig. 3.

TABLE 2 Presence of Tn*5801* types A and B in *Firmicutes* sequences deposited in the GenBank genome database as of January 2015

Species	No. of genomes studied	No. with Tn*5801* type:	
		A	B
Streptococcus agalactiae	301	1	42
Streptococcus mitis	23	1	
Streptococcus thoraltensis	1		1
Enterococcus faecalis	344	1	13
Enterococcus faecium	254	16	7
Enterococcus villorum	2	2	
Staphylococcus aureus	4,130	139	
Staphylococcus pseudintermedius	4	1	
Lactococcus garvieae	14	1	

Below we analyze in detail the features of variants of these clusters.

(i) Tn*5801* group A. Fourteen variants (A1 to A14) were identified among the available sequences (Fig. 1 and Table 3; see Table S2 in the supplemental material). Phylogenetic analysis allowed us to group them in different subgroups of highly related variants (Fig. 2).

Tn*5801* type A1 was considered the paradigm of the group. It was detected mostly in *Staphylococcus* (96%; $n = 107$ *S. aureus*/*S. pseudintermedius* isolates out of 112 isolates with Tn*5801* type A1) and was also the predominant Tn*5801* variant within this genus (76%; $n = 107/140$ *Staphylococcus* isolates carrying Tn*5801*). Variant A1 was also present in clonally unrelated *E. faecium* and *E. faecalis* isolates and species of *Streptococcus* (*S. agalactiae* and *S. mitis*), as reported previously (19). It is of note that most Tn*5801* type A1 isolates showing particular mutations were associated with specific geographical locations and/or from specific clonal backgrounds (Fig. 2; see Tables S2 and S3 in the supplemental material) (see below). For example, most *S. aureus* isolates (83%; $n = 88/106$) carrying Tn*5801* A1 belonged to CC8 (ST8, ST507, and ST609; many of them were recovered in the United States), and only a few isolates were associated with other clonal lineages (ST5 and ST664).

Variant A4 contained a 3,285-bp insertion within the *tet*(M)

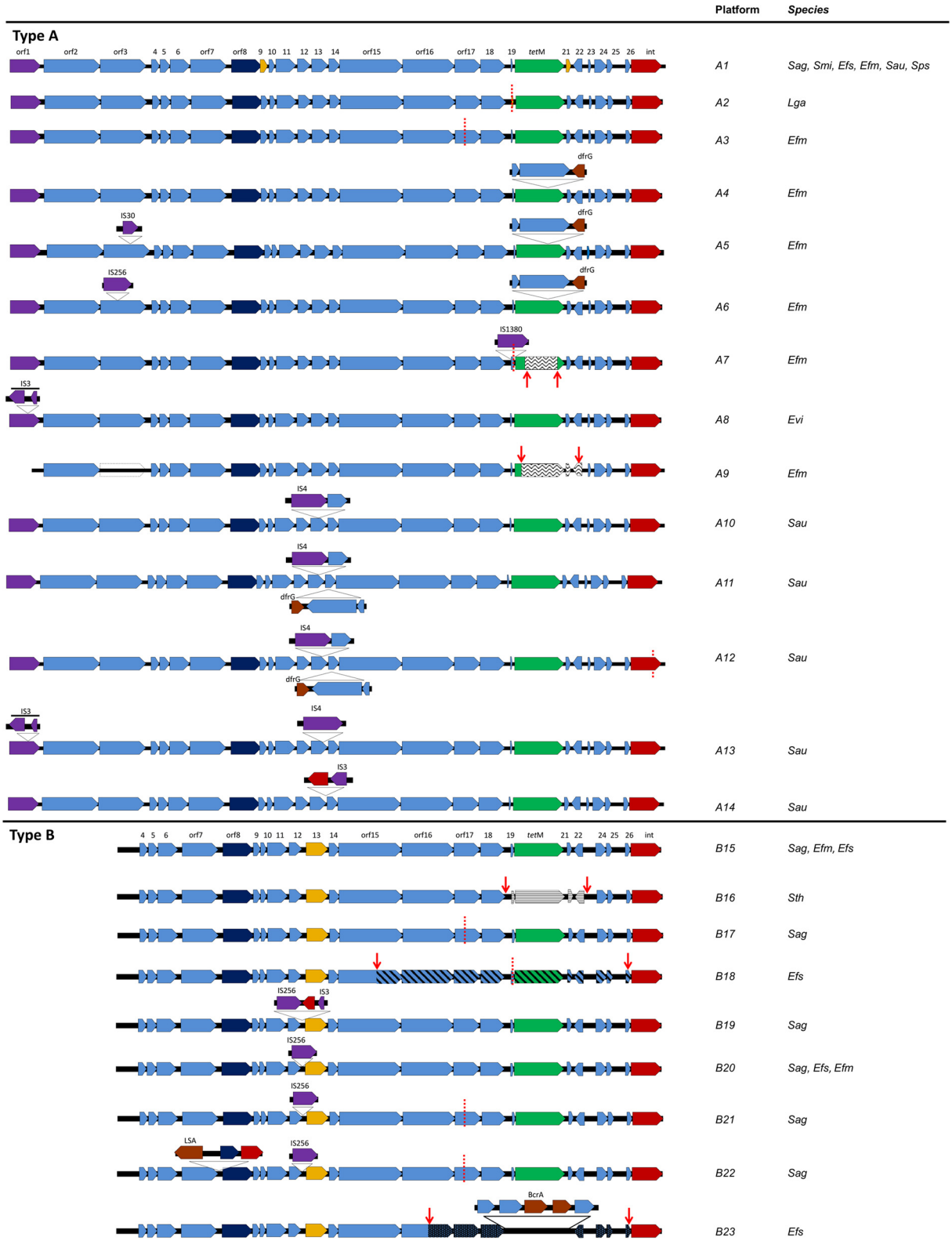


FIG 1 Diversity of Tn*5801*-like elements, types A and B, based on strains used in this study and sequences deposited in the GenBank genome database. *In silico* comparative analysis was made using BLAST and MUSCLE software available at the BLAST (http://www.ncbi.nlm.nih.gov/blast//) and MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/) websites. Dashed red lines, deletion sequences; red ORFs, ORFs encoding integrase-like proteins; dark blue ORFs, ORFs encoding replication proteins; light blue ORFs, ORFs encoding hypothetical proteins; green ORFs, tetracycline resistance gene *tet*(M); yellow ORFs, *orf*9 and *orf*21 are absent in *Staphylococcus aureus* Mu50 (reference genome) and *orf*13 in type B differs from *orf*13 in type A; brown ORFs, other resistance genes; purple ORFs, insertion sequences (IS). Wavy lines, recombination with Tn*5397*; straight lines, recombination with Tn*916*; diagonal lines, recombination with type A variant. ORFs with spots, B23 sequence differing from type B variants. *dfrG*, dihydrofolate reductase; LSA, lincosamides and streptogramins A; BcrA, bacitracin transport; *Sag*, *Streptococcus agalactiae*; *Smi*, *Streptococcus mitis*; *Efs*, *Enterococcus faecalis*; *Efm*, *Enterococcus faecium*; *Sau*, *Staphylococcus aureus*; *Sps*, *Staphylococcus pseudintermedius*. All IS correspond to IS families: IS*1062* (IS30 family, A5), IS*1542* (IS*256* family, A6), IS*1678* (IS*1380* family, A7), IS*Efa8* (IS*3* family, A8), IS*1542* (IS*256* family; B19, B20, B21, and B22), IS*Lgar5*/IS*Efm2* (IS*256* family, *orf*1).

gene that was flanked by direct and inverted repeats that comprised *dfrG* (encoding a dihydrofolate reductase) and two hypothetical protein sequences. Such an element confers resistance to trimethoprim, and it was also detected within *orf*14 of Tn*5801* types A11 and A12 (see below), in the Tn*916*-like transposon (Tn*6198*) of *Listeria monocytogenes*, in the plasmid pMG1 from *E. faecium*, and also in *Streptococcus pyogenes* isolates from India (35–37). The backbone of A4 was identical to that of Tn*5801* types A5 and A6, which also showed additional insertion sequences (IS) interrupting the *uvrD* (*orf*3) gene (IS*30* or IS*256*, respectively). All A4, A5, and A6 variants were found only among *E. faecium* ST18 isolates. Interestingly, the core sequences of A4, A5, and A6 variants clustered with type A1 sequences from human isolates of *E. faecium* and *S. aureus* from Japan and different European countries and shared specific mutations (G2336A_*orf*17, A3918T_*orf*16, A6574T_*orf*15, and T13303C_*orf*6), suggesting a common origin for them (Fig. 2; see Table S3 in the supplemental material).

Types A2, A7, A8, A13, and A14 and some A1 subvariants also clustered together (Fig. 2). Types A7 and A8 from enterococci and A2 from *L. garvieae* were highly similar according to the backbone consensus sequence (only 1 nucleotide [nt] of difference, G2644A_*orf*17). Types A2 and A7 (and also B18, an "A+B" Tn*5801* hybrid [see below]) shared a 125-bp deletion in *orf*19 (*orf*12$_{\text{Tn}916\text{-like}}$) that also appeared in the Tn*5801*-like element originally called CW459*tet*(M) from *C. perfringens* (16). In A7, such a deletion is upstream of *tet*(M) and next to an IS*1380*/IS*1678* insertion. A7 also showed a mosaic *tet*(M) gene with Tn*5397* *tet*(M) (GenBank accession number NG_034213.1). Variant A8 identified in a genome of *E. villorum* had the *orf*1 (identified as IS*Lgar5*/IS*Efm2* belonging to the IS*256* family) truncated by a fragment comprising two ORFs identified as IS*Efa8* (IS*3* family). Variants A13 (ST5) and A14 (ST247) from *S. aureus* isolates exhibited insertions of partial and complete copies of the IS*4* family and IS*Efa8* (IS3 family) sequences in *orf*14 and *orf*1 (variant A13) or an insertion with an integrase core domain within *orf*14 (variant A14). A2, A7, A8, A13, A14, and A1 shared 5 mutations, namely, G1457A and T1804C in *orf*18, G3043C_*orf*17, G70008T_*orf*15, and C13758T_*orf*4. Some isolates collected in common geographical areas also had specific additional amino acid changes (see Table S3 in the supplemental material).

Variants A10, A11, and A12 clustered together (Fig. 2) and contained insertions within *orf*13 and *orf*14 at different positions, which include an ATP binding protein and a hypothetical protein encoded within *orf*13 (A10, A11, and A12) or the cassette comprising *dfrG* mentioned above in *orf*14 (A11 and A12). A12 also had a 107-bp deletion in the integrase. They were mostly observed in *S. aureus* belonging to ST239 (84%; $n = 26/31$) and shared four nucleotide change mutations (T7199C_*orf*15, G7884A_*orf*14, T9310A_*orf*11, and C12423T_*orf*7) (see Table S3 in the supplemental material). Specific mutations were detected among isolates from specific geographical areas (variant A10 from China had an A4479G_*orf*16 mutation which was absent in A10 from the United States or Australia).

Variant A3, detected in a few isolates of *E. faecium* (ST50) collected in France during the 1990s, clustered separately and showed an 81-bp deletion in *orf*17 (Fig. 2). Variant A9, previously described as Tn*6086* (GenBank accession no. HM636636.1), was a chimera of Tn*5397* [region *tet*(M)-*orf*21-*orf*22] and Tn*5801* and also had an A/T mutation at nt 985 of the *uvrD* (*orf*3) gene that resulted in a stop codon and lack of *orf*1. It is of note that the A9 platform detected in *E. faecium* TC6 clustered with other A1 variants from *S. mitis* and *E. faecalis* isolated in Germany and Denmark, respectively, all sharing three distinct mutations (T3046A, T3586A, and C4861T). In fact, *E. faecium* TC6 is an *E. faecium* D344SRF (ST25) transconjugant from strain C68 that contains the same A9 variant (38), thus reflecting transferability of the element. Independent recombination events between Tn*5801* and Tn*5397* seem to have occurred in A9 and A7.

(ii) Tn*5801* group B. Group B comprises nine Tn*5801* elements identified among available sequences of streptococci and enterococci (B15 to B23) (Fig. 1 and Table 3; see Table S2 in the supplemental material). They split into three main clusters represented by platforms B15 to B22 (mainly in streptococci) and B18 and B23 (both in enterococci).

The B15 variant (also designated Tn*5801.Sag*) (20) was considered the paradigm of the Tn*5801* type and was detected among *E. faecalis* (this study), *E. faecium*, and *S. agalactiae* isolates. Some type B Tn*5801* elements shared a deletion of 81 bp in *orf*17 (B17, B21, and B22 from *S. agalactiae*; also detected in A3 from *E. faecium*) or an IS*256*-like insertion upstream of *orf*13 (B20, B21, and B22). The B19 sequence had this IS*256* insertion followed by an integrase core domain and an IS*3* family element. The B22 variant had an insertion of a 5,227 bp between *orf*7 and *orf*8, which comprises a phage-specific recombinase gene, a replication protein gene, and a *lsa*(C) gene encoding resistance to lincosamide, streptogramin A, and pleuromutilins (39). Such an insertion is identical to another found in *S. agalactiae* UCN70 (GenBank accession number HM990671.1) and similar to that of a Tn*916*-like element from *S. mitis*, although it lacks the *lsa*(C) gene (18, 40).

Type B23 has a region of five ORFs (encoding a transcriptional regulator, an histidine kinase, a bacitracin ABC transporter, a bacitracin transport permease, and an ABC transporter permease-like protein) instead of *orf*19-*tet*(M)-*orf*21. The sequence between *orf*26 and *orf*16 shared only a 92% similarity with the other type B sequences. This B23 variant was restricted to clonal complex 8 (CC8) (ST8, ST64, and ST90) *E. faecalis* isolates, one of the most ancient *E. faecalis* lineages (41). It included the oldest Tet^r strain described to date (an ST90 isolate from 1954, part of the historical collection of the Laboratory of Streptococcal Diseases at the National Institute of Allergy and Infectious Diseases in the United States) and ST8-related isolates of an emblematic strain widely used as a recipient (strains FA2 and JH2-2) which predate 1973 (41). The B18 variant is a mosaic of type A (region between *orf*26 and *orf*15) and type B (integrase, *orf*4 to part of *orf*15) platforms and exhibits the 125-bp deletion in *orf*19 also observed in type A variants (A2 and A7). The integrases of B18 and B23 share three unique nucleotide changes. Finally, variant B16, detected in an *S. thoraltensis* isolate (GenBank accession no. ARCI01000001.1), revealed a recombination event that resulted in the replacement of the *orf*19-*orf*22 region of Tn*5801* by that of Tn*916*.

Epidemiological background of Tn*5801* among enterococci. *int*$_{\text{Tn}5801}$ was detected in 10% of the enterococcal isolates analyzed ($n = 61/610$; 41 *E. faecium*, 13 *E. faecalis*, and 7 *Enterococcus* spp.) (Table 3). A large proportion of the *int*$_{\text{Tn}5801}$-positive isolates (84%; $n = 51/61$) were resistant to tetracycline, and most of them carried *tet*(M), *tet*(L), and/or *tet*(S) (77% [$n = 47/61$], 38% [$n = 23/61$], and 2% [$n = 1/61$], respectively). Specific transposon sequences such as *int*/*xis*$_{\text{Tn}916}$ (16%; $n = 10/61$), resolvase$_{\text{Tn}5397}$

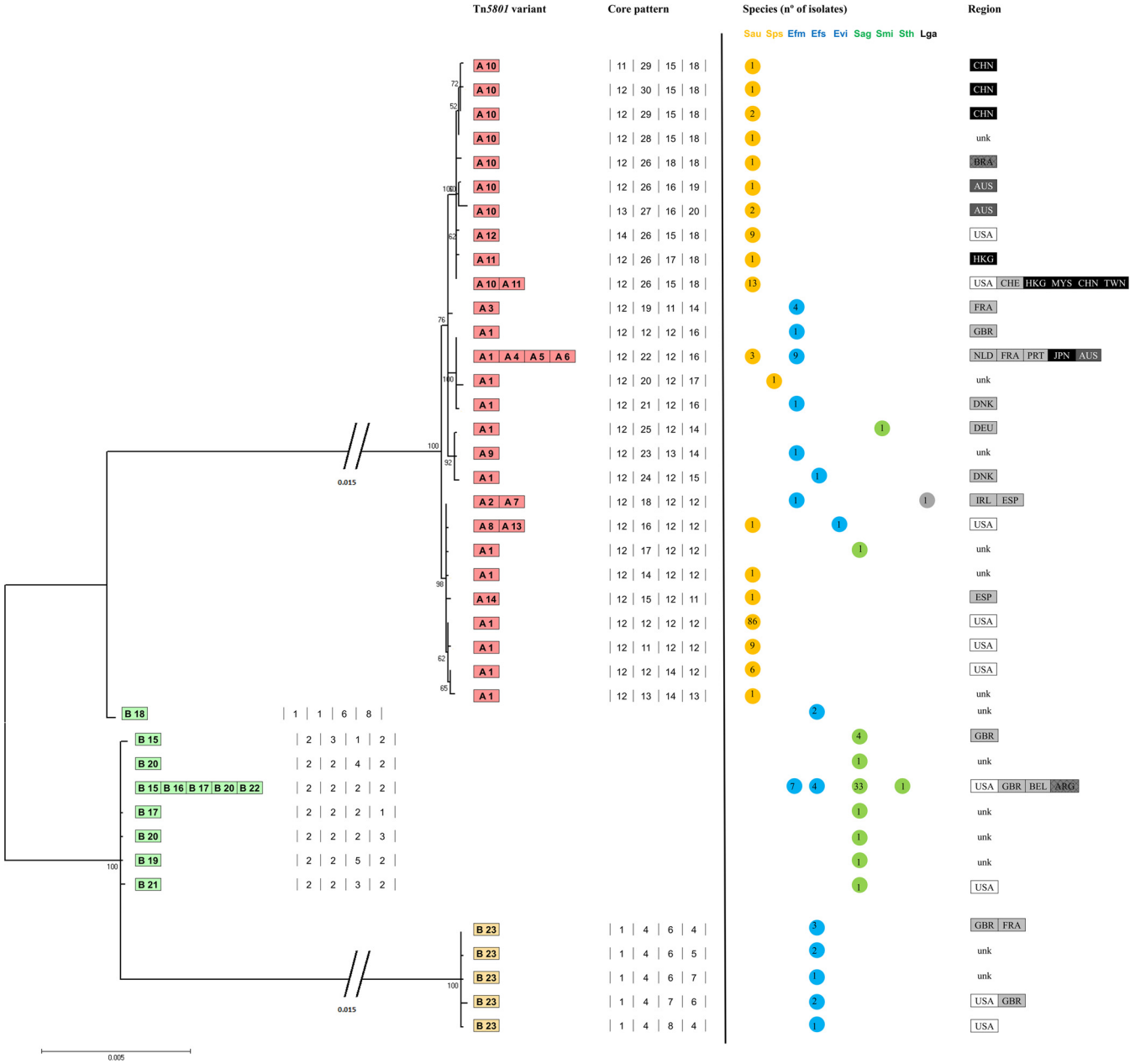


FIG 2 Phylogenetic analysis of Tn*5801* variants in *Firmicutes*. The 4 numbers separated by lines indicate the four core fragments (*int*Tn*5801*, *orf*18-14, *orf*12-8, and *orf*7-4) representing the common region in all Tn*5801* variants described in the GenBank database. These sequences were aligned, concatenated, trimmed, and further analyzed. A maximum-likelihood phylogenetic tree is represented, with 1,000 bootstrap replications. The *orf*26-19 fragment was excluded from this analysis because of its variability (resulting from both mutation and recombination events). The distribution of Tn*5801* variants by species and geographical regions is also represented. Countries within each continent are represented by different tones of gray, black and white. *Sau*, *Staphylococcus aureus*; *Sps*, *Staphylococcus pseudintermedius*; *Efm*, *Enterococcus faecium*; *Efs*, *Enterococcus faecalis*; *Evi*, *Enterococcus villorum*; *Sag*, *Streptococcus agalactiae*, *Smi*, *Streptococcus mitis*, *Sth*, *Streptococcus thoraltensis*; *Lga*, *Lactococcus garvieae*; ARG, Argentina; AUS, Australia; BEL, Belgium; BRA, Brazil; CHE, Switzerland; CHN, China; DEU, Germany; DNK, Denmark; FRA, France; GBR, Great Britain; HKG, Hong Kong; IRL, Ireland; JPN, Japan; MYS, Malaysia; NLD, The Netherlands; PRT, Portugal; ESP, Spain; TWN, Taiwan; USA, United States; unk, unknown.

(7%, $n = 4/61$), or *int*/*xis*$_{Tn6000}$ (2%; $n = 1/61$) were also observed. The Tn*5801* platforms identified in isolates of this collection correspond to A1, A4, and B15 variants. Epidemiological features of the *int*$_{Tn5801}$-positive isolates of *E. faecium* and *E. faecalis* are shown in Table 3.

Location of Tn*5801*. All Tn*5801* platforms identified in enterococci in this study were chromosomally located at the 3′ end of *guaA*, as previously observed for three emblematic Tn*5801* elements of staphylococci, *Clostridium*, and *S. agalactiae* (13, 28). In the available sequences of the Tn*5801* variants was the 11-bp sequence described previously as direct repeats (DRs) situated within the 3′ end of *guaA* (13). All *guaA* sequences within the same species had high similarity (99 to 100%), which was somewhat lower (70 to 80%) when comparing isolates from different species.

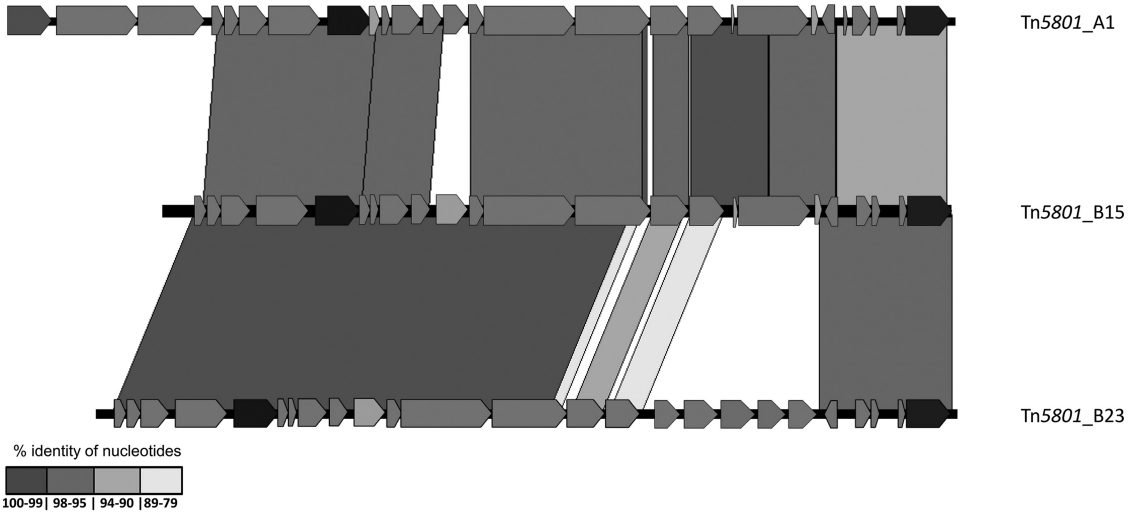


FIG 3 Alignment of selected transposons (Tn*5801*_A1, Tn*5801*_B15, and Tn*5801*_B23) by using the Artemis comparison tool (ACT). Only regions of >50 bp are represented. Gray areas mark matching regions of the same orientation among ICE from different species, indicating an identity between 79 and 100%. Only the areas with the highest percentages of similarity are shown.

Tn*5801* transfer in *Enterococcus*. The ability of Tn*5801* to be transferred was studied in mating assays using enterococcal isolates carrying the Tn*5801* variants A1, A4, and B15 as donors. Only Tn*5801* variant B15 was successfully transferred by conjugation to *E. faecalis* strains JH2-2 and OG1RF. The conjugation frequency was similar for the two receptors despite the presence of the B23 variant in the genome of the JH2-2 recipient (Table 4). Hybridization of SmaI-digested DNA from transconjugants with specific probes for *guaA*, int_{Tn5801}, and *tet*(M) allowed us to infer the location and the size of the transferred region (see Fig. S3 in the supplemental material). A single band of 150 kb that hybridized with the three probes tested was observed for all *E. faecalis* OG1RF transconjugants obtained in primary mating experiments. For *E. faecalis* JH2-2 transconjugants, the sizes of the bands that hybrid-

TABLE 3 Epidemiological features of the enterococcal isolates carrying Tn*5801*-like ICE[a]

Species	Tn*5801* platform (*n*)[b]	MLST (BAPS)	No. of isolates	Yr	Source(s) (*n*)	Country (*n*)	Tet^r phenotype(s) (*n*)	Tet^r gene(s) (*n*)	Other Tn(s) (*n*)
E. faecalis	B15 (2)	ST9/CC9	3	2001	HP	ESP, ARG	+	*tet*(M) (3), *tet*(L) (3)	
	A1	ST30/CC30	1	1997	HP	AUS	+	*tet*(M)	
	B15 (1)	ST318/CC9	2	ND	HP	BRA	+	*tet*(M)	
	B15	ST55/CC55	1	2001	HV	PRT	+	*tet*(M)	Tn*916*, Tn*6000*
	A1	ST445	1	2001	SP	PRT	+	*tet*(M), *tet*(L), *tet*(S)	
	B15 (3)	ND/ND	5	2001/2002	HP (2), HV (2), SW (1)	BRA (2), PRT (3)	+	*tet*(M)	Tn*916* (3)
E. faecium	+	ST17/CC17 (3.32)	7	1998–2005	HP (5), SW (2)	ESP (3), AUS (1), POL (1), PRT (2)	+ (6), − (1)	*tet*(M) (5), *tet*(L) (3)	
	+	ST16/CC17 (3.32)	2	1999	HP	AUS (2)	+	*tet*(M)	
	A1 (1), A4 (2)	ST18/CC17 (3.31)	9	1996–2005	HP (7), SW (2)	PRT (8), SER (1)	+	*tet*(M) (7), *tet*(L) (7)	Tn*916* (2), Tn5397 (2)
	+	ST64/CC17 (3.31)	4	2003–2004	HP	CHI	+	*tet*(M) (3)	
	+	ST173/CC17 (3.31)	1	2000	HP	AUS	+	*tet*(M), *tet*(L)	
	+	ST80/CC17 (2.1a)	1	2004	HP	HUN	+		
	+	ST132/CC17 (3.31)	1	2001	HP	PRT	+	*tet*(M), *tet*(L)	Tn*5397*
	A1	ST182/CC17 (7)	1	1992	HP	USA	+	*tet*(M)	
	+	ST202/CC17 (3.32)	1	2005	HP	POL	+	*tet*(M), *tet*(L)	
	A1	ST50/CC9 (2.1a)	1	NA	HP	BRA	+	*tet*(M)	
	+	ST366 (5)	1	2000	HP	PRT		*tet*(M)	Tn*916*
	+	ST368 (3.31)	1	2001	SW	PRT			
	+	ND/ND	11	2001–2002	HP (4), HV (1), SW (3), P (2), SP (1)	PRT (9), ESP (1), TUR (1)	+ (8), − (3)	*tet*(M) (8), *tet*(L) (5)	Tn*916* (2), Tn*5397* (1)
Enterococcus sp.	+	ND/ND	7	2001–2006	HV (4), SW (2), SP (1)	PRT	+ (3), − (4)	*tet*(M) (4), *tet*(L) (1)	Tn*916* (1)

[a] Abbreviations: ST, sequence type; BAPS, Bayesian analysis of population structure; HV, healthy volunteer; HP, hospital patient; SW, sewage; P, poultry; SP, swine/piggeries; ARG, Argentina; AUS, Australia; BRA, Brazil; CHI, China; HUN, Hungary; POL, Poland; PRT, Portugal; ESP, Spain; SRB, Serbia; TUR, Turkey; ND, not determined; Tet^r, tetracycline resistance; Tn(s), transposon(s).

[b] +, Tn*5801*-like platform which has not been characterized in the corresponding type.

TABLE 4 Results of filter mating assays with *Enterococcus faecalis* strains[a]

Transconjugant	Donor[b]	Recipient	No. of Tn*5801* copies detected	Hybridized SmaI-digested DNA band size (added fragment size), kb	Hybridized probe			Transfer rate per donor	Growth rate, mean/SD	Relative growth rate, % (transconjugant/recipient[c])	Tet MIC, μg/ml
					int	*tet*	*gua*				
JH2-2::Tn5801_1	*E. faecalis* Ef1	JH2-2	1	200 (50)	+	+	+	1.6×10^{-8}	0.15437/0.01096	3.66385	48
JH2-2::Tn5801_2	*E. faecalis* Ef1	JH2-2	2	250 (100)	+	+	+	1.6×10^{-8}	0.15489/0.01568	3.34281	96
OG1SS::Tn5801_1.1	JH2-2::Tn5801_1	OG1SS	1	150 (20)	+	+	+	1.81×10^{-7}	0.16787/0.00576	3.16674	32
OG1SS::Tn5801_1.2	JH2-2::Tn5801_1	OG1SS	2	340 (20)	+	+	−	1.01×10^{-8}	0.16773/0.00321	3.24749	96
				150 (20)	+	+	+				
OG1RF::Tn5801_1	*E. faecalis*Ef1	OG1RF	1	150 (20)	+	+	+	3.3×10^{-8}	0.16087/0.00769	0.16068	24
OG1RF::Tn5801_2	*E. faecalis* Ef1	OG1RF	1	150 (20)	+	+	+	3.3×10^{-8}	ND[d]	ND	48
OG1RF::Tn5801_3	*E. faecalis* Ef1	OG1RF	1	150 (20)	+	+	+	3.3×10^{-8}	ND	ND	64
OG1SS::Tn5801_2.1	JH2-2::Tn5801_2	OG1SS	1	150 (20)	+	+	+	2×10^{-7}	0.17217/0.00709	0.68706	96
OG1SS::Tn5801_2.2	JH2-2::Tn5801_2	OG1SS	2	150 (20)	+	+	+	2×10^{-7}	0.17717/0.00661	−2.19576	96
				120 (20)	+	+	−				
OG1SS::Tn5801_2.3	JH2-2::Tn5801_2	OG1SS	2	200 (50)	+	+	−	1.6×10^{-5}	0.17283/0.00776	0.30828	96
				170 (50)	+	+	+				
OG1SS::Tn5801_2.4	JH2-2::Tn5801_2	OG1SS	1	150 (20)	+	+	+	1.6×10^{-5}	0.16903/0.00645	2.50019	32

[a] Following the order presented in Fig. S3 in the supplemental material.
[b] *Enterococcus faecalis* Ef1 tetracycline MIC, 48 μg/ml.
[c] Recipient growth rates (means/standard deviations): 0.16024/0.01179 (*E. faecalis* JH2-2), 0.17336/0.00362 (*E. faecalis* OG1SS), and 0.16112/0.00771 (*E. faecalis* OG1RF).
[d] ND, not determined.

ized with the probes were 200 and 250 kb (the estimated size of the transferred region was ca. 20 kb for OG1RF and ca. 50 or 100 kb for the two JH2-2 transconjugants). Mating assays using these *E. faecalis* JH2-2::Tn*5801*_1 (50 kb) and *E. faecalis* JH2-2::Tn*5801*_2 (100 kb) transconjugants as donors and *E. faecalis* OG1SS as the recipient yielded *E. faecalis* OG1SS::Tn*5801* transconjugants with one or two copies of a ca. 20-kb transposable element. All transconjugants that contain two Tn*5801* copies showed positive hybridization of two bands with the int_{Tn5801} and *tet*(M) genes, but only one band hybridized with the *guaA* probe (see Fig. S3 in the supplemental material). Moreover, a PCR mapping assay demonstrated the presence of a B15 Tn*5801*-like element in all primary and secondary transconjugants, while amplification of specific B23 fragments was observed only for one primary and one secondary transconjugant, *E. faecalis* JH2-2::Tn*5801*_2 and *E. faecalis* OG1SS::Tn*5801*_2.3, respectively (Table 4; see Fig. S3 in the supplemental material).

Fitness cost of the transconjugants. The growth rates of *E. faecium* transconjugants containing B15 variants ranged from a 2.2% increase to a 3.7% reduction in the growth rate of the recipient strain (growth rates and standard deviations are indicated in Table 4). Detailed analysis of the data revealed that the secondary transconjugants (using JH2-2 as donor and OG1SS as recipient) that apparently acquired two entire Tn*5801* elements yielded different values of fitness increase or fitness reduction in comparison with the donor strain. They correspond to transconjugants that showed slightly different PFGE types, showed hybridization patterns that indicate insertion of the two acquired Tn*5801* at different genomic locations, and had a high MIC of tetracycline (96 μg/ml) (Table 4).

DISCUSSION

This study represents the first description of Tn*5801*-like elements in enterococci and one of the first studies addressing the evolutionary changes of antibiotic resistance transposons throughout decades. To date, only scarce Tn*5801*-like backbones of a few *S. aureus* and *Streptococcus* strains have been identified in separate studies (3, 15, 18–20), which did not analyze the clonal context of the isolates, thus limiting the understanding of the evolvability of this element.

Sequence analysis of type A and B variants revealed a common Tn*5801* backbone compatible with an *S. agalactiae* origin on the basis of the GC content (36%) and suggested the further acquisition of a 5-kb extra fragment by staphylococci (GC content = 32%). Such acquisition might have occurred through an IS*256* family insertion event (*orf*1 of the group A platforms) and would have resulted in the emergence of type A variants.

The apparent confinement of Tn*5801* type A or B in particular clonal backgrounds is remarkable. A recent study suggests that the introduction and massive use of tetracycline since the 1950s could have contributed to a positive selection of neonatal hypervirulent clonal lineages (CC17, CC23, or CC10) of *S. agalactiae* carrying *tet*(M)-CTn, as Tn*5801*, which we here designated Tn*5801* B type platforms (1). We were also able to detect Tn*5801* in predominant antibiotic-resistant clonal lineages associated with human infections, such as *S. aureus* (CC8, CC5, and ST239) and *E. faecium* (ST17 and ST18) from the 1990s, or early lineages of *E. faecalis* (CC8). However, our data and available knowledge necessitate a carefully interpretation of the relationship between the emergence and persistence of antibiotic resistance and selective processes of particular clones due to antimicrobial treatments (42). In this respect, we should take into account therapeutic strategies and relevant bacterial features that could have facilitated the selective acquisition of certain elements by particular clones. For example, anti-restriction-modification (anti-RM) occurring in Tn*916*-like elements often influences selective horizontal gene transfer between staphylococcal populations and other species of *Firmicutes* (43, 44). Putative anti-RM detected in Tn*5801* (*orf*12) acts against staphylococcal type I RM systems which are present in *S. aureus* CC5 and CC8 (the backgrounds in which Tn*5801* seems to have initially been detected in the late 1950s) (3, 15) and ST239 (an emergent hybrid lineage of ST8 and ST30) (10, 17, 45).

Deep sequence analysis is a powerful tool to accurately establish transmission of pathogens (46). In this study, the comprehensive analysis of Tn*5801* variants identified some types associated with sympatric populations of major clonal lineages (as reflected by the presence of specific nucleotide changes among isolates from particular geographical areas). This seems clear for *S. aureus* ST239 (which carries highly related A10, A11, and A12 variants) and ST8/ST609 human isolates from the United States recovered from 1994 to date (carrying A1), for which differences in the genome sequences of regional endemic clones have also been reported (17, 44). Besides clonal spread of strains carrying Tn*5801*, the presence of some variants in isolates of different species recovered in different countries and, eventually, different hosts indicates that horizontal gene transfer has also contributed to the current occurrence of Tn*5801*. This study demonstrates the transference of Tn*5801* between different *E. faecalis* backgrounds, even if the strain already harbored a Tn*5801* element. Although the *guaA*-associated islands theoretically integrate site specifically, tyrosine recombinases can be functional at secondary integration sites (11, 12). This would explain the insertion of elements at different genomic locations (transconjugants *E. faecalis* OG1SS::Tn*5801*_1.2, OG1SS::Tn*5801*_2.2, and OG1SS::Tn*5801*_2.3). Moreover, integration at the same site cannot be discarded if a consensus hot spot remains available for new integration events after the first Tn*5801* acquisition. For example, two *E. faecalis* genomes (*E. faecalis* Com1 and *E. faecalis* Com2; GenBank accession no. AJES01000002.1 and AJBL01000002.1, respectively) harbored a composite structure that comprises Tn*5801* B15 and a putative transposable element (GC content of 27.3%, compatible with a clostridial origin [47]), both located upstream of the *guaA* gene. The clostridial element seems to have been inserted at the 11-bp hot spot of a *guaA-int*$_{Tn5801}$ region. Finally, recombination cannot be discarded for some cases, as described for other elements in *E. faecalis* (48). Such events increase the number (if they include different genes) or the level of resistance (if the gene dosage is increased as we observed here), with a variable fitness cost to the cell, influencing the probability of being selected and stably maintained.

Although the classical JH2-2 lab strain was serendipitously useful for this work, as it allowed detection of the acquisition of more than one Tn*5801* element by a single strain, the presence of conjugative genetic elements that can interfere with the transferability of other elements under study opens discussion about the suitability of certain classic laboratory strains for use in transfer experiments.

CTn-*tet*(M) elements with the same genetic backbone may carry different determinants associated with resistance (e.g., to bacitracin, lincosamides/streptogramins, and trimethoprim in our study) and/or ISs (IS*3*, IS*4*, IS*30*, IS*256*, and IS*1380* families) which are frequently designated with different numbers in other studies (7, 20, 49). However, all these variants reflect the dynamics of a basic particular element circulating among bacterial communities. In this study, ISs with intact inverted repeats (IR) were observed within different genes (*orf*1, *orf*3, *orf*13, and *orf*19) of all Tn*5801* type A isolates of *S. aureus* and *E. faecium*, suggesting their recent acquisition. Among type B variants, only the mutator family IS*256* was consistently detected upstream of *orf*13 (types B20, B18, and B19), and this stability persisted even when some particular variants (e.g., B20) were disseminated across genera. Some ISs from the IS*256* family have been implicated as modulators of gene expression, suggesting that future studies might reveal the role of this IS in the function of type B Tn*5801*-like elements (50–53).

In summary, this study shows that Tn*5801*-like variants have been present among *Firmicutes* since at least the 1950s, being predominant in microorganisms obtained from human hosts (54). Different waves of expansion in human populations of certain clones of *Staphylococcus*, *Enterococcus*, and *Streptococcus* seem to play an important role in their occurrence and evolvability, although transfer events between different clonal backgrounds would also contribute to their spread, as demonstrated experimentally here for *E. faecalis*. The detection of genetic rearrangements in the four functional regions highlights once more the plasticity of CTns, potentially influencing microevolutionary events eliciting adaptive functions in microbial populations and communities in common or heterogeneous hosts and environments (55–58).

ACKNOWLEDGMENTS

We thank the Instituto de Salud Carlos III (PI12-01581) for the support of R.L.S. We also acknowledge the European Development Regional Fund "A way to achieve Europe" (ERDF) for cofounding the Plan Nacional de I+D+ I 2012-2015 (PI12-01581) and CIBER actions (CIBER in Epidemiology and Public Health, CIBERESP; CB06/02/0053).

FUNDING INFORMATION

European Commission, Seven Framework Program (EVOTARFP7-HEALTH-282004), provided support to Fernando Baquero and Teresa M. Coque. The Ministry of Economy and Competitiveness of Spain, Plan Nacional de I+D+ I 2012-2015, Instituto de Salud Carlos III (PI12-01581), provided support to Teresa M. Coque. The Ministry of Economy and Competitiveness of Spain, CIBERESP (CB06/02/0053), provided support to Fernando Baquero. The European Development Regional Fund "A way to achieve Europe" (ERDF) cofunds ISCIII and CIBER actions. The Regional Government of Madrid (PROMPT-S2010/BMD2414) provided support to Fernando Baquero. The Fundação para a Ciência e a Tecnologia, FCT UID/MULTI/04378/2013, provided support to Luisa Peixe and Carla Novais.

REFERENCES

1. **Da Cunha V, Davies MR, Douarre P-E, Rosinski-Chupin I, Margarit I, Spinali S, Perkins T, Lechat P, Dmytruk N, Sauvage E, Ma L, Romi B, Tichit M, Lopez-Sanchez M-J, Descorps-Declere S, Souche E, Buchrieser C, Trieu-Cuot P, Moszer I, Clermont D, Maione D, Bouchier C, McMillan DJ, Parkhill J, Telford JL, Dougan G, Walker MJ, Holden MTG, Poyart C, Glaser P.** 2014. Streptococcus agalactiae clones infecting humans were selected and fixed through the extensive use of tetracycline. Nat Commun **5:**4544. http://dx.doi.org/10.1038/ncomms5544.
2. **Edmond KM, Kortsalioudaki C, Scott S, Schrag SJ, Zaidi AKM, Cousens S, Heath PT.** 2012. Group B streptococcal disease in infants aged younger than 3 months: systematic review and meta-analysis. Lancet **379:**547–556. http://dx.doi.org/10.1016/S0140-6736(11)61651-6.
3. **de Vries LE, Christensen H, Skov RL, Aarestrup FM, Agersø Y.** 2009. Diversity of the tetracycline resistance gene tet(M) and identification of Tn916- and Tn5801-like (Tn6014) transposons in Staphylococcus aureus from humans and animals. J Antimicrob Chemother **64:**490–500. http://dx.doi.org/10.1093/jac/dkp214.
4. **Soge OO, Beck NK, White TM, No DB, Roberts MC.** 2008. A novel transposon, Tn6009, composed of a Tn916 element linked with a Staphylococcus aureus mer operon. J Antimicrob Chemother **62:**674–680. http://dx.doi.org/10.1093/jac/dkn255.
5. **Roberts AP, Mullany P.** 2011. Tn916-like genetic elements: a diverse group of modular mobile elements conferring antibiotic resistance. FEMS Microbiol Rev **35:**856–871. http://dx.doi.org/10.1111/j.1574-6976.2011.00283.x.
6. **Rice LB.** 1998. Tn916 family conjugative transposons and dissemination

of antimicrobial resistance determinants. Antimicrob Agents Chemother **42:**1871–1877.
7. **Clewell DB, Weaver KE, Dunny GM, Coque TM, Francia MV, Hayes F.** 2014. Extrachromosomal and mobile elements in enterococci: transmission, maintenance, and epidemiology. *In* Gilmore M, Clewell D, Ike Y, Shankar N (ed), Enterococci: from commensals to leading causes of drug resistant infection. Massachusetts Eye and Ear Infirmary, Boston, MA.
8. **Hegstad K, Mikalsen T, Coque TM, Werner G, Sundsfjord A.** 2010. Mobile genetic elements and their contribution to the emergence of antimicrobial resistant Enterococcus faecalis and Enterococcus faecium. Clin Microbiol Infect **16:**541–554. http://dx.doi.org/10.1111/j.1469-0691.2010.03226.x.
9. **Novais C, Freitas AR, Silveira E, Baquero F, Peixe L, Roberts AP, Coque TM.** 2012. Different genetic supports for the *tet*(S) gene in enterococci. Antimicrob Agents Chemother **56:**6014–6018. http://dx.doi.org/10.1128/AAC.00758-12.
10. **Smyth DS, McDougal LK, Gran FW, Manoharan A, Enright MC, Song J-H, de Lencastre H, Robinson DA.** 2010. Population structure of a hybrid clonal group of methicillin-resistant Staphylococcus aureus, ST239-MRSA-III. PLoS One **5:**e8582. http://dx.doi.org/10.1371/journal.pone.0008582.
11. **Smyth DS, Robinson DA.** 2009. Integrative and sequence characteristics of a novel genetic element, ICE6013, in Staphylococcus aureus. J Bacteriol **191:**5964–5975. http://dx.doi.org/10.1128/JB.00352-09.
12. **Song L, Pan Y, Chen S, Zhang X.** 2012. Structural characteristics of genomic islands associated with GMP synthases as integration hotspot among sequenced microbial genomes. Comput Biol Chem **36:**62–70. http://dx.doi.org/10.1016/j.compbiolchem.2012.01.001.
13. **Wipf JRK, Schwendener S, Nielsen JB, Westh H, Perreten V.** 2015. The new macrolide-lincosamide-streptogramin B resistance gene *erm*(45) is located within a genomic island in Staphylococcus fleurettii. Antimicrob Agents Chemother **59:**3578–3581. http://dx.doi.org/10.1128/AAC.00369-15.
14. **Boyd DA, Cabral T, Van Caeseele P, Wylie J, Mulvey MR.** 2002. Molecular characterization of the vanE gene cluster in vancomycin-resistant Enterococcus faecalis N00-410 isolated in Canada. Antimicrob Agents Chemother **46:**1977–1979. http://dx.doi.org/10.1128/AAC.46.6.1977-1979.2002.
15. **Kuroda M, Ohta T, Uchiyama I, Baba T, Yuzawa H, Kobayashi I, Cui L, Oguchi A, Aoki K, Nagai Y, Lian J, Ito T, Kanamori M, Matsumaru H, Maruyama A, Murakami H, Hosoyama A, Mizutani-Ui Y, Takahashi NK, Sawano T, Inoue R, Kaito C, Sekimizu K, Hirakawa H, Kuhara S, Goto S, Yabuzaki J, Kanehisa M, Yamashita A, Oshima K, Furuya K, Yoshino C, Shiba T, Hattori M, Ogasawara N, Hayashi H, Hiramatsu K.** 2001. Whole-genome sequencing of meticillin-resistant Staphylococcus aureus. Lancet **357:**1225–1240. http://dx.doi.org/10.1016/S0140-6736(00)04403-2.
16. **Roberts AP, Johanesen PA, Lyras D, Mullany P, Rood JI.** 2001. Comparison of Tn5397 from Clostridium difficile, Tn916 from Enterococcus faecalis and the CW459tet(M) element from Clostridium perfringens shows that they have similar conjugation regions but different insertion and excision modules. Microbiology **147:**1243–1251. http://dx.doi.org/10.1099/00221287-147-5-1243.
17. **Yamamoto T, Takano T, Higuchi W, Iwao Y, Singur O, Reva I, Otsuka Y, Nakayashiki T, Mori H, Reva G, Kuznetsov V, Potapov V.** 2012. Comparative genomics and drug resistance of a geographic variant of ST239 methicillin-resistant Staphylococcus aureus emerged in Russia. PLoS One **7:**e29187. http://dx.doi.org/10.1371/journal.pone.0029187.
18. **Brenciani A, Tiberi E, Tili E, Mingoia M, Palmieri C, Varaldo PE, Giovanetti E.** 2014. Genetic determinants and elements associated with antibiotic resistance in viridans group streptococci. J Antimicrob Chemother **69:**1197–1204. http://dx.doi.org/10.1093/jac/dkt495.
19. **Denapaite D, Brückner R, Nuhn M, Reichmann P, Henrich B, Maurer P, Schähle Y, Selbmann P, Zimmermann W, Wambutt R, Hakenbeck R.** 2010. The genome of Streptococcus mitis B6—what is a commensal? PLoS One **5:**e9426. http://dx.doi.org/10.1371/journal.pone.0009426.
20. **Mingoia M, Morici E, Tili E, Giovanetti E, Montanari MP, Varaldo PE.** 2013. Characterization of Tn5801.Sag, a variant of Staphylococcus aureus Tn916 family transposon Tn5801 that is widespread in clinical isolates of Streptococcus agalactiae. Antimicrob Agents Chemother **57:**4570–4574. http://dx.doi.org/10.1128/AAC.00521-13.
21. **Carver TJ, Rutherford KM, Berriman M, Rajandream M-A, Barrell BG, Parkhill J.** 2005. ACT: the Artemis comparison tool. Bioinformatics **21:**3422–3423. http://dx.doi.org/10.1093/bioinformatics/bti553.
22. **Tamura K, Stecher G, Peterson D, Filipski A, Kumar S.** 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol **30:**2725–2729. http://dx.doi.org/10.1093/molbev/mst197.
23. **Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lund O.** 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol **50:**1355–1361. http://dx.doi.org/10.1128/JCM.06094-11.
24. **Silveira E, Freitas AR, Antunes P, Barros M, Campos J, Coque TM, Peixe L, Novais C.** 2014. Co-transfer of resistance to high concentrations of copper and first-line antibiotics among Enterococcus from different origins (humans, animals, the environment and foods) and clonal lineages. J Antimicrob Chemother **69:**899–906. http://dx.doi.org/10.1093/jac/dkt479.
25. **CLSI.** 2014. Performance standards for antimicrobial susceptibility testing; twenty-second informational supplement. CLSI, Wayne, PA.
26. **Aarestrup FM, Agerso Y, Gerner-Smidt P, Madsen M, Jensen LB.** 2000. Comparison of antimicrobial resistance phenotypes and resistance genes in Enterococcus faecalis and Enterococcus faecium from humans in the community, broilers, and pigs in Denmark. Diagn Microbiol Infect Dis **37:**127–137. http://dx.doi.org/10.1016/S0732-8893(00)00130-9.
27. **Coque TM, Willems RJL, Fortún J, Top J, Diz S, Loza E, Cantón R, Baquero F.** 2005. Population structure of Enterococcus faecium causing bacteremia in a Spanish university hospital: setting the scene for a future increase in vancomycin resistance? Antimicrob Agents Chemother **49:**2693–2700. http://dx.doi.org/10.1128/AAC.49.7.2693-2700.2005.
28. **Tenover FC, Arbeit RD, Goering RV, Mickelsen PA, Murray BE, Persing DH, Swaminathan B.** 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J Clin Microbiol **33:**2233–2239.
29. **Willems RJL, Top J, van Schaik W, Leavis H, Bonten M, Sirén J, Hanage WP, Corander J.** 2012. Restricted gene flow among hospital subpopulations of Enterococcus faecium. mBio **3:**e00151-12. http://dx.doi.org/10.1128/mBio.00151-12.
30. **Tedim AP, Ruiz-Garbajosa P, Corander J, Rodríguez CM, Cantón R, Willems RJ, Baquero F, Coque TM.** 2015. Population biology of intestinal Enterococcus isolates from hospitalized and nonhospitalized individuals in different age groups. Appl Environ Microbiol **81:**1820–1831. http://dx.doi.org/10.1128/AEM.03661-14.
31. **Freitas AR, Tedim AP, Novais C, Ruiz-Garbajosa P, Werner G, Laverde-Gomez JA, Cantón R, Peixe L, Baquero F, Coque TM.** 2010. Global spread of the hyl_{Efm} colonization-virulence gene in megaplasmids of the Enterococcus faecium CC17 polyclonal subcluster. Antimicrob Agents Chemother **54:**2660–2665. http://dx.doi.org/10.1128/AAC.00134-10.
32. **Werner G, Freitas AR, Coque TM, Sollid JE, Lester C, Hammerum AM, Garcia-Migura L, Jensen LB, Francia MV, Witte W, Willems RJ, Sundsfjord A.** 2011. Host range of enterococcal vanA plasmids among Gram-positive intestinal bacteria. J Antimicrob Chemother **66:**273–282. http://dx.doi.org/10.1093/jac/dkq455.
33. **Hall BG, Acar H, Nandipati A, Barlow M.** 2014. Growth rates made easy. Mol Biol Evol **31:**232–238. http://dx.doi.org/10.1093/molbev/mst187.
34. **Foucault M-L, Depardieu F, Courvalin P, Grillot-Courvalin C.** 2010. Inducible expression eliminates the fitness cost of vancomycin resistance in enterococci. Proc Natl Acad Sci U S A **107:**16964–16969. http://dx.doi.org/10.1073/pnas.1006855107.
35. **Bergmann R, Sagar V, Nitsche-Schmitz DP, Chhatwal GS.** 2012. First detection of trimethoprim resistance determinant dfrG in Streptococcus pyogenes clinical isolates in India. Antimicrob Agents Chemother **56:**5424–5425. http://dx.doi.org/10.1128/AAC.01284-12.
36. **Sekiguchi J, Tharavichitkul P, Miyoshi-Akiyama T, Chupia V, Fujino T, Araake M, Irie A, Morita K, Kuratsuji T, Kirikae T.** 2005. Cloning and characterization of a novel trimethoprim-resistant dihydrofolate reductase from a nosocomial isolate of Staphylococcus aureus CM.S2 (IMCJ1454). Antimicrob Agents Chemother **49:**3948–3951. http://dx.doi.org/10.1128/AAC.49.9.3948-3951.2005.
37. **Bergmann R, van der Linden M, Chhatwal GS, Nitsche-Schmitz DP.** 2014. Factors that cause trimethoprim resistance in Streptococcus pyogenes. Antimicrob Agents Chemother **58:**2281–2288. http://dx.doi.org/10.1128/AAC.02282-13.
38. **Rice LB, Lakticová V, Carias LL, Rudin S, Hutton R, Marshall SH.** 2009. Transferable capacity for gastrointestinal colonization in Enterococcus

faecium in a mouse model. J Infect Dis **199:**342–349. http://dx.doi.org/10.1086/595986.
39. **Douarre P-E, Sauvage E, Poyart C, Glaser P.** 2015. Host specificity in the diversity and transfer of lsa resistance genes in group B Streptococcus. J Antimicrob Chemother **70:**3205–3213. http://dx.doi.org/10.1093/jac/dkv277.
40. **Malbruny B, Werno AM, Murdoch DR, Leclercq R, Cattoir V.** 2011. Cross-resistance to lincosamides, streptogramins A, and pleuromutilins due to the *lsa*(C) gene in Streptococcus agalactiae UCN70. Antimicrob Agents Chemother **55:**1470–1474. http://dx.doi.org/10.1128/AAC.01068-10.
41. **McBride SM, Fischetti VA, Leblanc DJ, Moellering RC, Gilmore MS.** 2007. Genetic diversity among Enterococcus faecalis. PLoS One **2:**e582. http://dx.doi.org/10.1371/journal.pone.0000582.
42. **Day T, Huijben S, Read AF.** 2015. Is selection relevant in the evolutionary emergence of drug resistance? Trends Microbiol **23:**126–133. http://dx.doi.org/10.1016/j.tim.2015.01.005.
43. **Roberts GA, Chen K, Bower EKM, Madrzak J, Woods A, Barker AM, Cooper LP, White JH, Blakely GW, Manfield I, Dryden DTF.** 2013. Mutations of the domain forming the dimeric interface of the ArdA protein affect dimerization and antimodification activity but not antirestriction activity. FEBS J **280:**4903–4914. http://dx.doi.org/10.1111/febs.12467.
44. **Roberts GA, Houston PJ, White JH, Chen K, Stephanou AS, Cooper LP, Dryden DTF, Lindsay JA.** 2013. Impact of target site distribution for type I restriction enzymes on the evolution of methicillin-resistant Staphylococcus aureus (MRSA) populations. Nucleic Acids Res **41:**7472–7484. http://dx.doi.org/10.1093/nar/gkt535.
45. **Sadykov MR.** 2014. Restriction-modification systems as a barrier for genetic manipulation of Staphylococcus aureus, p 9–23. *In* Bose JL (ed), The genetic manipulation of staphylococci. Springer, New York, NY.
46. **Worby CJ, Lipsitch M, Hanage WP.** 2015. Shared genomic variants: identification of transmission routes using pathogen deep sequence data. bioRxiv http://dx.doi.org/10.1101/032458.
47. **Lassalle F, Périan S, Bataillon T, Nesme X, Duret L, Daubin V.** 2015. GC-content evolution in bacterial genomes: the biased gene conversion hypothesis expands. PLoS Genet **11:**e1004941. http://dx.doi.org/10.1371/journal.pgen.1004941.
48. **Rice LB, Carias LL, Marshall SH, Hutton-Thomas R, Rudin S.** 2007. Characterization of Tn5386, a Tn916-related mobile element. Plasmid **58:**61–67. http://dx.doi.org/10.1016/j.plasmid.2007.01.002.
49. **Palmieri C, Mingoia M, Massidda O, Giovanetti E, Varaldo PE.** 2012. Streptococcus pneumoniae transposon Tn1545/Tn6003 changes to Tn6002 due to spontaneous excision in circular form of the *erm*(B)- and aphA3-containing macrolide-aminoglycoside-streptothricin (MAS) element. Antimicrob Agents Chemother **56:**5994–5997. http://dx.doi.org/10.1128/AAC.01487-12.
50. **Maki H, Murakami K.** 1997. Formation of potent hybrid promoters of the mutant llm gene by IS256 transposition in methicillin-resistant Staphylococcus aureus. J Bacteriol **179:**6944–6948.
51. **Benson MA, Ohneck EA, Ryan C, Alonzo F, Smith H, Narechania A, Kolokotronis S-O, Satola SW, Uhlemann A-C, Sebra R, Deikus G, Shopsin B, Planet PJ, Torres VJ.** 2014. Evolution of hypervirulence by a MRSA clone through acquisition of a transposable element. Mol Microbiol **93:**664–681. http://dx.doi.org/10.1111/mmi.12682.
52. **McEvoy CRE, Tsuji B, Gao W, Seemann T, Porter JL, Doig K, Ngo D, Howden BP, Stinear TP.** 2013. Decreased vancomycin susceptibility in Staphylococcus aureus caused by IS256 tempering of WalKR expression. Antimicrob Agents Chemother **57:**3240–3249. http://dx.doi.org/10.1128/AAC.00279-13.
53. **Fléchard M, Gilot P.** 2014. Physiological impact of transposable elements encoding DDE transposases in the environmental adaptation of Streptococcus agalactiae. Microbiology **160:**1298–1315. http://dx.doi.org/10.1099/mic.0.077628-0.
54. **Agersø Y, Pedersen AG, Aarestrup FM.** 2006. Identification of Tn5397-like and Tn916-like transposons and diversity of the tetracycline resistance gene tet(M) in enterococci from humans, pigs and poultry. J Antimicrob Chemother **57:**832–839. http://dx.doi.org/10.1093/jac/dkl069.
55. **Bellanger X, Payot S, Leblond-Bourget N, Guédon G.** 2014. Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. FEMS Microbiol Rev **38:**720–760. http://dx.doi.org/10.1111/1574-6976.12058.
56. **Bellanger X, Roberts AP, Morel C, Choulet F, Pavlovic G, Mullany P, Decaris B, Guédon G.** 2009. Conjugative transfer of the integrative conjugative elements ICESt1 and ICESt3 from Streptococcus thermophilus. J Bacteriol **191:**2764–2775. http://dx.doi.org/10.1128/JB.01412-08.
57. **Burrus V, Pavlovic G, Decaris B, Guédon G.** 2002. Conjugative transposons: the tip of the iceberg. Mol Microbiol **46:**601–610. http://dx.doi.org/10.1046/j.1365-2958.2002.03191.x.
58. **Roberts AP, Mullany P.** 2009. A modular master on the move: the Tn916 family of mobile genetic elements. Trends Microbiol **17:**251–258. http://dx.doi.org/10.1016/j.tim.2009.03.002.