environmental microbiology

Environmental Microbiology (2019) **21**(8), 3046–3062 doi:10.1111/1462-2920.14702

Phylogenomics of *Enterococcus faecalis* from wild birds: new insights into host-associated differences in core and accessory genomes of the species

Ricardo León-Sampedro,[1,2] **Rosa del Campo**,[1,3] **Mercedes Rodriguez-Baños**,[1] **Val F. Lanza**,[1,2] **María José Pozuelo**,[4] **Carlos Francés-Cuesta**,[5,6] **Ana P. Tedim**,[1] **Ana R. Freitas**,[7] **Carla Novais**,[7] **Luísa Peixe**,[7] **Rob J. L. Willems**,[8] **Jukka Corander**,[9] **Fernando González Candelas**,[2,5,6] **Fernando Baquero**[1,2,10] **and Teresa M. Coque**[1,2,10]*

[1]*Department of Microbiology, University Hospital Ramón y Cajal, Ramón y Cajal Health Research Institute (IRYCIS), Madrid, Spain.*

[2]*Biomedical Research Networking Center for Epidemiology and Public Health (CIBER-ESP), Madrid, Spain.*

[3]*Spanish Network for Research in Infectious Diseases (REIPI), Health Institute Carlos III, Madrid, Spain.*

[4]*Department of Biology, Pharmacy Faculty, University San Pablo-CEU, Boadilla del Monte, Spain.*

[5]*Infection and Public Health Unit, FISABIO/University of Valencia, Spain.*

[6]*Institute for Integrative Systems Biology, I*^{2}*SysBio, CSIC-University of Valencia, Valencia, Spain.*

[7]*UCIBIO/REQUIMTE, Department of Biological Sciences, Microbiology Laboratory, Pharmacy Faculty, University of Porto, Porto, Portugal.*

[8]*Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, the Netherlands.*

[9]*Department of Biostatistics, University of Oslo, Oslo, Norway.*

[10]*Antibiotic Resistance and Bacterial Virulence Unit Associated with the Superior Council of Scientific Investigations (CSIC), Madrid, Spain.*

Summary

Wild birds have been suggested to be reservoirs of antimicrobial resistant and/or pathogenic *Enterococcus faecalis* (*Efs*) strains, but the scarcity of studies and available sequences limit our understanding of the population structure of the species in these hosts. Here, we analysed the clonal and plasmid diversity of 97 *Efs* isolates from wild migratory birds. We found a high diversity, with most sequence types (STs) being firstly described here, while others were found in other hosts including some predominant in poultry. We found that pheromone-responsive plasmids predominate in wild bird *Efs* while 35% of the isolates entirely lack plasmids. Then, to better understand the ecology of the species, the whole genome of fivestrains with known STs (ST82, ST170, ST16 and ST55) were sequenced and compared with all the Efs genomes available in public databases. Using several methods to analyse core and accessory genomes (AccNET, PLACNET, hierBAPS and PANINI), we detected differences in the accessory genome of some lineages (e.g. ST82) demonstrating specific associations with birds. Conversely, the genomes of other *Efs* lineages exhibited divergence in core and accessory genomes, reflecting different adaptive trajectories in various hosts. This pangenome divergence, horizontal gene transfer events and occasional epidemic peaks could explain the population structure of the species.

Introduction

Enterococcus faecalis is a member of the gastrointestinal flora of mammals, reptiles, insects and birds, and it causes mild-to-severe infections in these hosts (Lebreton *et al*., 2017). Transmission of both genetic elements encoding antibiotic resistance and strains between animals and humans relies on the genetic similarity/identity between donor and recipient (Ruiz-Garbajosa *et al*., 2006a,b; McBride *et al*., 2007; Freitas *et al*., 2009, 2017; Larsen *et al*., 2011; Kuch *et al*., 2012; Oravcova *et al*., 2013, 2014; Lozano *et al*., 2015; Zischka *et al*., 2015). Although sporadically analysed, wild birds might also be vectors of genes encoding resistance to first-line antibiotics (Oravcova *et al*., 2017; Ben Yahia *et al*., 2018) and possibly, highly virulent strains for poultry (Petersen *et al*., 2008). Unfortunately, there are very few strains from non-human sources either included in whole-

Received 30 August, 2018; revised 20 May, 2019; accepted 1 June, 2019. *For correspondence. E-mail mariateresacoque.hrc@salud.madrid.org or E-mail teresacoque@gmail.com; Tel. (+34) 913368832.

genome databases or included in analysis of population structure by classical methods [7.7% of the available genomes at the time of our analysis (39/441); 23 from farm animals, 2 from insects, 10 from non-flying birds (two penguins, one chicken, one duck and one turkey), one tunicate, one mouse, two pets (cat, dog)]. This, along with the low specificity of typing methods, precludes our understanding of the ecology and transmission dynamics of relevant adaptive traits in *E. faecalis.* In addition, the diversity and abundance of mobile genetic elements of *E. faecalis* remain largely unexplored (Lanza *et al*., 2015).

Enterococcus faecalis population structure analyses based on multilocus sequence typing (MLST) data (both the sequence-based gene trees and the allele-based population snapshots obtained by the goeBURST) or on the phylogenomic analysis of concatenated core genes have revealed a high level of recombination and a low proportion of strains with identical genotypes (Ruiz-Garbajosa *et al*., 2006a,b; Guzman Prieto *et al*., 2016). Attempts to establish ecotypes or natural host-adapted clusters using Bayesian analyses of population structure (BAPS) yielded incongruent groups in studies on various sample sizes (Tedim *et al*., 2015). Recently, neutral models that incorporate microepidemics and migration, which are considered the ecological factors possibly limiting transmission between subpopulations, revealed that the *E. faecalis* population structure maintains hallmarks of a high mutation rate and/or drift (Numminen *et al*., 2016). These models suggest frequent inter-host migration, both vertical and horizontal, which would continuously homogenize the species stratification. This homogenization would mix the various populations among hosts and sources thus explaining the apparently limited host specialization of this bacterial species (Numminen *et al*., 2016).

Here, we have analysed *E. faecalis* strains from wild migratory birds (97 isolates from 35 bird species) by MLST and whole genome sequencing (WGS) and compared the results with available MLST and WGS databases. We used a combination of phylogenetic and recently published tools to analyse population structure, which allowed us to infer the core and accessory genomes. The results further suggest the apparently limited host specialization of this enterococcal species (Ruiz-Garbajosa *et al*., 2006a,b; Tedim *et al*., 2015; Numminen *et al*., 2016), with the accessory genome reflecting the boundaries of some populations, probably in accordance with their historical evolutionary path. Furthermore, an analysis of plasmid diversity of this species is provided.

Results

We analysed 97 *E. faecalis* recovered from 103 faecal samples of wild birds at two Spanish veterinarian centres for nature conservation in the Central and South West of the Iberian Peninsula (2001–2010). The sampled animals were migratory and raptor birds, which represents the diversity of Iberian Peninsula wildlife (33 species of 10 orders with various lifestyle and dietary habits, see. Table 1).

The combined analysis of the eco-epidemiological investigation (including data on antibiotic resistance and gene order), MLST-based population structure analyses, and the study of plasmid replicon diversity of the 97 *E. faecalis* strains, suggested a high degree of gene flow among populations of this bacterial species. Further phylogenomic analyses using the five genomes from wild birds in comparison with the 441 *E. faecalis* genomes available in genome databases also suggested broad genetic promiscuity within the species.

Antibiotic resistance and virulence factors of E. faecalis *from wild birds*

Sixty-six of the 97 isolates (68%) were resistant to one or more antibiotics, including tetracycline (67%), chloramphenicol (42%) and erythromycin (28%). In addition, we discovered high level of streptomycin (26%), kanamycin (19%) and gentamicin (5%) (Supporting Information Table S1). All the isolates were susceptible to ampicillin and glycopeptides, whereas one-third (32%) were susceptible to all antibiotics tested. Resistance profiles are shown in the Supporting Information Table S1. We detected the presence of various putative virulence genes coding for *gelE* (93%), *asa1* (76%), *esp* (60%) and *cylA* (35%). The presence of virulence genes was not associated with any wild bird group. The presence/absence of these genes in the *E. faecalis* WGS database is represented in the Supporting Information Tables S2 and S3.

Population structure of E. faecalis *from wild birds in the context of the population structure of the species*

MLST-based comparative analysis. MLST of the 56 isolates of wild birds representing unique pulsed-field gel electrophoresis (PFGE) types revealed 40 STs, most (26/40 = 65%) of which 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). Using goeBURST, we also identified known STs that grouped into nine clonal complexes (CCs), 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) and CC19 (ST339). To identify and localize these STs in the global population structure of *E. faecalis*, we analysed the concatenated sequences of the seven MLST genes of all 1452 *E. faecalis* isolates deposited in the MLST database by

Table 1. Sources of the 103 faecal samples used in this study.

Order	Number of birds	Common name	Scientific name	Number of isolates
Falconiformes	42	Griffon Vulture	*Gyps fulvus*	6
		Cinereus Vulture	*Aegypius monachus*	3
		Black Vulture, Monk Vulture		
		Red Kite	*Milvus milvus*	3
		Black Kite	*Milvus migrans*	3
		Egyptian Vulture/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	10	Drakes	*Anas platyrhynchos*	10
Ciconiiformes	19	Cattle Egret	*Bubulcus ibis*	2
		White Stork	*Ciconia ciconia*	17
Charadriiformes	6	Stone Curlew	*Burhinus oedicnemus*	6
		Lesser Black-backed Gull	*Larus fuscus*	
Caprimulgiforme	1	European Nightjar	*Caprimulgus europaeus*	1
Columbiformes	2	Rock Dove	*Columba livia*	2
Passeriformes	10	Red-billed Chough	*Pyrrhocorax pyrrhocorax*	2
		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	3	Great Bustard	*Otis tarda*	3
Strigiformes	6	Brown Owl	*Strix aluco*	1
		Owl	*Different species*	1
		Little Owl	*Athene noctua*	4
Coraciformes	4	Hoopoe	*Upupa epops*	1
		Unknown		3

hierarchical Bayesian Analysis of the Population Structure (hierBAPS). This analysis identified three first-level hierBAPS populations, here called 'mlstEFS' (multilocus sequence type *E. faecalis*) followed by a capital letter (mlstEFS-'*N*') and nine hierBAPS subgroups called 'mlstEFS' followed by a number (mlstEFS-'*n*') after a second-level analysis (Supporting Information Fig. S1). Most STs from wild birds fall within the hierBAPS subgroups mlstEFS-4, -5 and -6, and strains of this mlstEFS-6 subgroup appear to be distributed in various distant branches in the tree of concatenated MLST alleles (Supporting Information Fig. S1).

The decision to apply hierBAPS to facilitate clustering of the sequences into genetically distinct groups and study the admixture (https://github.com/gtonkinhill/rhierbaps) (Supporting Information Fig. S1) was based on a new analysis considering the increase in the number of STs included in the *E. faecalis* database (https://pubmlst.org/efaecalis/) in the last 2 years (i.e. 218 new STs with a total of 1452 isolates corresponding to 741 STs). The results had an excellent correlation (r = 0.956) with those obtained in a previous study by our group, excluding ST80, which was identified as the distorting outgroup both for BAPS and phylogenetic analysis (Willems *et al*., 2012; Tedim *et al*., 2015).

Whole-genome comparative analysis: core genome. In an attempt to analyse the genetic variability of *E. faecalis* STs in different hosts, we obtained the WGS of five strains from wild birds of disparate lifestyles, namely *Ciconia ciconia* (omnivore; AE8-ST170 and AE63-ST55), *Gyps fulvus* (scavenger; AE65-ST82), *Strix aluco* (carnivorous and nocturnal; AE60-ST82) and *Falco tinnunculus* (carnivorous; AE51-ST16). The five sequenced genomes from birds isolates corresponded to STs widely described in humans (ST55), both in humans and farm animals (ST16), or from poultry and other birds (ST82). A comparative analysis of these sequences with the 441 sequences of *E. faecalis* available at the GenBank database, was performed (Supporting Information Table S2).

The genome analysis of the 441 *E. faecalis* from GenBank revealed an average genome size of 3.3 Mb, and an average core genome of 1 526 083 bp that includes 102 161 SNPs (the alignment of the SNPs core genome is available under DOI 10.17605/OSF.IO/U6APN in Open Science Framework repository). Further analysis of the SNP

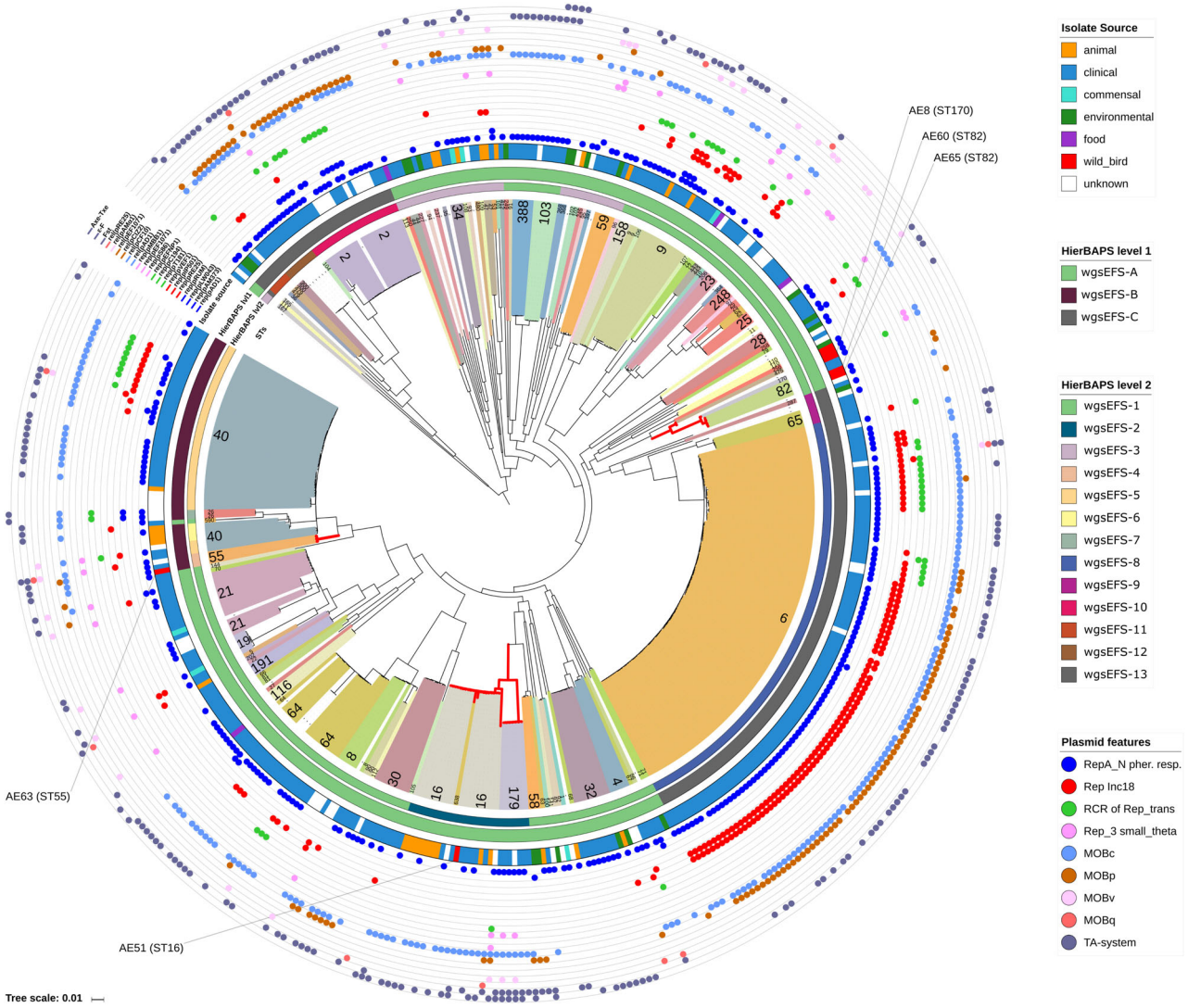

Fig. 1. Phylogenetic tree of *E. faecalis* core genomes deposited in the databases (from human clinical samples, human commensal, farm and domestic animals, environmental samples, and food, n = 441) plus the five *E. faecalis* isolates from wild birds examined in this study. The inner circle indicates the STs, with different shades of colours indicating different STs. The second and third circles indicate the clusters obtained groups and subgroups revealed with hierBAPS analysis. The fourth circle represents the origin of the isolates (animal, human, environmental). Red branches correspond to clades detected in wild bird isolates of the present study. The 22 outer circles marked by coloured dots indicate the presence or absence of plasmids identified by the *in silico* PCR screening. Each colour indicates a different plasmid family. [Color figure can be viewed at wileyonlinelibrary.com]

core genome with hierBAPS showed three WGS-based EFS groups (wgsEFS) designated with capital letters (Fig. 1). We detected three major groups (wgsEFS A, B, C) comprising various subgroups. Within the largest and highly heterogeneous hierBAPS wgsEFS-A group (three subgroups), isolates from wild birds fell into subgroups wgsEFS-1 (ST82-like isolates AE8-ST170, AE60-ST82, AE65-ST82, in addition to one from a human and two from penguins in the same branch) and wgsEFS-2 (which comprised one isolate of this study and isolates of ST16 and SLV ST179 from humans and animals).

Within hierBAPS wgsEFS-B, the subgroups wgsEFS-4, wgsEFS-5 and wgsEFS-7 were linked to the phylogenetic clusters ST55 (including the isolate AE63-ST55), ST40 (the sub-branch linked to humans), and ST26/ST590 respectively. Finally, we did not detect any isolate from wild birds in the wgsEFS-C group, which only comprises human isolates. The hierBAPS wgsEFS-C split into two large and phylogenetically distant *E. faecalis* branches, comprising two (wgsEFS-8 and wgsEFS-9) and four subgroups (wgsEFS-10, -11, -12, -13).

Whole-genome comparative analysis: accessory genome. The analysis of the genomes available at the GenBank database or SRA using the AccNET tool (Lanza *et al*., 2017), yielded a set of 9481 proteins representative

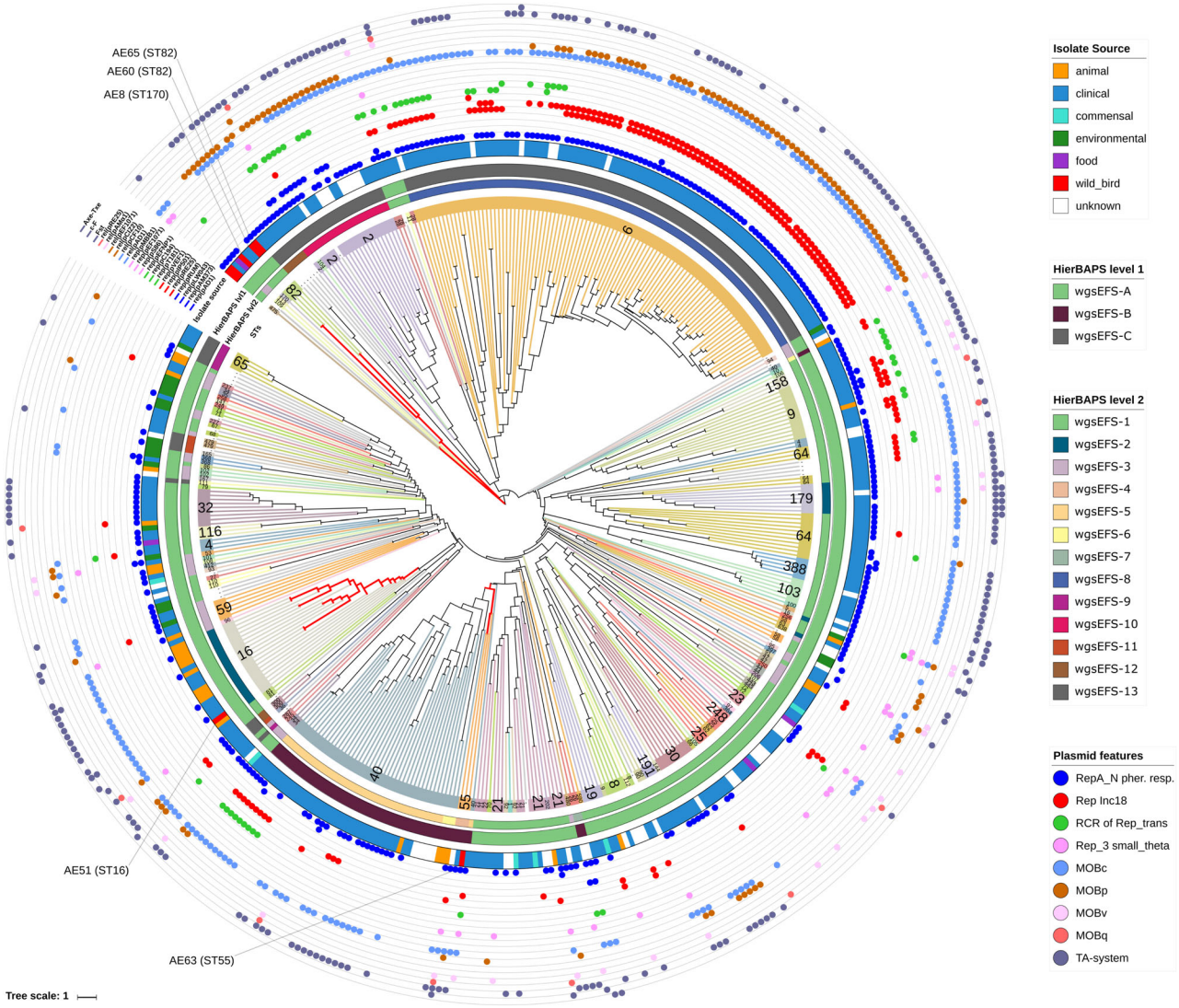

Fig. 2. Accessory genome tree constructed by a distance matrix built from the accessory network of *E. faecalis* genomes deposited in the databases (from human clinical samples, human commensal, farm and domestic animals, environmental samples, and food, n = 441) plus five *E. faecalis* isolates from wild birds examined in this study. A set of 11 630 representative proteins of the accessory genome from the available fully sequenced strains were classified using OrthoMCL (Fischer *et al*., 2011), genomes and proteins (nodes) being clustered according to the accessory genome profiles (95% similarity between genome unit nodes and protein nodes). The inner circle indicates the STs: different shades of colours represent different STs. The second and third circles from inside indicate the clusters obtained with the first and second level of hierBAPS analysis, which are marked in different colours. The outer circle represents the origin of the isolates. Red branches correspond to the clade corresponding to the common ancestor of wild bird isolates of our collection in a determined phylogenetic distance. The 22 outer circles marked by coloured dots indicate the presence or absence of plasmids by the *in silico* PCR screening. Each colour indicates a plasmid family. [Color figure can be viewed at wileyonlinelibrary.com]

of the accessory genome (Supporting Information Table S3), which were classified using OrthoMCL (Fischer *et al*., 2011). In the accessory network, genomes and proteins (nodes) were clustered according to the accessory genome profiles (95% similarity between genome unit nodes and protein nodes) (Figs. 2 and 3).

Figure 2 shows the accessory genome tree of *E. faecalis*, which had a strong positive correlation with the core genome tree represented in Fig. 1 (p = 0.001 by the Mantel test) despite some incongruences between some predominant lineages within wgsEFS-A (ST16 and its SLV ST179, both within the wgsEFS-2 subgroup) and wgsEFS-C (ST6 and ST2 of the subgroups wgsEFS-8 and wgsEFS-9 respectively).

A detailed analysis of the STs from wild birds sequenced here revealed some features of these lineages. ST16 was widely distributed in humans and farm animals, especially poultry. Its single variant ST179, which was particularly prevalent among humans (Bortolaia *et al*., 2016), had a similar core but showed disparate accessory genomes.

Accessory genomes of the ST179 isolates were similar to those of ST64, one of the oldest and more abundant *E. faecalis* lineages in the human gastrointestinal tract (Moles *et al*., 2015). The 'root' position of the ST82 strains in the tree indicates the 'uniqueness' of the ST82 accessory genome in the context of the genomes analysed (Fig. 2), and its accessory genome network revealed common and unique features among strains (Fig. 3). The ST82 isolates exhibited a set of 89 proteins specific to the group (Supporting Information Table S3), with only 38 of 89 with a known function and belonging to different function groups according to the clusters of orthologous groups (COG) database (carbohydrate transport and metabolism, 16%; transcription, 13%; replication, 13%; recombination and repair, 13%; defence mechanisms, 13%; mobilome 10%; energy production, 5%; translation and ribosomal structure, 5%; and membrane proteins, 3%). A small cluster of 10 hypothetical proteins was also shared with ST256 and ST475 from penguin isolates.

Although we did not find any isolate of the wgsEFS-C group, it is worth noting the high degree of admixture of ST6 with other *E. faecalis* lineages, especially ST2. ST6 was the most common ST identified in clinical isolates of the species (Figs. 2 and 3).

The PANINI application (Supporting Information Fig. S2) confirmed the main clusters and incongruences for some of the STs observed in the accessory genome cladogram obtained with AccNET. These results can be viewed online at https://leonsampedro.github.io/AccNET_WB/Figure_AccWB.html (AccNET) and https://microreact.org/project/Skt1ZFaYf (PANINI).

Plasmid diversity in wild bird isolates. In our sample of 97 *E. faecalis* from wild birds, the number of plasmids per strain varied between 0 and 4 (sizes ranging from 3 to 290 kb), with a notable absence of plasmids in 35% of the 97 isolates studied. The screening of replicases (*rep*), relaxases (*rel*) and toxin–antitoxin systems (TA) further provided a snapshot of the plasmid diversity in our sample and in the whole genome database (Supporting Information Fig. S3).

- *Pheromone-responsive plasmids.* These plasmids predominate in wild birds as reflected by the occurrence of *rep*9 genes (78%, which can split into *rep*9A$_{pAD1}$, *rep*9B$_{pBEE99}$, *rep*9C$_{pCF10}$; 23%, 34%; 43% respectively), and relaxases of the MOBP (PcfG$_{pCF10/pBEE99}$, 63%; rel9$_{pCF10}$, 12%; rel2$_{pEF1071}$, 2%) and MOBC (rel5$_{pAD1}$, 51%) families (Supporting Information Fig. S3; Tables 2 and 3). When we analysed the five genomes sequenced, we identified a single pheromone-responsive plasmid in four strains (no plasmids were detected in the remaining sample).

The two ST82 *E. faecalis* isolates from the disparate species *Gyps fulvus* and *Strix aluco* harboured an identical plasmid (pAE60 and pAE65, Fig. 4) (100% coverage and 99% similarity, rep$_{pAD1}$, and a rel$_{pCF10}$) which entirely matched the partial sequences of the WGS of the *E. faecalis* strain RMC5, a clinical isolate of 1954 (McBride *et al*., 2007) (sequence IDs: ASDD01000025.1, ASDD01000029.1, ASDD01000030.1 and ASDD01000031.1) and another partial WGS of the *E. faecalis* isolate C6 (Sequence ID: OANZ01000011.1). Some regions of this plasmid are highly homologous to regions of plasmids pBEE99 and pEF123.

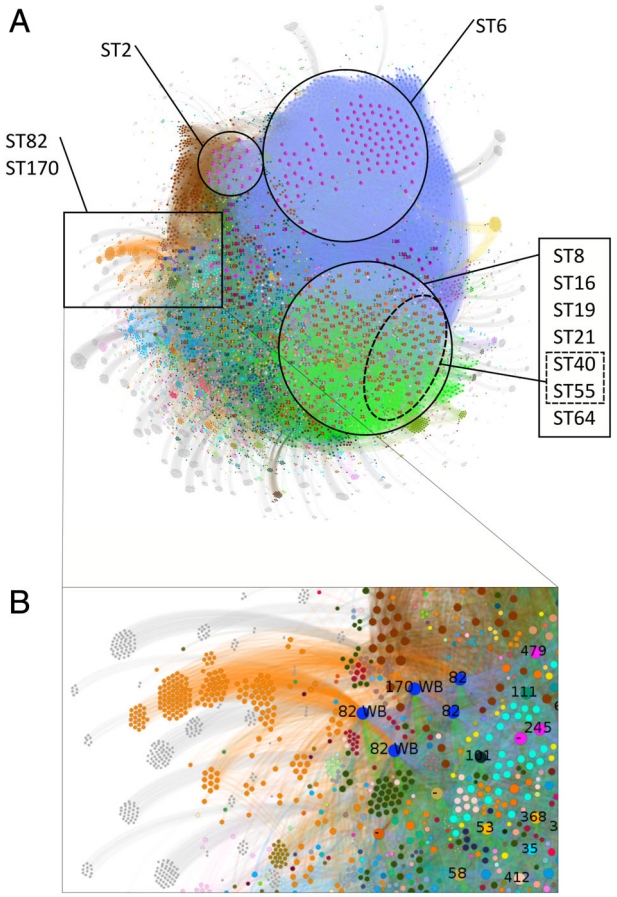

Fig. 3. A. Representation of the accessory genome network of available *E. faecalis* genomes (this study and deposited in public database) by using the ACCNET tool. Clusters of isolates (larger nodes) as well as proteins based on 95% similarity can be visualized according the colours of the nodes. The network shows a major cluster of pink (isolates) and violet (proteins) nodes that correspond to ST2, ST6 and ST9 isolates. The green cluster comprises proteins common to isolates distributed throughout the network. This explains why the violet cluster is integrated in the network despite of it appears as a specific cluster with pink isolates. The orange isolate cluster includes a variety of STs comprising ST16 and ST55 of the wild bird isolates. B. The orange protein clusters show specific proteins belonging to wild bird isolates (ST82). [Color figure can be viewed at wileyonlinelibrary.com]

The plasmid pAE63 (58.6 kb) (Fig. 5) from an ST55 *E. faecalis* isolate from a *Ciconia ciconia* host was a pheromone plasmid derivative of two pAD1-like, pB (Sequence ID: CP019514.1; Length: 65920) and pA plasmids (Sequence ID: CP019513.1; Length: 66602).

Table 2. Replicases plasmid-related genes content in the *Efs* genome database (PCR *in silico*) and in the wild bird collection.

	RCR of Rep_trans			Inc18			Rep3_small theta					RepA_N subgroups of pheromone responsive plasmids			
	rep14 pEFNP1	rep13 pC194	rep7 pT181	rep1 pIP501	rep2 pRE25	rep21 pVEF1	rep6 pAMα/pS86	rep18a pEF418	rep18b pEF1071	rep18c pCIZ2	rep4 pMBB1	rep8 pAM373	rep9 pAD1	rep17 pRUM	rep15 pLW043
Animal (n = 31)	0.00	0.00	3.23	12.90	6.45	0.00	0.00	0.00	9.68	0.00	0.00	0.00	45.16	0.00	0.00
Clinical (n = 323)	0.00	3.72	12.69	25.70	35.60	0.00	5.88	0.00	0.62	0.00	0.00	4.33	70.28	0.00	2.79
Environmental (n = 22)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.55	0.00	0.00	4.55	36.36	0.00	0.00
Wild birds (n = 97)*	2.00	5.00	55.26	0.00	10.31	5.00	33.00	4.12	3.09	3.00	0.00	0.00	78.35	8.25	0.00
Unknown (n = 56)	0.00	8.93	8.93	10.71	12.50	0.00	8.93	0.00	0.00	0.00	3.50	0.00	78.57	0.00	3.57

*The 97 isolates from wild birds were screened by PCR.

pA and pB were previously identified as two independent plasmids in another ST53 *E. faecalis* clinical isolate (GenBank CP019512.1). The pAE60, pAE65 and pAE63 contain bacteriocins and restriction–modification systems that are also present in other *E. faecalis* plasmids.

Finally, the pAE8 (77 kb) (Fig. 6) from an ST170 *E. faecalis* isolate from a *Ciconia ciconia* host is a chimera of Inc18 and pheromone-responsive plasmids ($rep1_{pIP501}$, $rep2_{pRE25}$, rep_{pAD1}, rel_{pAD1}), with various insertion sequence (IS) elements (IS*256*, IS*1216*), reflecting the origin in different hosts. The plasmid also contains genes encoding resistance to erythromycin, chloramphenicol and aminoglycosides (*str*/*aadE*) and a truncated copy of Tn*6248* (*tet*[M]-*tet*[L]-*catA9*), which lacks the *cat* gene, as well as a gene encoding resistance to cadmium.

The screening of pheromone plasmids in the *E. faecalis* genome database revealed differences in the abundance and diversity of these plasmids in the three hierBAPS wgsEFS groups. Most strains belonging to hierBAPS wgsEFS-C (ST6 and ST2) had both MOBP and MOBC relaxases (44% of the strains), whereas genomes within the wgsEFS-B (ST40) group either harboured plasmids with a single *rel* of MOBC, or did not contain any element of this plasmid family (Fig. 1). Of note is the lack of (or scarce) representation of pheromone plasmids in old lineages of hierBAPS wgsEFS-A, such as ST16, ST8 or ST21, whereas plasmids with MOBC were predominant in other STs (Fig. 1).

- *Other plasmid families.* Among the wild bird isolates, sequences of plasmid families associated with resistance to antimicrobials, such as Inc18 ($rep1/2_{pRE25}$, $rep1_{pVEF1}$, ε-ζ_{Inc18}) and pRUM ($rep17_{pRUM}$) were detected in <10%. Rolling-circle replicating (RCR) plasmids ($rep7_{pT181}$-like, $rep13_{pC194}$, $rep14_{pEFNP1}$ and $rel_{pAM\alpha}$) were detected in <5% of those isolates. Among them, the frequencies of Rep3_small theta plasmids sequences range from 3% ($rep18_{pEF418}$, $rep11_{pEF1071}$ and $repUn_{pCIZ2}$) to 33% ($rep6_{pAM\alpha1/pS86}$,) (Tables 2–3 and Fig. S3). These plasmids were more frequent in isolates from other hosts (Tables 2–3 and Supporting Information Fig. S3).

Discussion

This study represents the first analysis of *E. faecalis* isolates from wild birds in the context of the global population structure of the species. The first full genomes of isolates from *E. faecalis* obtained from wild birds were analysed, including their plasmid diversity. Although the number of genomes from wild birds' species was low, the combination of tools used to explore the core and the accessory genomes reinforce previous theoretical models

Table 3. Relaxases and toxin–antitoxin (TA) system plasmid-related genes content in the *Efs* genome database (PCR *in silico*) and in the wild bird collection.

	MOBV	MOBQ	MOBC	MOBP			TA		
	relpAMα	relpRE25	relpAD1	relpCF10	relpClZ2	relpEF1071	*par*	ε-ζ	Axe-Txe
Animal (*n* = 31)	3.23	9.68	54.84	9.68	0.00	0.00	22.58	35.48	0.00
Clinical (*n* = 323)	8.67	1.55	65.33	29.72	0.00	0.62	53.25	8.36	1.24
Environmental (*n* = 22)	0.00	0.00	9.09	22.73	0.00	0.00	22.73	0.00	0.00
Wild birds (*n* = 97)*	2.00	0.00	50.52	63.16	12.37	2.00	23.00	7.00	0.00
Unknown (*n* = 56)	12.50	5.36	55.36	25.00	0.00	0.00	39.29	7.14	0.00

that suggested frequent inter-host and inter-species horizontal and vertical migration. This would explain the apparently limited host specialization assumed for this bacterial species (Numminen *et al*., 2016).

Enterococcus. faecalis from wild birds belongs to STs broadly distributed in other hosts and thus, we can analyse the effects of transmission, migration and adaptation to various environments. A remarkable result was the distinct phylogenetic location of ST82, an ST previously associated with birds. Slight phylogenetic differences were noted by comparing the whole genome of available ST82 strains (2 ST82 and 1 ST170 analysed in this study, ST82 from penguins and other single locus variants from poultry) with other *E. faecalis* clusters. However, the study of the accessory genome revealed a group of highly specific genes. The *E. faecalis* ST82 is a common cause of amyloid arthropathy in poultry, often followed by systemic amyloidosis. Outbreaks caused by a single ST82 clone have been documented in Denmark, France, Germany and the US (Petersen *et al*., 2008), and transmission from wild birds has been included among the hypotheses to explain this global clonal expansion (Zekarias *et al*., 2000; Petersen *et al*., 2009; Fertner *et al*., 2011). Although it is tempting to suggest a relationship between the specific metabolic features of ST82 and host (birds) specificity, the lack of additional ST82 genomes precludes establishing their contribution to severe disease/habitat/host adaptation. Our results will stimulate future directed recovery of this clonal complex, knowing the specificity of a set of genes associated with ST82-ST170 (Fig. 3).

Much less host/habitat specificity was observed for other STs detected in wild birds and other hosts. ST16 is considered a 'zoonotic' *E. faecalis* lineage, due to the frequent recovery of similar ST16 strains from humans and farm animals (Bortolaia *et al*., 2016). It is often detected in strains causing farm outbreaks, contaminating commercial poultry meat and colonizing swine slots, in addition to the faecal isolates from wild birds (Freitas *et al*., 2009). Differences in the accessory genomes of ST16 and ST179 (a SLV of ST16 confined to humans) reflect different patterns of recombination. Differences in accessory genomes were also observed for members of the hierBAPS wgsEFS-B that encompasses ST55 (wgsEFS-4) and ST40 (wgsEFS-5, wgsEFS-6), which has been detected in wild birds and humans, and also ST26 (wgsEFS-7), which is common in healthy humans (Freitas *et al*., 2009; Chowdhury *et al*., 2014; Tedim *et al*., 2015; Zischka *et al*., 2015). The third large hierBAPS group, wgsEFS-C, was overrepresented in the human-associated ST6 and ST2 lineages, but has also been recovered from foodborne animals (Freitas *et al*., 2009) and rodents (Lozano *et al*., 2015). It is of note that the core phylogenetic distances contrast with the similarity of the accessory genomes (including the presence of plasmids belonging to pheromone-responsive, Inc18 and RepTrans families). This finding could be interpreted as a case of evolutionary convergence of various *E. faecalis* lineages to exploit the same or closely related niches resulting from local gene capture from ecologically related populations, but the low number of strains from other non-human hosts precludes sound conclusions.

The AccNET analysis allowed us, for the first time, to explore in depth the accessory genome of the *E. faecalis* species, which is relevant considering the role of mobile genetic elements (MGEs) as a major source of clonal diversification (Manson *et al*., 2010; Palmer and Gilmore, 2010). Until now, the accessory genome and, more specifically, the plasmidome of *E. faecalis* has been scarcely been analysed (Clewell *et al*., 2014; Lanza *et al*., 2015). The absence of plasmids in one-third of the isolates from wild birds supports the classical view that the carriage of plasmids increases with proximity to settings inhabited by humans as consequence of anthropogenic activities, including the use of antimicrobials (Souza and Eguiarte, 1997). However, the acquisition and maintenance of MGEs by *E. faecalis* depends on a variety of adaptive strategies that were not investigated here. These adaptations include the presence of CRISPR-Cas in the chromosome (Palmer and Gilmore, 2010; Price *et al*., 2016), restriction-modification (RM) and anti-RM systems, which would act against certain plasmids (e.g. pCF10); or lineage-specific bacteriocins carried by pheromone-responsive plasmids (Kurushima *et al*., 2016) or conjugative transposons (Tn*916*-like elements) that might influence the transfer of

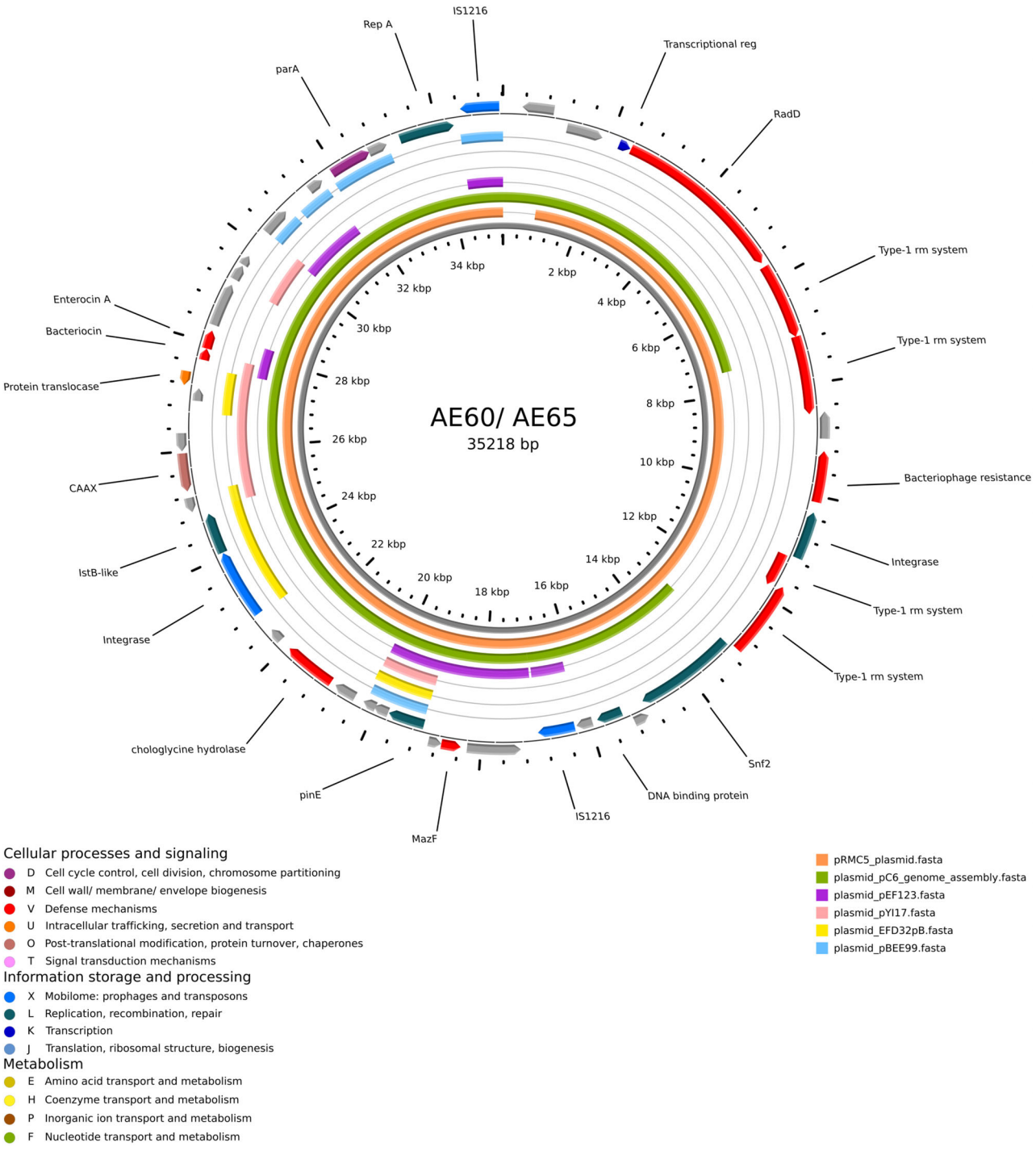

Fig. 4. Schematic representation drawn by NuRIG tool of plasmid pAE60 and pAE65. The inner black circle represents the pAE60 and pAE65 plasmid sequences, and the outer coloured circles represent the reference plasmids indicated in the key and aligned with the represented plasmid [*E. faecalis* EnGen0363 strain RMC5; *E. faecalis* isolate C6 genome assembly, contig: contig_11_length_43945_cov_30, (OANZ01000011.1); *E. faecalis* strain EF123 plasmid pEF123, (KX579977.1); *E. faecalis* plasmid pYI17 DNA, (D78257.1); *E. faecalis* D32 plasmid EFD32pB, (CP003728.1); *E. faecalis* plasmid pBEE99 (NC_013533.1)]. Arrows represent genes and indicate the orientation. Grey arrows represent hypothetical proteins of unknown function; the remaining colours represent the gene function according to the COGs database classification indicated in the key. [Color figure can be viewed at wileyonlinelibrary.com]

genetic material within and between lineages (Manson *et al*., 2010; Laverde Gomez *et al*., 2011; León-Sampedro *et al*., 2016; Price *et al*., 2016).

Pheromone-responsive plasmids are the most common plasmids of *E. faecalis*, and they are involved in the acquisition and transmission of a variety of adaptive

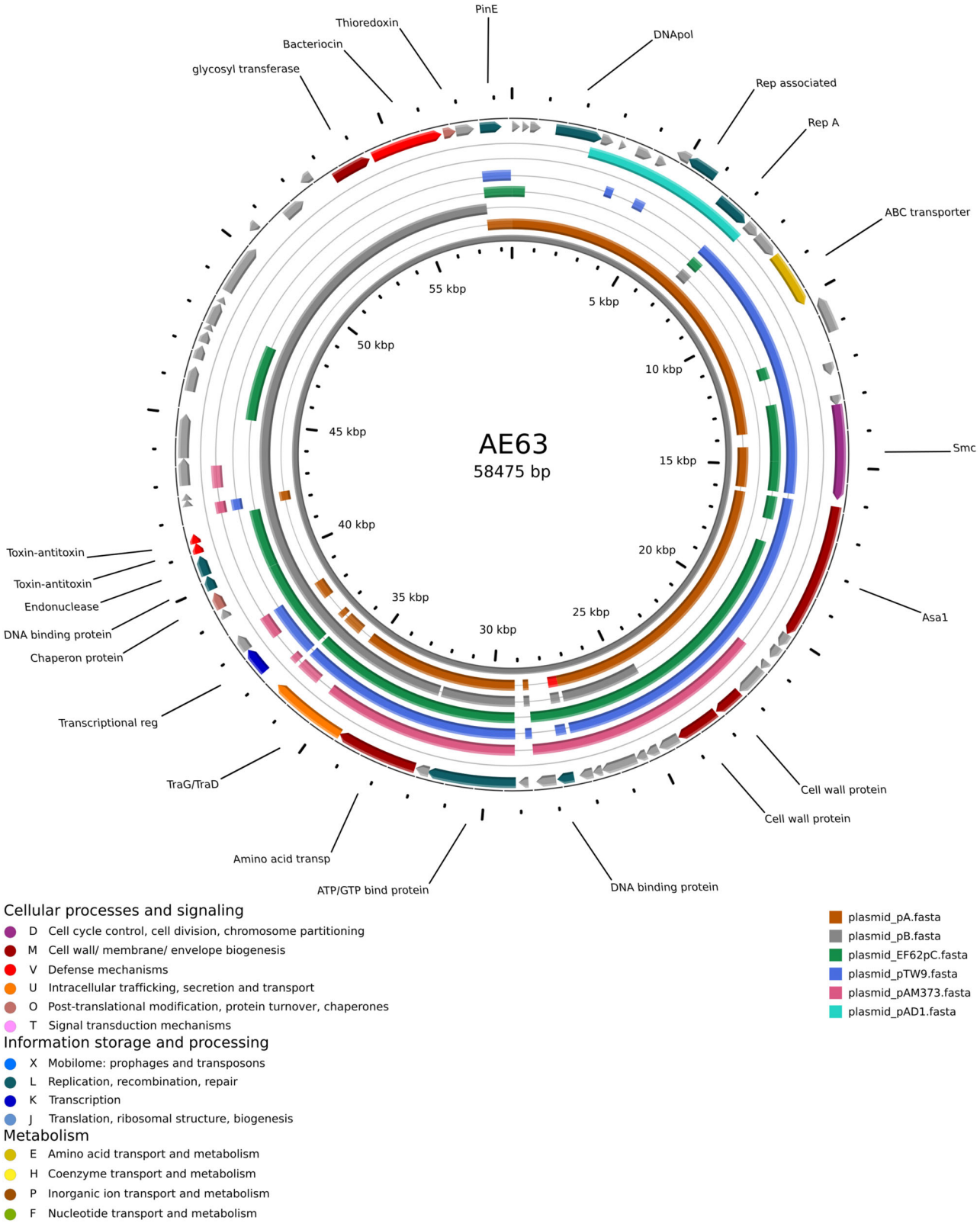

Fig. 5. Schematic representation drawn by NuRIG tool of plasmid pAE63. The inner black circle represents the pAE63 plasmid sequence, and the outer coloured circles represent the reference plasmids indicated in the key and aligned with the represented plasmid [*E. faecalis* strain CLB21560 plasmid pA, (CP019513.1); *E. faecalis* strain CLB21560 plasmid pB, (CP019514.1); *E. faecalis* 62 plasmid EF62pC, (CP002494.1); *E. faecalis* plasmid pTW9 DNA, (AB563188.1); *E. faecalis* plasmid pAM373, (AE002565.1); *E. faecalis* plasmid pAD1, (L01794.1)]. Arrows represent genes and indicate the orientation. Grey arrows represent hypothetical proteins of unknown function; the remaining colours represent the gene function according to the COGs database classification indicated in the key. [Color figure can be viewed at wileyonlinelibrary.com]

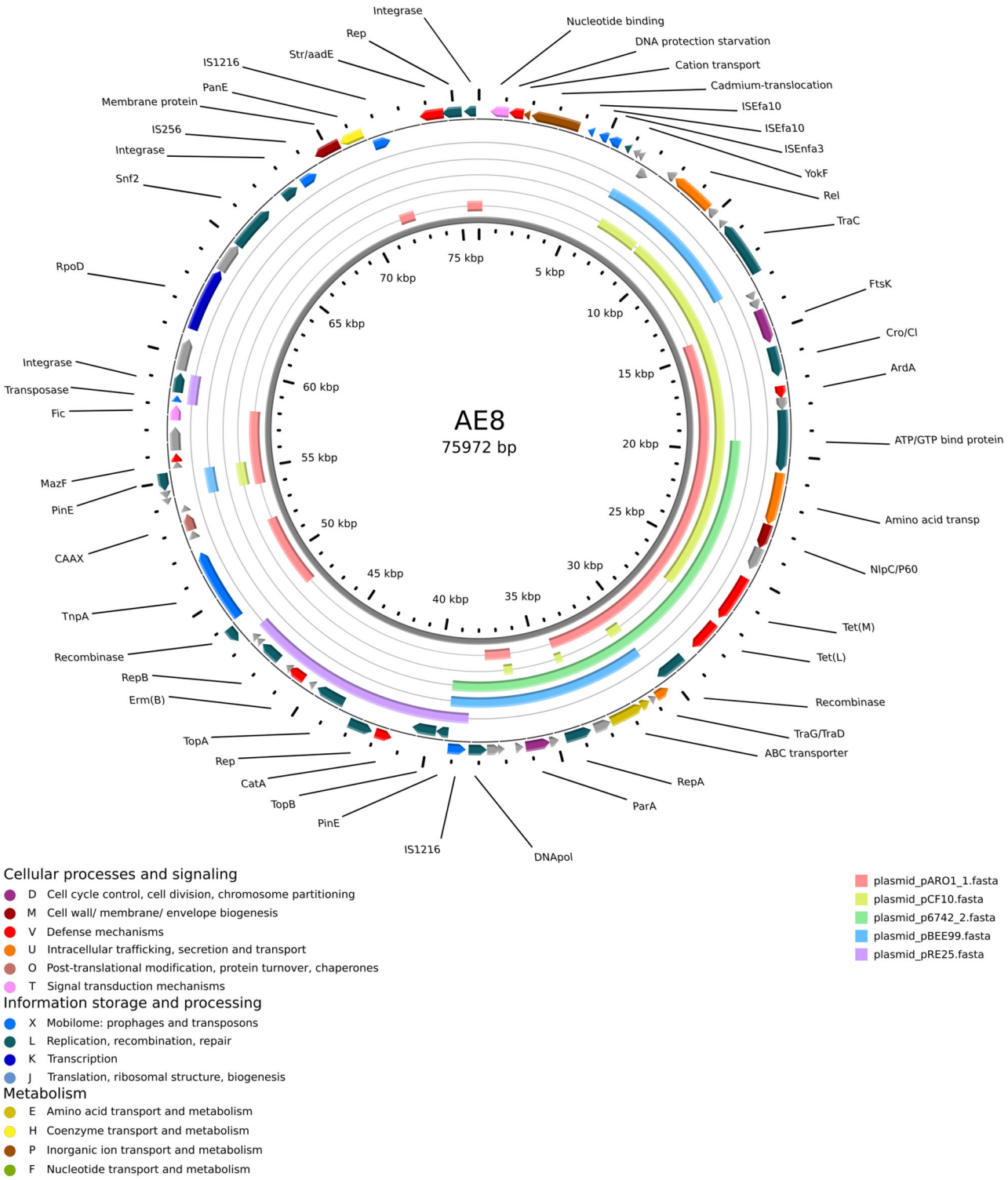

Fig. 6. Schematic representation drawn by NuRIG tool of plasmid pAE8. The inner black circle represents the *pAE8* plasmid sequence, and the outer coloured circles represent the reference plasmids indicated in the key and aligned with the represented plasmid [*E. faecalis* plasmid pCF10, (AY855841.2); *E. faecalis* ARO1/DG plasmid pARO1.1, (CP022484.1); *E. faecalis* strain 6742 plasmid p6742_2, (KY513281.1); *E. faecalis* plasmid pBEE99, (NC_013533.1); *E. faecalis* RE25 plasmid pRE25, (NC_008445.1)]. Arrows represent genes and indicate the orientation. Grey arrows represent hypothetical proteins of unknown function; the remaining colours represent the gene function according to the COGs database classification indicated in the key. [Color figure can be viewed at wileyonlinelibrary.com]

traits. Known members of this plasmid family vary in size (e.g. ca 70 kb vs. 40 kb for pCF10 and pAD1 derivatives respectively), in their response to pheromones from plasmids of different bacterial species (enterococci, staphylococci or streptococci) and the genetic cargo (virulence vs. antibiotic resistance). These differences might reflect the evolution and interplay of the major groups pAD1, pPD1 and pCF10 in different bacterial communities as has been suggested for humans (Hirt *et al*., 2002, 2005; Dunny, 2007; Gaca and Gilmore, 2016; Mansfield *et al*., 2017). Inc18 and chimeras of Inc18 and pheromone responsive plasmids are frequently involved in the acquisition and transfer of antibiotic resistance genes in the clinical setting (Freitas *et al*., 2013; Clewell *et al*., 2014; Lanza *et al*., 2015). The recovery of Inc18 from wild birds in close contact with humans (e.g. *Ciconia ciconia*) indicates that some birds are involved in the natural space of transmission of *E. faecalis* genes and strains.

In summary, core genome divergence, together with convergent gene introgression events in the accessory genome and occasional or historical epidemic peaks enriching specific populations, would help explain the population structure of *E. faecalis.* The inclusion of genomes from wild bird isolates in this study illustrates the apparent host specificity of some lineages (ST82), the incongruences between the phylogenetic trees of core and accessory genomes of other lineages (ST16/179, ST6, ST2), and the varying trajectories in hosts of still others (ST40, ST16). These results suggest alternating periods of genetic isolation and gene flow in different hosts for some *E. faecalis* lineages from wild birds. A reassessment of the population structure of the species through increased sampling of nonhuman isolates is therefore necessary to further identify host-specific populations suggested here and elsewhere (He *et al*., 2018). The presence and diversity of pheromone-responsive plasmids in most birds, as well as in human isolates, confirms the relevance of this plasmid family in the biology of *E. faecalis* as a species. Genome research has facilitated our understanding of the evolutionary history of opportunistic pathogens based on their interactions with their natural hosts. This approach should provide highly valuable information for human, animal and environmental health.

Experimental procedures

Samples and bacterial strains

We collected 103 faecal swabs from native wild birds in two Spanish veterinarian centers for nature conservation in the Central and South West of the Iberian Peninsula: the Wildlife Hospital Group for the Rehabilitation of Native Fauna [*Grupo para la Rehabilitación de la Fauna Autóctona* (GREFA)]) in the Madrid Region (40°28′22″N 3°52′20″O) and the Wildlife Recovery Center 'Action for the Wild World' in Extremadura (38°33′41″N 6°20′21″O), respectively (2001–2010). The sampled animals were migratory and raptor birds that suffered collisions with vehicles or power pylons, gunshot wounds or that were orphaned chicks. Table 1 describes the wild bird sample, which represents the diversity of Iberian Peninsula wildlife (33 species of 10 orders with various lifestyle and dietary habits).

Individual faecal swabs were sent on Amies transport medium (Biomedics, Spain) to the laboratory within 24 h after sampling and were preserved at −80°C until processing. Samples were screened for *E. faecalis* using m-*Enterococcus* selective media and were further identified by biochemical tests (Facklam *et al*., 2002). Antimicrobial susceptibility tests against eight antibiotics (ampicillin, gentamicin, kanamycin, streptomycin, tetracycline, erythromycin, vancomycin and chloramphenicol) were performed by the agar dilution method according to Clinical and Laboratory Standards Institute (CLSI) guidelines (Humphries et al., 2018). The identification of the isolates was confirmed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS; Bruker Daltonics, Billerica, MA, USA) and the antibiotic susceptibility by disc diffusion disc when isolates were subcultured for further analysis. The presence of putative virulence genes coding for enterococcal surface protein, cytolysin/hemolysin, gelatinase and aggregation substance was screened for using polymerase chain reaction (PCR) (Coque *et al*., 1995).

Population structure analysis

Clonal relatedness was established by PFGE using *Sma*I as a restriction enzyme and multilocus sequence typing (MLST; www.mlst.net) (Freitas *et al*., 2013). Clustering of all *E. faecalis* STs available in the database was performed with goeBURST (www.phyloviz.net). Briefly, the separation of large DNA molecules for the PFGE was achieved CHEF-DRIII device using different running settings depending on the expected fragments size: *Sma*I [1–20s, 38 h (*E. faecalis*)] for the clonal relatedness, and S1 nuclease (5–25 s for 6 h followed by 30–45 s for 18 h) for the plasmid characterization. After electrophoresis through agarose gels, DNA was staining with the GelRed™ dye and visualized by illumination with 300-nm UV light. Clonal relationships were established by comparison of DNA restriction patterns (digested with *Sma*I) following the Tenover *et al*. guidelines (Tenover *et al*., 1995; Roberts and Mullany, 2011).

A maximum likelihood phylogenetic tree was built with the sequences of the seven concatenated MLST genes by using the IQ-TREE software (Nguyen *et al*., 2015) and the best model autodetection feature. We used the Interactive Tree of Life (iTOL) tool (Letunic and Bork, 2011) to

represent the obtained tree and the clonal complexes obtained with goeBURST and the STs found in the wild bird collection. Next, the population structure was analysed with hierBAPS, using as queries either the concatenated sequences of the seven MLST loci or complete *E. faecalis* genome sequences (Cheng *et al*., 2013).

Characterization of plasmids

Plasmid profiling was determined using a modified Barton's method protocol (Barton *et al*., 1995; Freitas *et al*., 2016). Plasmid categorization was based on the diversity of genes encoding rep-initiator proteins (*rep*), relaxases (*rel*) and toxin–antitoxin (TA) systems of Gram-positive plasmids using PCR-typing schemes (Supporting Information Table S4) (Jensen *et al*., 2010; Wardal *et al*., 2013; Clewell *et al*., 2014; Freitas *et al*., 2016). Characterization of pheromone-responsive plasmids was confirmed by Southern blot hybridization of S1-digested genomic DNA with specific probes for *rel*, *rep* and TA genes obtained by PCR using DNA from reference plasmids as positive controls. Briefly, the Southern blot hybridization requires different steps: DNA transfer and fixation; DNA labelling and hybridization; post-hybridization washes; signal detection (Southern, 1992). Chemiluminescence detection was carried out using a dioxetane substrate (CDP-Star reagent) which is catalysed by the probe-bound alkaline phosphatase generating a rapid light output. The signal was detected by ChemiDoc™ XRS (Bio-rad). The DNA immobilized on a nylon membrane was hybridized with different DNA probes at different times without significant loss of signal strength as described. We assessed the genomic location of genes with specific replicase gene probes under high stringency conditions (Supporting Information Table S4).

The plasmid content (number and diversity of *rep*, *rel* and TA sequences) of *E. faecalis* isolates from the wild birds in this study was compared with those of the 441 genomes available in public databases (31 from animals, 323 from hospitalized patients, 22 from environmental samples, 6 from nonhospitalized patients, 4 from food and 55 from unknown sources) (Supporting Information Table S2). To this end, we modified an *in silico* PCR pipeline applying the same PCR schemes mentioned above (*rep*/*rel*/TA) (https://github.com/leonsampedro/primersearch/blob/master/parse_primersearch_mod.R). Briefly, we implemented the PrimerSearch EMBOSS tool with a homemade script written in the R environment; we transferred the results to an easy-to-handle table and further analysed the data obtained. The *E. faecalis* genomes from public databases were classified according to their origin (human, animal and environmental), with information obtained from the Pathosystems Resource Integration Center database (Wattam *et al*., 2014). Finally, we represented the plasmid features found in the phylogenetic trees derived from the core and accessory genome using the iTOL tool.

Comparative genomics

Five *E. faecalis* isolates from wild bird species with disparate lifestyles were fully sequenced. Total DNA was extracted from 5 ml of overnight cultures using the Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA) and DNA concentration was measured using a Qubit™ Fluorometer and Nanodrop 1000 (Thermo Scientific, Waltham, MA, USA). The library preparation was carried out with the Nextera DNA Flex Library Prep Kit (Illumina, San Diego, CA). Sequencing was performed using a standard 2 × 100 base protocol in a Genome Analyser IIx Illumina HiSeq 2500 platform (Illumina). The paired-end reads were *de novo* assembled using SPAdes (v.3.5.0) (Bankevich *et al*., 2012). The quality of the genome assemblies was evaluated by QUAST (Supporting Information Table S5) (Gurevich *et al*., 2013).

Both the core and accessory genomes of the sequenced strains were compared with the 441 *E. faecalis* genomes available at the National Centre for Biotechnology Information (NCBI) database. Sequencing reads of the five genomes from wild birds were trimmed, filtered by quality and aligned and mapped to the reference genome V583 (Accession no. NC_004668.1). Variants were called using Snippy v2.5 (https://github.com/tseemann/snippy), requiring a minimum base quality of 20 and a minimum read coverage of 10×.

For the analysis of the core genome (the set of homologous nucleotides present in all the isolates when mapped against the same reference), an alignment of the single nucleotide polymorphisms (SNPs) obtained with Snippy v2.5 was used to infer a phylogeny (Supporting Information file S1). A maximum-likelihood tree was generated using IQ-TREE with the feature of automated detection of the best evolutionary model (Nguyen *et al*., 2015). The tree was plotted using the iTOL tool (Letunic and Bork, 2011).

For the analysis of the accessory genome, we used AccNET, a tool that allows us to infer the accessory genome from the proteomes and cluster them based on 95% protein similarity (Lanza *et al*., 2017). The set of representative proteins was used to build a binary matrix (presence/absence of proteins in the accessory genome) in the R-environment (http://www.r-project.org) and a cladogram to classify the strains according to the accessory genomes. The Euclidean distance was calculated by the *dist* function and a hierarchical clustering was performed with UPGMA using the *hclust* function in the R environment. This accessory genome binary matrix was also analysed by Pangenome Neighbour Identification for Bacterial Populations (PANINI; http://panini.wgsa.net/), a

computationally scalable method for identifying the neighbours of each isolate in a data set with the aim of confirming the results obtained by the cladogram (Abudahab *et al*., 2017). Protein functions were inferred by interrogating the COG database and using OrthoMCL, a tool for genome-scale screening (Fischer *et al*., 2011).

To analyse the congruence between the core and the accessory trees, we performed correlations between the distance matrices by the application of the Mantel test. Mantel regressions were performed using the package Vegan in the R environment (http://www.R-project.org).

Finally, the plasmid content of the 5 sequenced strains was inferred by combining the PLACNET (Lanza *et al*., 2014) and plasmidSPAdes (v.3.5.0) tools (Antipov *et al*., 2016) to analyse short reads. The plasmid contigs obtained by both methods were annotated with Prokka v1.12 (Seemann, 2014) and further compared with available sequences at the NCBI database by BLASTN and BLASTP. Afterwards, plasmid backbones were drawn with Easyfig v2.2.2 (Sullivan *et al*., 2011) and Nurig tools (https://github.com/valflanza/nurig).

Statistical analysis

Differences in the plasmid content of hosts/sources were considered significant when $p < 0.05$. The Lilliefors test was used to test the normality of the plasmid data distribution, including the complete genome database and the collection of wild bird isolates analysed. The nonparametric Kruskal–Wallis test and the post hoc Tukey's range test were used to compare data from different sources. Statistical tests were performed within the R environment.

SRA accession numbers

The five genomes sequence in this work were deposited to the Sequence Read Archive (SRA) database with the accession numbers SRR8587711 (*Efs* AE60), SRR8587709 (*Efs* AE8), SRR8587712 (*Efs* AE63), SRR8587710 (*Efs* AE51) and SRR8587713 (*Efs* AE65).

Acknowledgements

This study was supported by the Joint Programming Initiative in Antimicrobial Resistance (JPIAMR Third call, STARCS, JPIAMR2016-AC16/00039), the Instituto de Salud Carlos III of Spain/Ministry of Economy and Competitiveness- and the European Development Regional Fund 'A way to achieve Europe' (ERDF) for co-founding the Spanish R&D National Plan Estatal de I + D + i 2013-2016 (PI15-0512), CIBER (CIBER in Epidemiology and Public Health, CIBERESP; CB06/02/0053), and the Regional Government of Madrid (InGeMICS- B2017/BMD-3691). RLS was further funded by a Research Grant of ISCIII (Instituto de Salud Carlos III of Spain) and the Sociedad Española de Enfermedades Infecciosas y Microbiología Clínica (SEIMC).

Conflict of interest

The authors have no conflict of interest to declare.

References

Abudahab, K., Prada, J.M., Yang, Z., Bentley, S.D., Croucher, N. J., Corander, J., and Aanensen, D.M. (2017) PANINI: Pangenome neighbor identification for bacterial populations. *Microb Genom* **5**: e000220.

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

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., *et al*. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**: 455–477.

Barton, B.M., Harding, G.P., and Zuccarelli, A.J. (1995) A general method for detecting and sizing large plasmids. *Anal Biochem* **226**: 235–240.

Bortolaia, V., Espinosa-Gongora, C., and Guardabassi, L. (2016) Human health risks associated with antimicrobial-resistant enterococci and *Staphylococcus aureus* on poultry meat. *Clin Microbiol Infect* **22**: 130–140.

Cheng, L., Connor, T.R., Siren, J., Aanensen, D.M., and Corander, J. (2013) Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Mol Biol Evol* **30**: 1224–1228.

Chowdhury, S.A., Nallapareddy, S.R., Arias, C.A., Murray, B.E., and Gilligan, P.H. (2014) The majority of a collection of U.S. endocarditis *Enterococcus faecalis* isolates obtained from 1974 to 2004 lack capsular genes and belong to diverse, non-hospital-associated lineages. *J Clin Microbiol* **52**: 549–556.

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

Coque, T.M., Patterson, J.E., Steckelberg, J.M., and Murray, B.E. (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**: 1223–1229.

Dunny, G.M. (2007) The peptide pheromone-inducible conjugation system of *Enterococcus faecalis* plasmid pCF10: cell-cell signalling, gene transfer, complexity and evolution. *Philos Trans R Soc B Biol Sci* **362**: 1185–1193.

Facklam, R., Carvalho, M., and Teixeira, L. (2002) History, taxonomy, biochemical characteristics and antibiotic susceptibility testing of enterococci. In *The Enterococci: Pathogenesis, Molecular Biology and Antibiotic Resistance*, Gilmore, M., Clewell, D., Courvalin, P., Dunny, G., Murray, B.,

and Rice, L. (eds). Washington, DC, USA: American Society for Microbiology, pp. 1–54.

Fertner, M.E., Olsen, R.H., Bisgaard, M., and Christensen, H. (2011) Transmission and genetic diversity of *Enterococcus faecalis* among layer chickens during hatch. *Acta Vet Scand* **53**: 56.

Fischer, S., Brunk, B.P., Chen, F., Gao, X., Harb, O.S., Iodice, J.B., *et al*. (2011) Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Curr Protoc Bioinformatics* **6** Unit 6.12.1-19.

Freitas, A.R., Elghaieb, H., León-Sampedro, R., Abbassi, M. S., Novais, C., Coque, T.M., *et al*. (2017) Detection of optrA in the African continent (Tunisia) within a mosaic *Enterococcus faecalis* plasmid from urban wastewaters. *J Antimicrob Chemother* **72**: 3245–3251.

Freitas, A.R., Novais, C., Ruiz-Garbajosa, P., Coque, T.M., and 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**: 1104–1111.

Freitas, A.R., Novais, C., Tedim, A.P., Francia, M.V., Baquero, F., Peixe, L., and Coque, T.M. (2013) Microevolutionary events involving narrow host plasmids influences local fixation of vancomycin-resistance in enterococcus populations. *PLoS One* **8**: e60589.

Freitas, A.R., Tedim, A.P., Francia, M.V., Jensen, L.B., Novais, C., Peixe, L., *et al*. (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.

Gaca, A.O., and Gilmore, M.S. (2016) Killing of VRE *Enterococcus faecalis* by commensal strains: evidence for evolution and accumulation of mobile elements in the absence of competition. *Gut Microbes* **7**: 90–96.

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

Guzman Prieto, A.M., van Schaik, W., Rogers, M.R.C., Coque, T.M., Baquero, F., Corander, J., and Willems, R.J. L. (2016) Global emergence and dissemination of enterococci as nosocomial pathogens: attack of the clones? *Front Microbiol* **7**: 788.

He, Q., Hou, Q., Wang, Y., Li, J., Li, W., Kwok, L.-Y., *et al*. (2018) Comparative genomic analysis of *Enterococcus faecalis*: insights into their environmental adaptations. *BMC Genomics* **19**: 527.

Hirt, H., Manias, D.A., Bryan, E.M., Klein, J.R., Marklund, J. K., Staddon, J.H., *et al*. (2005) Characterization of the pheromone response of the *Enterococcus faecalis* conjugative plasmid pCF10: complete sequence and comparative analysis of the transcriptional and phenotypic responses of pCF10-containing cells to pheromone induction. *J Bacteriol* **187**: 1044–1054.

Humphries, R.M., Ambler, J., Mitchell, S.L., Castanheira, M., Dingle, T., Hindler, J.A., *et al*. (2018) CLSI methods development and standardization working group best practices for evaluation of antimicrobial susceptibility tests. *J Clin Microbiol* **56**(4): e01934–17CLSI (2014).

Hirt, H., Schlievert, P.M., and Dunny, G.M. (2002) In vivo induction of virulence and antibiotic resistance transfer in *Enterococcus faecalis* mediated by the sex pheromone-sensing system of pCF10. *Infect Immun* **70**: 716–723.

Jensen, L.B., Garcia-Migura, L., Valenzuela, A.J.S., Løhr, M., Hasman, H., and Aarestrup, F.M. (2010) A classification system for plasmids from enterococci and other Gram-positive bacteria. *J Microbiol Methods* **80**: 25–43.

Kuch, A., Willems, R.J.L., Werner, G., Coque, T.M., Hammerum, A.M., Sundsfjord, A., *et al*. (2012) Insight into antimicrobial susceptibility and population structure of contemporary human *Enterococcus faecalis* isolates from Europe. *J Antimicrob Chemother* **67**: 551–558.

Kurushima, J., Ike, Y., and Tomita, H. (2016) Partial diversity generates effector immunity specificity of the Bac41-like Bacteriocins of *Enterococcus faecalis* clinical strains. *J Bacteriol* **198**: 2379–2390.

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

Lanza, V.F., Tedim, A.P., Martínez, J.L., Baquero, F., and Coque, T.M. (2015) The Plasmidome of Firmicutes: impact on the emergence and the spread of resistance to antimicrobials. *Microbiol Spectr* **3**: PLAS-0039-2014.

Lanza, V.F., de Toro, M., Garcillán-Barcia, M.P., Mora, A., Blanco, J., Coque, T.M., and 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.

Larsen, J., Schønheyder, H.C., Singh, K.V., Lester, C.H., Olsen, S.S., Porsbo, L.J., *et al*. (2011) Porcine and human community reservoirs of *Enterococcus faecalis* , Denmark. *Emerg Infect Dis* **17**: 2395–2397.

Laverde Gomez, J.A., Hendrickx, A.P.A., Willems, R.J., Top, J., Sava, I., Huebner, J., *et al*. (2011) Intra- and interspecies genomic transfer of the *Enterococcus faecalis* pathogenicity Island. *PLoS One* **6**: e16720.

Lebreton, F., Manson, A.L., Saavedra, J.T., Straub, T.J., Earl, A.M., and Gilmore, M.S. (2017) Tracing the enterococci from Paleozoic origins to the hospital. *Cell* **169**: 849–861 e13.

León-Sampedro, R., Novais, C., Peixe, L., Baquero, F., and Coque, T.M. (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.

Letunic, I., and Bork, P. (2011) Interactive tree of life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* **39**: W475–W478.

Lozano, C., González-Barrio, D., García, J.T., Ceballos, S., Olea, P.P., Ruiz-Fons, F., and Torres, C. (2015) Detection of vancomycin-resistant *Enterococcus faecalis* ST6-vanB2 and *E. faecium* ST915-vanA in faecal samples of wild *Rattus rattus* in Spain. *Vet Microbiol* **177**: 168–174.

Mansfield, J.M., Herrmann, P., Jesionowski, A.M., and Vickerman, M.M. (2017) Streptococcus gordonii pheromone s.g.cAM373 may influence the reservoir of antibiotic resistance determinants of *Enterococcus faecalis* origin in the oral metagenome. *J Med Microbiol* **66**: 1635–1639.

Manson, J.M., Hancock, L.E., and Gilmore, M.S. (2010) Mechanism of chromosomal transfer of *Enterococcus*

faecalis pathogenicity island, capsule, antimicrobial resistance, and other traits. *Proc Natl Acad Sci* **107**: 12269–12274.

McBride, S.M., Fischetti, V.A., Leblanc, D.J., Moellering, R. C., and Gilmore, M.S. (2007) Genetic diversity among *Enterococcus faecalis*. *PLoS One* **2**: e582.

Moles, L., Gómez, M., Jiménez, E., Fernández, L., Bustos, G., Chaves, F., *et al*. (2015) Preterm infant gut colonization in the neonatal ICU and complete restoration 2 years later. *Clin Microbiol Infect* **21**: 936.e1–936.e10.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B. Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**: 268–274.

Numminen, E., Gutmann, M., Shubin, M., Marttinen, P., Méric, G., van Schaik, W., *et al*. (2016) The impact of host metapopulation structure on the population genetics of colonizing bacteria. *J Theor Biol* **396**: 53–62.

Oravcova, V., Ghosh, A., Zurek, L., Bardon, J., Guenther, S., Cizek, A., and Literak, I. (2013) Vancomycin-resistant enterococci in rooks (*Corvus frugilegus*) wintering throughout Europe. *Environ Microbiol* **15**: 548–556.

Oravcova, V., Svec, P., and Literak, I. (2017) Vancomycin-resistant enterococci with *vanA* and *vanB* genes in Australian gulls. *Environ Microbiol Rep* **9**: 316–318.

Oravcova, V., Zurek, L., Townsend, A., Clark, A.B., Ellis, J. C., Cizek, A., and Literak, I. (2014) American crows as carriers of vancomycin-resistant enterococci with vanA gene. *Environ Microbiol* **16**: 939–949.

Palmer, K.L., and Gilmore, M.S. (2010) Multidrug-Resistant Enterococci Lack CRISPR-cas. *MBio* **1**: e00227-10–e00227-19.

Petersen, A., Chadfield, M.S., Christensen, J.P., Christensen, H., and Bisgaard, M. (2008) Characterization of small-colony variants of *Enterococcus faecalis* isolated from chickens with amyloid arthropathy. *J Clin Microbiol* **46**: 2686–2691.

Petersen, A., Christensen, H., Philipp, H.-C., and Bisgaard, M. (2009) Clonality of *Enterococcus faecalis* associated with amyloid arthropathy in chickens evaluated by multilocus sequence typing (MLST). *Vet Microbiol* **134**: 392–395.

Price, V.J., Huo, W., Sharifi, A., and Palmer, K.L. (2016) CRISPR-Cas and restriction-modification act additively against conjugative antibiotic resistance plasmid transfer in *Enterococcus faecalis*. *mSphere* **1**: e00064–e00016.

Roberts, A.P., and Mullany, P. (2011) Tn916-like genetic elements: a diverse group of modular mobile elements conferring antibiotic resistance. *FEMS Microbiol Rev* **35**: 856–871.

Ruiz-Garbajosa, P., Bonten, M.J.M., Robinson, D.A., Top, J., Nallapareddy, S.R., Torres, C., *et al*. (2006a) 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.

Ruiz-Garbajosa, P., Cantón, R., Pintado, V., Coque, T.M., Willems, R., Baquero, F., and del Campo, R. (2006b) Genetic and phenotypic differences among *Enterococcus faecalis* clones from intestinal colonisation and invasive disease. *Clin Microbiol Infect* **12**: 1193–1198.

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

Southern, E.M. (1992) Detection of specific sequences among DNA fragments separated by gel electrophoresis. 1975. *Biotechnology* **24**: 122–139.

Souza, V., and Eguiarte, L.E. (1997) Bacteria gone native vs. bacteria gone awry?: plasmidic transfer and bacterial evolution. *Proc Natl Acad Sci U S A* **94**: 5501–5503.

Sullivan, M.J., Petty, N.K., and Beatson, S.A. (2011) Easyfig: a genome comparison visualizer. *Bioinformatics* **27**: 1009–1010.

Tedim, A.P., Ruiz-Garbajosa, P., Corander, J., Rodríguez, C. M., Cantón, R., Willems, R.J., *et al*. (2015) Population biology of intestinal enterococcus isolates from hospitalized and nonhospitalized individuals in different age groups. *Appl Environ Microbiol* **81**: 1820–1831.

Tenover, F.C., Arbeit, R.D., Goering, R.V., Mickelsen, P.A., Murray, B.E., Persing, D.H., and 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.

Wardal, E., Gawryszewska, I., Hryniewicz, W., and Sadowy, E. (2013) Abundance and diversity of plasmid-associated genes among clinical isolates of Enterococcus faecalis. *Plasmid* **70**: 329–342.

Wattam, A.R., Abraham, D., Dalay, O., Disz, T.L., Driscoll, T., Gabbard, J.L., *et al*. (2014) PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res* **42**: D581–D591.

Willems, R.J.L., Top, J., van Schaik, W., Leavis, H., Bonten, M., Sirén, J., *et al*. (2012) Restricted gene flow among hospital subpopulations of *Enterococcus faecium*. *MBio* **3**: e00151–e00112.

Ben Yahia, H., Chairat, S., Hamdi, N., Gharsa, H., Ben Sallem, R., Ceballos, S., *et al*. (2018) Antimicrobial resistance and genetic lineages of faecal enterococci of wild birds: emergence of vanA and vanB2 harbouring *Enterococcus faecalis*. *Int J Antimicrob Agents* **52**: 936–941.

Zekarias, B., Landman, W.J., Tooten, P.C., and Gruys, E. (2000) Leukocyte responses in two breeds of layer chicken that differ in susceptibility to induced amyloid arthropathy. *Vet Immunol Immunopathol* **77**: 55–69.

Zischka, M., Künne, C.T., Blom, J., Wobser, D., Sakınç, T., Schmidt-Hohagen, K., *et al*. (2015) Comprehensive molecular, genomic and phenotypic analysis of a major clone of *Enterococcus faecalis* MLST ST40. *BMC Genomics* **16**: 175.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Maximum likelihood phylogenetic tree derived from a concatenated sequence of the seven MLST loci and the STs (and CCs) of each isolate in the complete *E. faecalis* MLST database (last update October 2017) (https://pubmlst.org/efaecalis/), including the 97 isolates from wild birds analysed here. Data were analysed by the electronic Based Upon Related Sequence Types (eBURST) obtaining the ST complexes represented in both the outer circle of the figure and the coloured clades within the tree. The second and third circles represent the groups and subgroups detected using hierarchical BAPS (hierBAPS). Dots at the branches tips match with colours of the hierBAPS subgroups. STs of wild bird strains identified in this study are marked with red arrows.

Previous studies of the *E. faecalis* population structure that analysed MLST datasets using BAPS did not yield consistent results (Tedim *et al*., 2015). Due to the increase in the number of STs included in the *E. faecalis* database (https://pubmlst.org/efaecalis/) in the last two years (i.e., 218 new STs with a total of 1452 isolates corresponding to 741 STs), we repeated the analysis, excluding ST80 once again, given that it remains as an outgroup that distorts the results from both BAPS and phylogenetic analyses. In this new analysis, we observed a strong, positive correlation (r = 0.956) between the new results and those obtained in a previous study (Tedim *et al*., 2015). The consistency of the results with BAPS between both analyses led us to apply hierBAPS, an upgraded version of the BAPS that allows clustering the sequences into genetically distinct groups and studying admixture (https://github.com/gtonkinhill/rhierbaps).

Fig. S2. Annotated output of the PANINI algorithm applied to the accessory genome binary matrix (presence/ausence of accessory proteins) of 446 *E. faecalis* isolates. Each node represents an isolate, each of which is coloured according to its ST. Clusters of isolates belonging to the same sequence type are circled and annotated.

Fig. S3. Plasmid gene content of *E. faecalis* genomes from the NCBI whole genome database. Each colour represents the source of the isolates analysed: blue, animal (n = 31); red, clinical (n = 323); green, environmental (n = 22) and purple, genomes of unknown origin (n = 56). Data were obtained by a homemade '*in silico*' PCR pipeline written in the R environment and using the Jensen *et al*. (2010) PCR scheme. PrimerSearch EMBOSS tool was implemented and modified to an easy to handle script. The genomes from the database were classified by source after extracting isolate information from the Pathosystems Resource Integration Center (PATRIC) database (Wattam *et al*., 2014). Abbreviations: Reps = replicases; Rels = relaxases; TA = toxin-antitoxin systems.

Table S1. Classification of *E. faecalis* obtained from wild birds according to PFGE, MLST and susceptibility to antibiotics analyses.

Table S2. Main features of the genome sequences analysed, including obtained from wild birds and from GenBank database.

Table S3. Accessory genome network obtained from the 446 whole genome sequences studied. In red, specific proteins belonged to the ST82 lineage.

Table S4. Oligonucleotides used in this study. Primers used for the identification of relaxases will be available upon request (Goicoechea P, Romo M, Coque TM, *et al*. Identification of enterococcal plasmids by multiplex-PCR-based relaxase typing (Abstract number: P1669). In: 18th European Congress of Clinical Microbiology and Infectious Diseases. Barcelona, Spain, 2008.)

Table S5. Assembly quality of the 5 genomes sequenced in this study obtained by QUAST.