Efagins evolved independently to target the enterococcal cell wall

Janira Prichula,[1,2,3]* Abigail L. Manson,[3] Joshua T. Smith,[3] Natia Karumidze,[4,5] Gracie Richards,[3] Suelen S. Mello,[1,2] Artur Czajkowski,[6] Mohammed Kaplan,[6] Ashlee M. Earl,[3]* and Michael S. Gilmore,[1,2,3]*

[1]Department of Ophthalmology, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA, USA

[2]Department of Microbiology, Harvard Medical School, Boston, MA, USA

[3]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA

[4]Eliava Institute of Bacteriophages, Tbilisi, Georgia

[5]European University, Tbilisi, Georgia

[6]University of Chicago, Chicago, IL, USA

***Co-corresponding authors:** jprichula@meei.harvard.edu (J.P) and michael_gilmore@meei.harvard.edu (M.S.G.)

ABSTRACT

Enterococci are major causes of multidrug-resistant hospital infections, underscoring the need for new antibacterial strategies. Here, we describe a previously unrecognized inhibitor class, efagins (***E**nterococcus **f**aecalis* ph**ag**e-related **in**hibitor**s**). Efagins are intrinsic to *E. faecalis*, the most widely distributed generalist among enterococci, and are structurally reminiscent of phage tails yet show only distant evolutionary relatedness to known phages or phage-like elements. The 14.6-kb efagin gene cluster is highly conserved except for a tail fiber–like encoding region that varies across five distinct classes. These classes correlate with differences in the rhamnose-rich Epa cell wall polysaccharide, the efagin binding target as demonstrated by domain-swap experiments. We identified at least 20 *E. faecalis epa* genotypes, and the efagins characterized here inhibit strains representing 17 of them. Importantly, efagin activity extends beyond *E. faecalis*: they inhibit multiple enterococcal species, including lineages of vancomycin-resistant *E. faecium*, highlighting their potential as precision antibacterials.

KEYWORDS

Enterococcus faecalis; generalist species, efagin, phage-related elements; antimicrobial agents; Epa; Rha-CWPS; hospital acquired infection, VRE; *Enterococcus faecium*

INTRODUCTION

The genus *Enterococcus* includes two leading causes of multidrug-resistant (MDR) infections globally, *E. faecalis* and *E. faecium.*[1–3] Projected to have evolved at the time animals first terrestrialized 400 – 500 MYA, their emergence paralleled the gain of functions that facilitate transmission among invertebrate and vertebrate hosts on land, but now also contribute to survival in the harsh environment of modern hospitals.[4] Through coevolution with land animals, including the most numerically abundant, diverse, and ecologically dispersed group, arthropods, enterococci are naturally exposed to a vast range of selections, from diverse innate defenses[5] to natural antimicrobial substances occurring in the diets of those hosts. In a recent planetary survey of enterococcal distribution, the species *E. faecalis* was recovered from the guts of the widest range of hosts, spanning insects to humans, constituting by far the greatest generalist of the genus.[6] This contrasts sharply with the discovery of many rare, apparently highly specialized enterococci, including 18 newly described species isolated from the guts of arthropods and animals that

consume them.[6] This raised the question as to what unique features of *E. faecalis*– one of the earliest species of *Enterococcus* to emerge at a time when terrestrial life was dominated by arthropods[4]– allow it to successfully invade diverse gut consortia and compete with more specialized species as it is conveyed up and down the food chain.

Comparing genomes of >3,000 sequenced *E. faecalis*, we observed that despite the ancient emergence and subsequent diversification of the species, a 14.6 kb cluster of genes previously annotated as encoding a small or partial phage,[7] which could not be ascribed phage function even in the presence of helper phages,[8] was highly conserved in all genomes, with reading frames well maintained. We interpreted this as evidence that the product of this block of genes contributes importantly to fitness in the highly diverse environments in which *E. faecalis* occurs.

Here we show that this cluster of genes, which shares only a distant sequence relationship to an inferred non-contractile tail sequence from an *E. faecalis* phage, encodes an independently derived entity with antimicrobial activity that targets the rhamnose-rich enterococcal cell wall polysaccharide (Rha-CWPS) of enterococci and possibly other microbes. We term the product of this gene cluster *efagin* (***E***. ***f****aecalis* ph**ag**e-related **in**hibitor). Although sharing very limited sequence identity, efagins exhibit a functional relationship to a group of independently derived elements termed tailocins,[9,10] which include pyocins of *Pseudomonas aeruginosa*,[11–16] monocins of *Listeria monocytogenes*,[17–20] and diffocins of *Clostridiodes difficile*.[21,22] Efagin activity extends beyond *E. faecalis* and includes other enterococcal species, including some vancomycin-resistant *E. faecium.* Efagins are programmable through targeting domain swaps and may constitute useful new tools for controlling the proliferation of specific, highly antibiotic-resistant hospital-adapted clones, while sparing a wide range of commensal enterococci and other microbes. Moreover, efagins represent a newly identified but ubiquitous selection that likely drives Rha-CWPS variation in enterococci, affecting phage susceptibility and contributing to strain succession in the guts of its wide-ranging hosts.

RESULTS

Efagins are encoded by all E. faecalis genomes

To assess the ubiquity of efagin in *E. faecalis*, we analyzed a dataset of 3,068 genome sequences. This included 2,676 publicly available genomes (National Center for Biotechnology Information [NCBI][23]), and 392 more sequenced by our research group (including 38 newly sequenced for this study [**Table S1**]). Every *E. faecalis* genome encodes all genes associated with efagin expression, except for five publicly available, low quality draft genomes for which partial hits were found. All high-quality drafts (fewer than 75 contigs, N=2278) possess the entire element. Moreover, across all major lineages spanning the species phylogeny (N=493), all efagin associated genes are well conserved with reading frames intact. We interpret this to mean that efagin is core to the *E. faecalis* species, and its function is under selection in diverse ecosystems.

Efagins are unique to E. faecalis

To determine if efagin or anything analogous to it occurred elsewhere, we searched a dataset of 15,854 *Enterococcaceae* genomes, excluding those of *E. faecalis*. No significant identity to efagin (>90% coverage; E-value <10) was found outside of the *E. faecalis* species. A search of all draft prokaryotic genomes in NCBI[23], approximately 1.2 million submissions at the time of the search (Oct 2022), similarly did not identify significant identities outside of *E. faecalis*.

To detect an analogous but potentially independently derived element within the genomes of other enterococci or related taxa, we also systematically screened them for phage-related structures meeting two key criteria: **i)** presence of at least a partial phage, and **ii)** presence in at least 50% of strains of a species. Partial phages occurring within the set of 15,854 *Enterococcaceae* genomes were identified using the Genomad provirus detection tool,[24] which detects clusters enriched in hallmark viral genes. Candidate prophage protein annotations were confirmed using Phastest.[25] Across all surveyed *Enterococcus* species and closely related *Melissococcus plutonius*, *Tetragenococcus halophilus*, and *Vagococcus fluvialis*, the efagin gene cluster of *E. faecalis* was the only example fitting these criteria. Our

search did identify an unrelated incomplete *Melissococcus plutonius* prophage present in 89.5% of that species' genomes (17/19 representative genomes), which is the subject of ongoing study. That partial phage is of length 11.6 kb (PHASTest[25]) to 20kb (Genomad[24]) **(Tables S2** and **S3**); however, its structural characteristics diverge from those of efagin as the *M. plutonius* element is predicted to encode a capsid and scaffold proteins suggestive of a defective phage or gene transfer agent[26] (**Table S3**). In contrast, efagins are characterized by a complete absence of phage head related-genes (**Figure S1A and S1B)**.

Efagins possess limited similarity to Gram-positive and Gram-negative phage-related antimicrobial agents and are independently derived

Other phage-related elements lacking capsids have been identified in select bacterial species, including the R- and F-type pyocins of *Pseudomonas aeruginosa*,[11–13,15,16] diffocin in *Clostridium difficile,*[21,22] and monocin in *Listeria monocytogenes.*[17–20] Though efagin shares little to no sequence identity to these elements, the genetic organization is broadly similar with operons structured such that genes encoding regulatory functions, followed by tail-like structural genes, and baseplate-like and release factor encoding genes (**Figure S1)**. Of the tailocins, efagin and monocin are the most organizationally similar, likely as a result of sharing parallel ancestry with non-contractile phage tail-like elements of Gram-positive microbes.[18–20] At the amino acid sequence level, evidence of sequence identity only occurs between three efagin proteins and monocin counterparts, with 40% amino acid identity over 90% of length for *ef1276*, an apparent regulatory element; 21% over 59% of length for *ef1288*, and 38% over 49% of the length of *ef1292* which are likely structural and lysis components (**Table S4**).

Because efagin is core to all *E. faecalis*, we sought to determine whether the above tail-like factors are also core to the species in which they had been described, perhaps filling a similar ecological role. We therefore assessed the prevalence of pyocins, diffocin and monocin across the phylogenetic diversity of their species of origin. Nucleotide sequences of each element were blasted against representative Refseq complete genomes that spanned the diversity of *L. monocytogenes* [N=74], *P. aeruginosa* [N=343], and *C. difficile* [N=32]. Monocin was found intact in 82.4% of *L. monocytogenes* genomes (**Table S5**). Interestingly, the remaining 17.6% (n=13) showed a fractional residual of the monocin sequence that reflected identical deletions of most (66%) of the monocin encoding unit, including all 11 structural genes.[18–20] Regulatory and lysis genes, however, remained intact (**Table S6**), suggesting that they may govern other coordinately expressed activities that remain under selection in that species. This deletion is present across multiple phylogenetically distinct lineages of *Listeria,* indicating that it may be a specific programmed event that has occurred repeatedly (**Figure S2**). At the point of deletion, all such monocin-derived elements possess identical insertions of 12-nucleotides (TTACATGCCCAT), providing some insight into a specific deletion mechanism. This high-frequency site-specific deletion is consistent with environments where selection favors variability.

The distribution of pyocins across *P. aeruginosa* genomes showed even greater variability. R-type (contractile) pyocins were identified in 75.8% of genomes, with the entire prototype reference sequence detected in 21.3% of representative genomes (**Figure 1A** and **Table S7**), and a variant with divergent components of the tail fiber and its chaperone detected in 54.5% more (**Table S8**). F-type (non-contractile) pyocins were identified in 35.5% of genomes, with the reference type present in 19.5% of genomes (**Table S9**), and a large variant in genes related to target specificity identified in 16% more (**Table S10**). In *C. difficile*, diffocin (R-type) was detected in 87.5% of genomes. Of those, 31.2% encoded the full length of the reference diffocin (**Table S11**), while 56.3% exhibited a large sequence variation in genes involved in structural assembly or receptor-binding specificity (**Table S12**). In contrast to efagin which is highly conserved and intact in all strains for which high quality genome sequences are available, these tailocins appear variable and often occur in alternate hybrid forms.

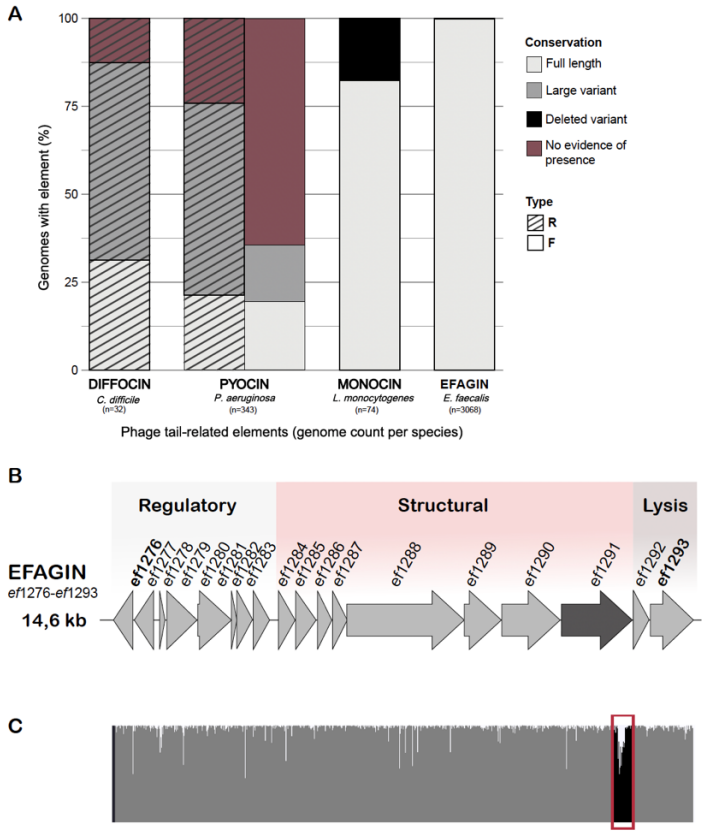


Figure 1. Efagin is core to *E. faecalis* and highly conserved except for a discrete variable region
(A) Distribution of contractile (R-type) and non-contractile (F-type) phage tail-like elements among genomes from Gram-positive and Gram-negative bacterial species.
(B) Genetic organization of the 14.6kb efagin gene cluster, and gene designations in prototype *E. faecalis* V583.
(C) Sequence conservation across the 14.6 kb efagin determinant in 2,774 *E. faecalis* genomes, aligned using MUSCLE.[27] The largest cluster of variability occurs in the C-terminal domain-encoding sequence of *ef1291* (highlight).

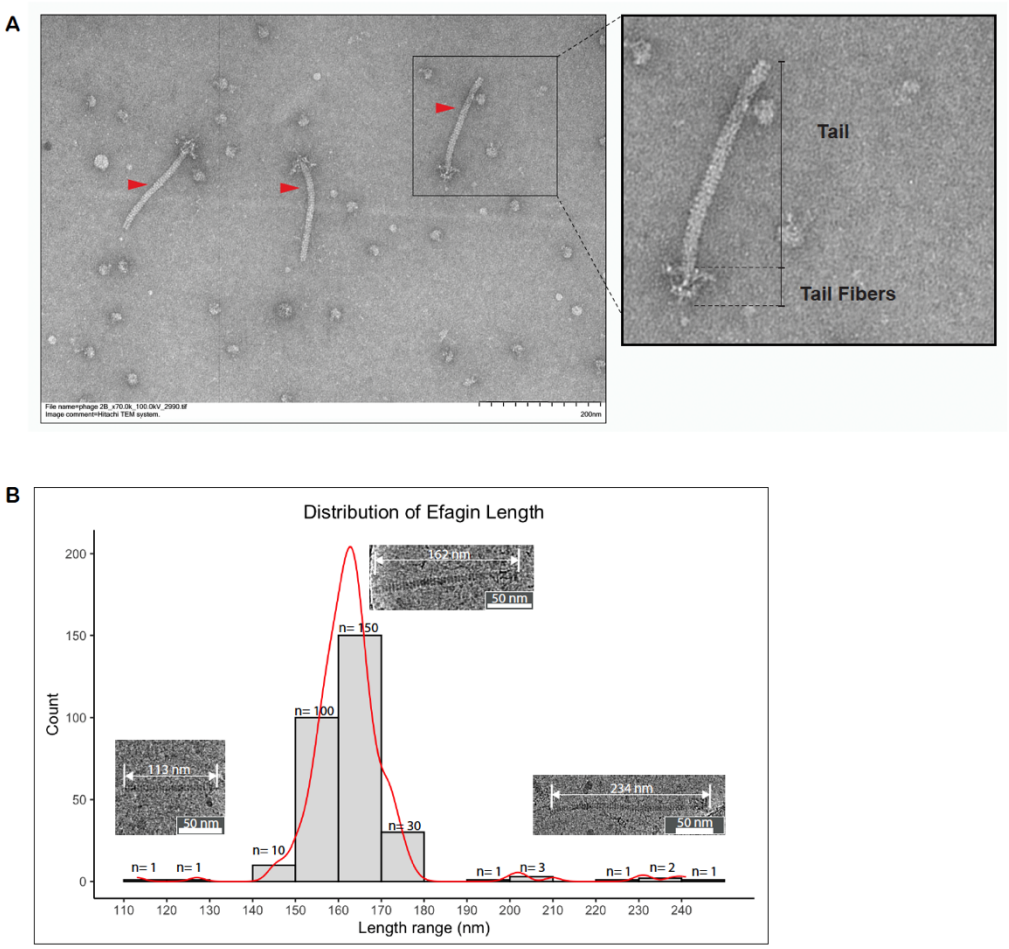
Efagin derived independently, sharing a common ancestor with a Siphoviridae phage
Because efagins exhibit limited identity to known bacterial phage tail-like elements, it was of interest to determine its closest extant relative. We previously noted that the gene cluster now identified as encoding efagin shared detectable sequence identity with Siphoviridae phages infecting microbes outside of the *Enterococcus* genus based on the limited sequence data available at the time.[28] When re-examined with current data, *E. faecalis* phage vB_EfaS_IME197 (41.3 kb)[29] now exhibits the greatest degree of identity. Six efagin proteins share 24.6% to 70.7% amino acid sequence identity with inferred gene products for this phage (**Table S13**). That level of similarity is far lower than sequence divergence within shared genes across the breadth of the *Enterococcus* genus,[4] indicating that efagin may not have derived directly from a common ancestor with that phage during the 400+ million timescale of *Enterococcus* existence.

Efagin structure and function
Efagin shows no evidence of mobility within the *E. faecalis* chromosome, as its genomic neighborhood is highly conserved (**Figure S3).** Further, the sequences of all efagin genes are highly conserved except for a localized region of variability near the 3' end of one gene (*ef1291*), suggesting selection for variation in the function encoded by that sequence (**Figure 1C**).

Previous work provided insights into the possible functionality of the efagin gene cluster. Matos and colleagues[8] systematically explored seven putative prophages moored within the prototype *E. faecalis* V583 chromosome, including the annotated partial phage described here as efagin. All phages – alone or

170 with a helper phage – produced viable particles capable of infecting new cells except for the efagin
171 element. Independently, as a byproduct of work assessing vesicle formation by *E. faecalis*, Afonina and
172 collegues[30] observed spurious particles that copurified with the vesicles of interest, which they attributed
173 to genes of the efagin cluster. Because **i)** efagin reading frames are well maintained, and **ii)** efagin
174 particles appeared to be expressed by the strain of *E. faecalis* OG1RF in studies of vesicle formation,[30]
175 we undertook a more exhaustive search to detect efagin activity. OG1RF, derived from a pediatric oral *E.*
176 *faecalis* isolate[31,32] and free of known bacteriocin-encoding genetic elements or phages[33] (aside from a
177 factor EntV known to be antagonistic for eukaryotic cells[34]) represented an ideal starting point for this
178 work.

179 We first reproduced the observations of Afonina et al.[30] by precipitating efagin using PEG and NaCl.
180 Transmission electron microscopy (TEM) revealed similar phage tail-like particles[30] absent in culture
181 supernatants derived from isogenic OG1RF*Δefg* in which 14,360 bp of the efagin gene cluster had been
182 deleted by allelic replacement in this work. Efagin particles were found to range from 113-234 nm, with an
183 average size of 163 ± 11.8 nm (mean ± SD), including tail and baseplate-like structures (**Figure 2A** and
184 **2B**).

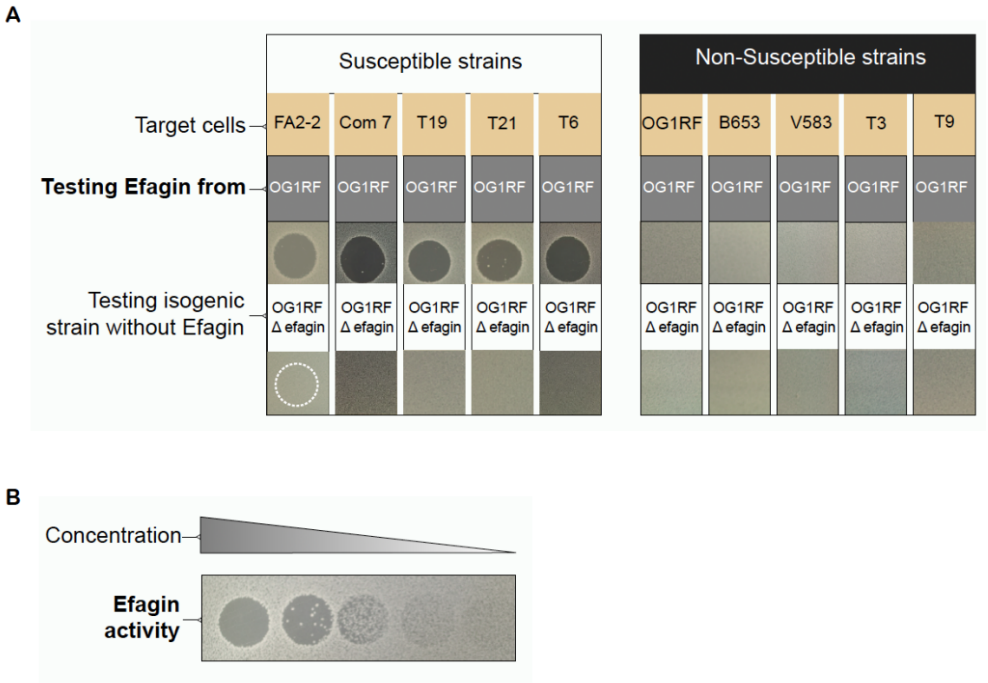


185

186 **Figure 2. TEM analysis of efagin particles**
187 (A) Partially purified OG1RF efagin (red arrow).
188 (B) Size distribution of 300 particles examined. Insets illustrate cryo-electron microscopy images of efagins across the
189 size spectrum.
190

191 To detect efagin activity, sterile filtrates of OG1RF or OG1RF*Δefg* were assayed for potential antagonistic
192 activity against a diverse panel of *E. faecalis* lineages.[28] Efagin-specific growth inhibition was observed
193 for some *E. faecalis* but not others (**Figure 3A**). Notably, efagin expressed by OG1RF did not kill the

194 producer itself (**Figure 3A**; right panel) but was active against select other *E. faecalis* strains (**Figure 3A**;
195 left panel). Furthermore, when serially diluted, efagin inhibitory activity was found to be lost evenly (rather
196 than forming fewer well-resolved individual plaques, as is typical of phages capable of self-replication and
197 reinfection). Instead, as for some antibiotics, breakthrough growth can be seen at lower concentrations
198 (**Figure 3B**). This indicates that the mechanism does not involve a phage-type infection and reinfection
199 cycle. Furthermore, even for the undiluted preparation, killing was confined to the area of the spot and
200 does not appear to be highly diffusible unlike most small molecule antibiotics.



201

202 **Figure 3. Efagin possesses select growth inhibitory activity**
203 (A) *E. faecalis* OG1RF efagin selectively targets *E. faecalis* strains (left panel) but not others (right panel).
204 (B) Efagin dilution does not result in discrete plaque formation.
205

206 ***Structural similarity between EF1291 and a known phage carbohydrate-recognition domain***
207 The observed region of sequence variation at the 3' end of *ef1291* (**Figure 1C**) was investigated further.
208 With limited inferred amino acid sequence identity to any protein of known function, *ef1291* was
209 annotated as encoding a hypothetical protein. We therefore used Phyre2[35] to detect possible conserved
210 higher order structural motifs (**Table S14)**. The greatest-confidence match was to the receptor-binding
211 domain of the Listeria phage PSA protein gp15, which encompassed 152 inferred C-terminal amino acids
212 of efagin with 96.7% confidence (representing 26% of the predicted efagin protein length). The crystal
213 structure of phage protein gp15 was previously solved, and the homologous domain was shown to bind
214 carbohydrates of the *L. monocytogenes* wall teichoic acid (WTA), thereby conferring target cell
215 specificity.[36] Because the amino acids of gp15 making contacts with *L. monocytogenes* WTA were
216 mapped,[36] we used AlphaFold to make a comparative efagin EF1291 model (**Figure S4).** Listeriophage
217 gp15 was solved in its trimeric form,[36] a quaternary structure common to other phage target binding
218 proteins,[37–39] so EF1291 was modeled as a trimer as well. This revealed substantial structural similarity
219 between them (**Figure S4)**, including amino acids of gp15 known to bind the wall carbohydrates of target
220 cells.[36] This provided initial evidence that EF1291 may function as a target cell receptor-binding protein,
221 possibly also targeting an analogous cell wall carbohydrate in enterococci, and that the observed pattern
222 of variability in EF1291 may correspond to variability in target cell specificity.
223
224 The rhamnose-rich cell wall polysaccharide (Rha-CWPS) in the cell wall of *E. faecalis* is referred to as
225 Enterococcal Polysaccharide Antigen (Epa). [40,41] This is known to be a variable trait[4]**.** Because of **i)** the

226 observed strain selectivity of OG1RF efagin**, ii)** structural similarities between efagin EF1291 and
227 listeriophage gp15, and **iii)** the observation that efagin does not appear to target the cell that produces it,
228 we hypothesized that efagin variability might correlate with Rha-CWPS/Epa variability. To detect a
229 relationship between variation in EF1291 sequence and Epa*,* we first tested for correlation between
230 *ef1291* single nucleotide polymorphisms (SNPs) and the presence/absence of specific Rha-CWPS/*epa*
231 operon genes in *E. faecalis* genomes that span the species diversity. We observed significant correlation
232 between specific *ef1291* SNP positions and the presence/absence of four *epa* genes in the *epa* gene
233 cluster variable region[42]: *ef*2166 (-0.729, p=1.13e-91), *ef*2167 (-0.729, p=1.13e-91), *ef*2168 (-0.723,
234 p=1.46e-89); and *ef*2169 (-0.698, p=6.05e-81) (**Figures S5** and **S6)**.

235 **Efagin types**
236 Next, we examined the sequence variation in the inferred C-terminal domain of EF1291 of 1,972 *E.*
237 *faecalis* genomes with high quality sequences. Nearly all variants fell into one of six sequence patterns,
238 termed efagin types A-E, and type X (**Figure 4** and **Table S15)**. Compared to arbitrarily chosen type A,
239 variations in types B, C and D clustered around specific amino acid positions with others being invariant.
240 However, the C-terminus of the EF1291 homolog of efagin type E is comparatively unrelated (19.1%
241 amino acid identity with type A) and appears to have arisen through a naturally occurring allelic
242 substitution. Although the EF1291 N-terminus of type E efagin shows extensive identity with that of types
243 A-D, the divergent C-terminal sequence shows identity outside of *Enterococcus* only to sequences from
244 the *Caudoviricetes* group (**Table S16**). The presence of this alternatively-derived C-terminal domain
245 provided specific evidence that it is likely to be independently functional. The sixth EF1291 sequence
246 variation, which we term type X, includes 49 subvariants of type C, two of type B, and two of type D, all
247 having an insertion, deletion, or substitution resulting in a premature stop codon. This is predicted to
248 result in a truncated EF1291 protein of 220 amino acids (*e.g.*, T21 strain) compared to 589 for type C
249 efagin encoded by OG1RF. We have yet to identify a susceptible target for type X.
250
251 When EF1291 amino acid variations are mapped onto the structural model derived from comparation with
252 listeriophage PSA gp15, including **i)** specific variants occurring in each efagin type and **ii)** amino acids
253 encoded by nucleotide positions that covary with the presence/absence of variable *epa* genes, they
254 cluster in regions corresponding to areas of gp15 known to determine target cell range. This includes one
255 amino acid (EF1291 amino acid 515 of type C efagin) which coincides precisely with a site on gp15,
256 S302, known to alter polysaccharide binding specificity[36] (**Figures 4A-F**).

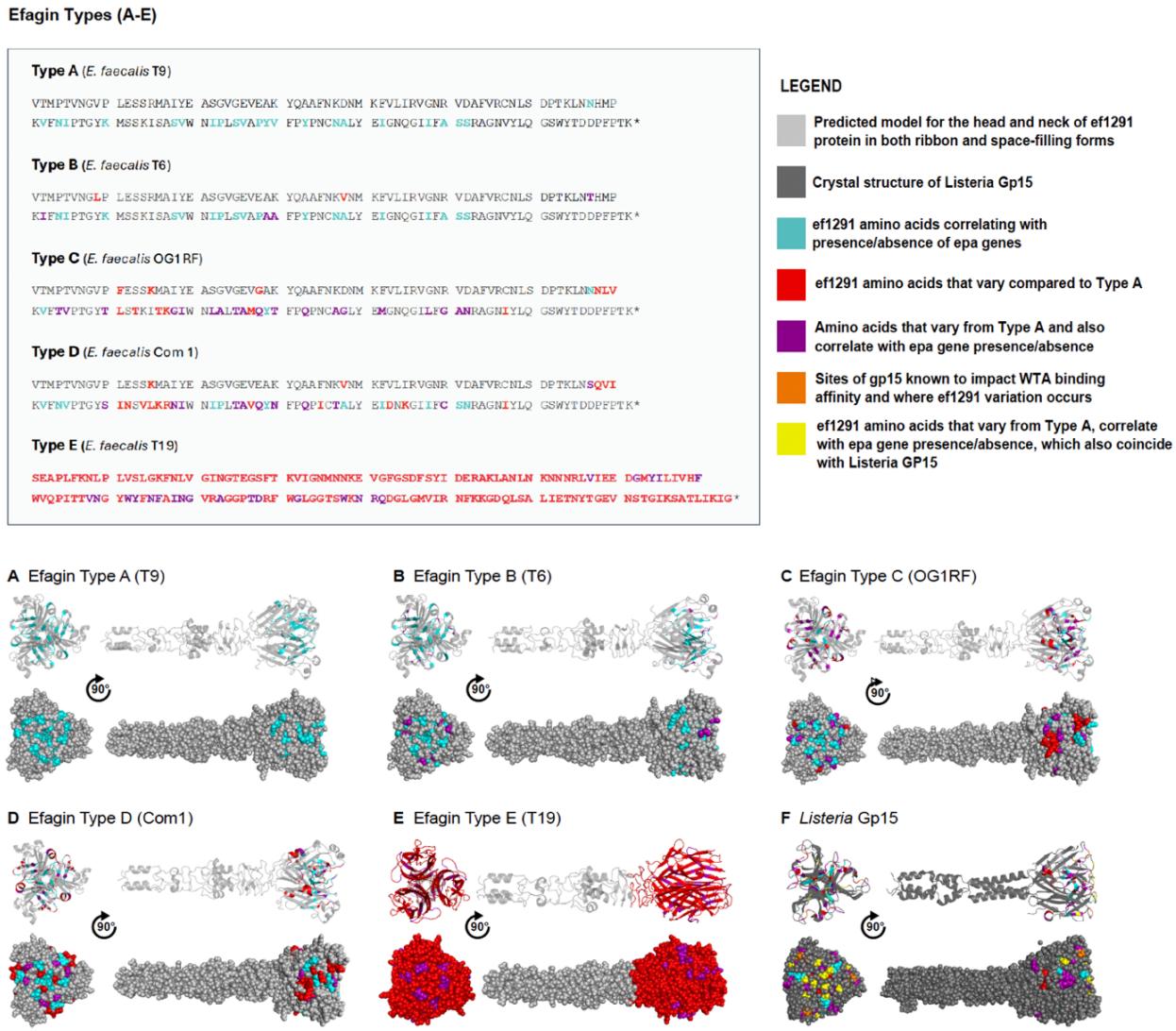


Figure 4. Structural similarities between EF1291 and listeriophage gp15 includes known carbohydrate-recognition domain

EF1291 C-terminal sequence variation patterns in efagin types A-E (amino acids 448 to 619). Amino acids that differ from efagin type A are indicated in red. Amino acids that correlate with the presence/absence of epa genes (correlation coefficient < -0.6) are colored cyan. Amino acids that vary in both variation and correlation analysis are colored purple.

(A-E) Structural models of efagin types, colored as in primary amino acid sequences above.

(F) Crystal structure of Listeria gp15. Amino acids known to impact WTA binding in Listeria gp15 are colored yellow.[36] Orange amino acids are those known to impact WTA binding in gp15 and also vary in EF1291. Additional positions homologous to those that vary in efagin, but for which data on target cell binding does not exist, are colored as in A-E.

To experimentally verify the efagin target, we selected and sequenced efagin-resistant mutants of susceptible strains. First, partially purified and concentrated type C efagin from OG1RF was spotted onto lawns of susceptible *E. faecalis* strains FA2-2[43] and X98.[44] These targets were chosen because both are susceptible to type C efagin but enigmatically were previously shown to be of different *epa* operon types.[42] Four FA2-2 mutants (FA2-2R1-R4) and two X98 mutants (X98R1 and -R2) were chosen for genome sequence analysis along with parental cultures as controls. All type C efagin resistant mutants of both lineages possessed mutations in *epa* operon genes (**Figure S7 and Table S17**), providing direct experimental evidence that despite differences in *epa* genotype, Epa is necessary for efagin activity.

Given the proliferation of genome sequence data since our prior analysis,[42] we re-examined the diversity of *epa* operons in the species *E. faecalis* by comparing 1,972 high quality *E. faecalis* genomes constituting a 100-fold increase in scope. The *E. faecalis epa* cluster consists of a conserved core of 18 genes (*epa*A–*epa*R), followed by a downstream region of variation (which includes *ef2164* to *ef2198* in prototype VRE *E. faecalis* strain V583).[41,42] As before[42], using conserved non-*epa* anchor genes flanking the operon, we located the Epa-encoding region of each genome, which on average spans 30+/-1.4 genes (**Table S19)**. With core genes largely conserved, operons were clustered based on shared homologs in the variable region, identifying 20 *epa* genotypes (**Figure 5** and **Table S15**). All four genes whose presence or absence correlated with sequence variations in *ef1291* (*ef2166–ef2169*) occurred in the *epa* type-determining variable region (**Figure 5**).

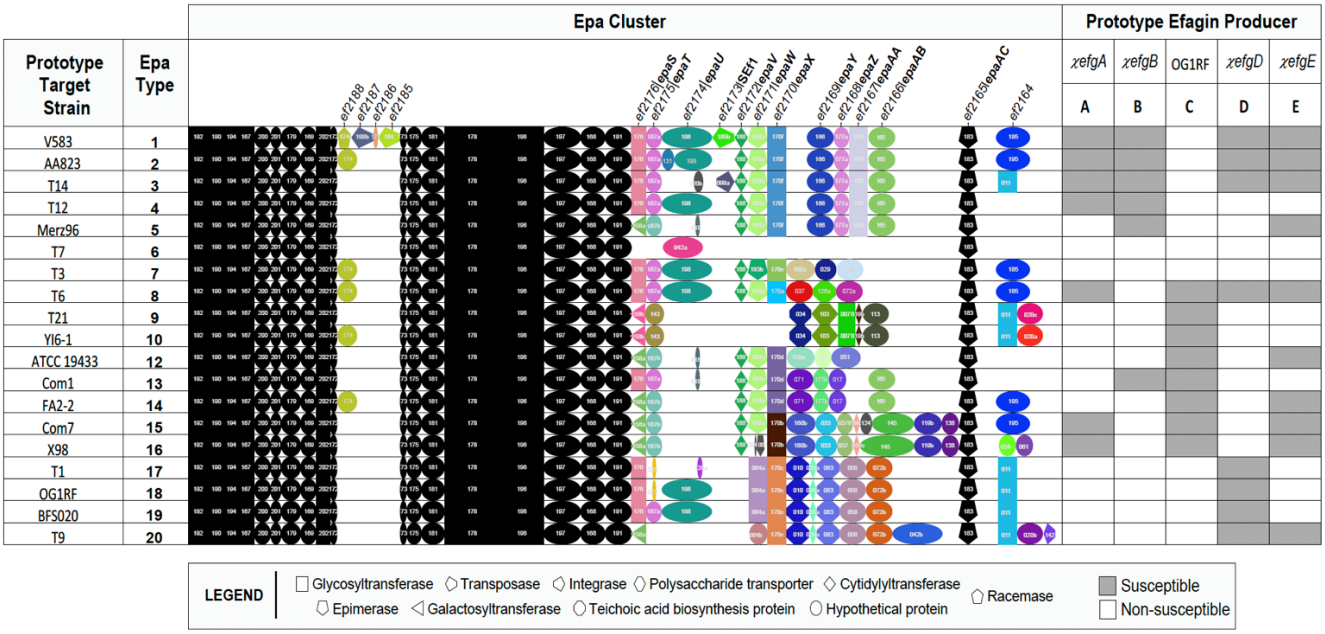


Figure 5. Epa genotypes and their susceptibility to being targeted by efagin types
Representative operons for 19 of the 20 identified *epa* genotypes found among the strains available to us are shown schematically. Conserved core epa genes appear in black, while variable genes are color-coded according to orthogroup membership (numbers shown inside each gene), with shapes corresponding to predicted function. Additionally, the susceptibility to the five types of efagin (A-E) is indicated for prototype strains harboring each epa genotype. The right-hand heat maps display efagin activity assays from five different donor strain: OG1RF (type C efagin) and OG1RF chimeras: *χefgA* (type A), *χefgB* (type B), *χefgD* (type D), and *χefgE* (type E). Chimeric constructs are denoted using the Greek letter "*χ*" (chi), and efagin is abbreviated as "*efg*."

The C-terminal domain determines target cell specificity

To prove that variation in the C-terminal domain of EF1291 determines target cell specificity, chimeras possessing C-terminal domain swaps from type A, B, D, and E donors were constructed on the efagin type C background of OG1RF by allelic replacement.[45] This strategy additionally allowed us to assess efagin activity free of other potential antagonistic activities or phages often present and variable in wild-type natural producers. Generation of an efagin type D EF1291 C-terminal chimera proved initially difficult. We reasoned that efagin type D may target OG1RF, making such a construct suicidal and direct assessment of the supernatant of the type D C-terminus donor strain, Com1, supported this prospect (although other potentially Com1-expressed inhibitory activities were not controlled for). Nevertheless, we concentrated a Com1 filtrate to select for resistant OG1RF mutants to serve as the host for constructing this chimera. Two Com1-inhibition resistant colonies of OG1RF were picked for genome sequence analysis, OG1RF-R1 and OG1RF-R2. Consistent with expectation, each possessed a different mutation in the OG1RF *epa* operon variable region (**Figure S7** and **Table S17**). Strain OG1RF-R2 was arbitrarily selected as the host for construction and expression of the type D chimera, which was obtained without

further difficulty. Once obtained, the type D chimera was verified as targeting OG1RF, but not OG1RF-R1 and -R2 mutants. As an additional proof of principle, the type A chimera was used to select for resistant mutants of its prototype strain target, V583. Again, two mutants were obtained and sequenced, V583-R1 and -R2, and again both also possessed mutations in variable region genes of its *epa* operon (**Figure S7** and **Table S17**). Each chimeric construction was verified by sequencing. Construction of EF1291 chimeras did not detectably affect efagin assembly or structure (*e.g.*, OG1RF*χefgD* as visualized by TEM (**Figure S8**).

Each efagin chimera was systematically examined for altered target cell specificity using available representatives of 19 of the 20 *epa* genotypes identified (**Figure 5**). We were unable to obtain a representative of *epa* genotype 11 for testing (**Figure S9)**. Notably, many *epa* types align with serovars described by Maekawa and collegues[46] without knowledge of *epa* variation. Chimeric efagins of types A, B, D, and E exhibit target cell specificity that differ from native OG1RF type C, showing both functionality and altered specificity. Efagins show overlapping activity patterns, but each type displays unique activities. This indicates that each efagin targets a unique aspect of the variable and complex Rha-CWPS/Epa structure.[47,48]

We never detected a population collapse associated with efagin production by an *E. faecalis* strain, indicating that its ecological role is not likely fratricide.[49] Further, direct testing showed that chimeric efagins do not kill the strains from which the corresponding C-terminal encoding sequence derived (**Figure 5**). An expanded analysis of 54 *E. faecalis* strains representing all efagin types, and 19 out of 20 *epa* types provided further support for a model where efagins target Epa types heterologous to that of the producer (Fisher's exact test $p < 0.00001$) (**Figure S10** and **Table S18)**. However, exceptions were observed, both in unexpected susceptibility as well as resistance (**Figure S10**). For example, strain SS7 (*epa* type 9) was found to be resistant to killing by efagin type C, whereas other members of *epa* type 9 were susceptible. Close examination of the SS7 genome shows that it harbors an additional gene within the *epa* operon (**Figure S10**, orthogroup 151; pink), potentially modifying its Epa structure. Other exceptions without obvious gene gains or losses (*e.g.*, strain Fly2, *epa* operon type 1 but resistant to efagin type D) likely have more subtle changes such as point mutations that affect expression of specific genes within their *epa* operons, similar to those selected for in deriving resistant mutants, but this remains to be shown.

Gene clusters encoding efagin (in V583 *ef*1276 – *ef*1293) and Epa (*ef*2198 – *ef*2164) are separated by approximately 1/3rd of the *E. faecalis* genome, so are not physically linked. We examined patterns of co-occurrence of efagin and *epa* types in 493 E. *faecalis* genomes that represent all major phylogenetic branches across the species (**Figure S11).** The results reveal a non-random pattern, with certain pairings occurring at high frequency (*e.g.*, efagin type B and *epa* type 8; or efagin type C and *epa* type 1), while others appear less common but still permissible. As these are scattered across the whole genome phylogeny (**Figure S6** and **S12)**, these co-occurrences support a functional linkage and possible co-evolution between efagin and *epa*, likely shaped by selective pressures related to self-immunity.

Efagins and phages select for cross-resistance

Chatterjee and colleagues[50] observed that mutants of *E. faecalis* strain X98 gained immunity to infection to some phages after acquiring mutations or deletions in *epa* genes. We obtained two mutant strains (51RS1 and 51RS2) from that study[50]. X98 is targeted by efagin types A, C and E (**Figure S7**), but we observed that phage-resistant X98 mutant 51RS1, which carries a deletion in the *epa* gene EFOG_0058,[50] retains susceptibility to efagin types A and E, but is no longer susceptible to type C. In contrast, phage-resistant X98 mutant 51RS2, which harbors a missense mutation in a different *epa* gene (EFOG_0052),[50] was found to be resistant to all efagin types (**Figure S7)**. This shows that phage selection likely contributes to variation in efagin susceptibility in an *epa* genetic background otherwise predicted to be incompatible, and conversely, that efagin selection likely affects phage susceptibility patterns.

Efagin target range extends beyond E. faecalis to other enterococcal species

We previously identified structural similarities between the *epa* operons of *E. faecalis*, *E. faecium*[42] and related operons in other *Enterococcus* species.[51] It therefore was of interest to determine whether efagin inhibitory activity extends beyond the species *E. faecalis*. As our previous analysis of *epa* operons of *E.*

363 *faecium* was done when few genomes were available,[42] we reexamined *epa* operons in a collection of 58
364 *E. faecium* genomes that span the species diversity[52] and assessed each for efagin susceptibility. Similar
365 to *E. faecalis*, *E. faecium epa* operons possess a conserved and then variable region with substantial
366 genetic diversity. The entire Epa rhamnose backbone (*epa*A–R) is core to *E. faecalis*, while a distinct but
367 conserved set of genes is core to *E. faecium (***Table S20***)*. Sequence identity among Epa-like proteins
368 encoded by *E. faecalis* V583 and *E. faecium* Aus0004 orthologs ranges from 31.2% (*epa*Q) to 92.1%
369 (*epa*E), whereas the variable decoration genes are entirely different, reflecting extensive divergence in
370 their evolutionary origin (**Figure S13** and **Table S21**). Of those 58 diverse strains, 35 were found to be
371 susceptible to at least one efagin type (**Figure 6**). Interestingly, 7 out of 18 VRE *E. faecium* strains (39%)
372 are targeted by an *E. faecalis* efagin, whereas 28 out of 40 (70%) VSE *E. faecium* are susceptible. Only 1
373 of 16 strains (6%) of the highly successful and often VRE clonal complex 17 (CC17) lineage is targeted
374 by an efagin. As both *E. faecalis* and *E. faecium* species co-occur in the human gut,[53] this raises the
375 prospect that evolving surfaces resistant to efagin killing may have contributed to the ecological success
376 of many members of the CC17 VRE *E. faecium* lineage.

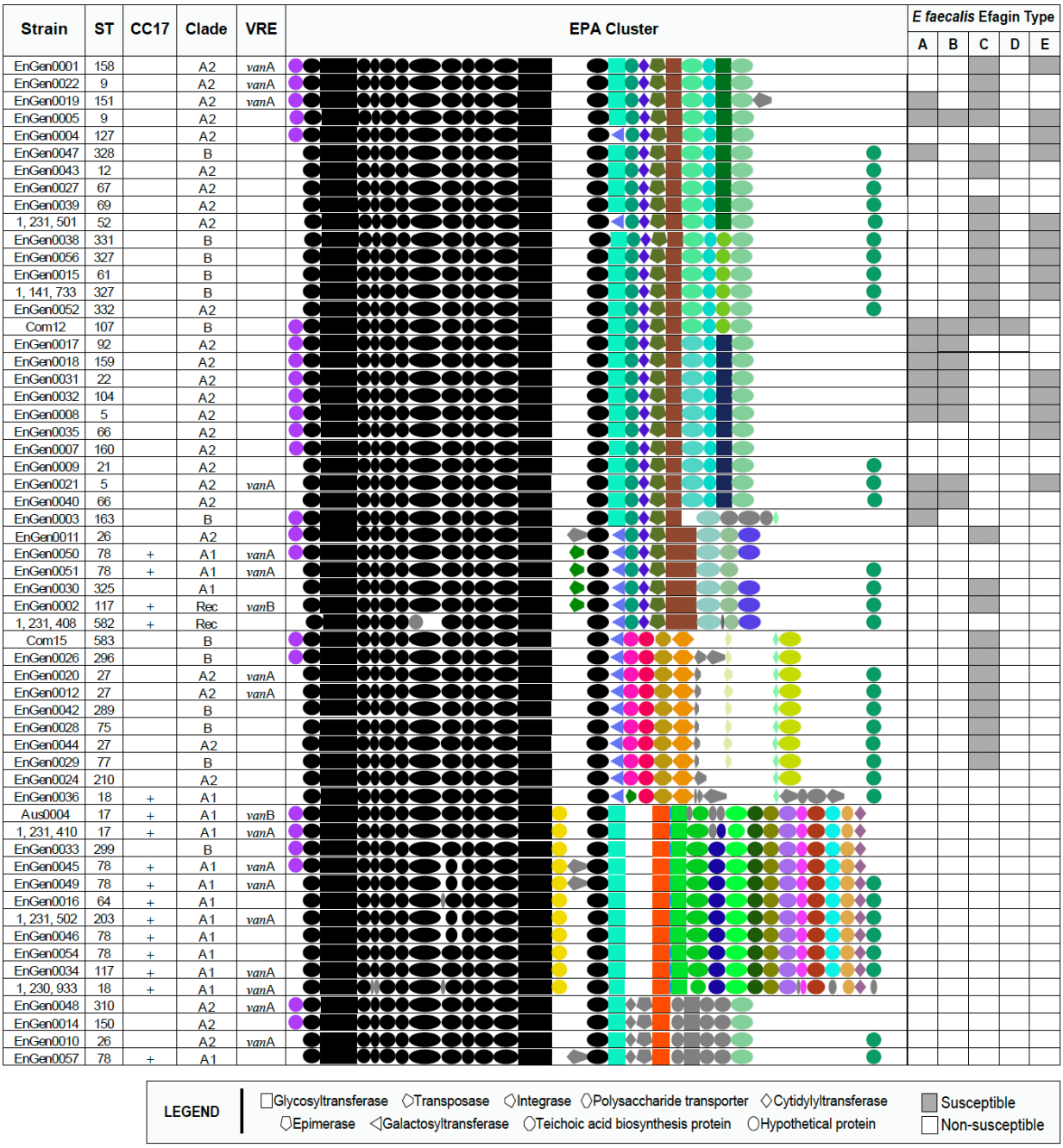


Figure 6. ***Structural diversity of epa operons across E. faecium strains and correlation with efagin susceptibility***
Strain identifier, sequence type (ST), clonal complex (CC), *E. faecium* clade assignments (A1, A2, or B[52]) and vancomycin resistance genotype in relation to the organization of *epa* operons. Rec indicates strains that are hybrids between clades. Conserved core genes are shown in black, while variable genes are colored according to orthogroup membership, with shapes corresponding to predicted function. The rightmost panel shows the efagin susceptibility profile. In the Epa cluster, black represents core genes found in all the genomes, and grey indicates rare genes found in less than 5%.

In addition to *E. faecium*, 40 other *Enterococcus* species spanning all four phylogenetic clades[4] were evaluated for efagin susceptibility (**Figure 7** and **Table S22**). Efagins were found to be effective against many species within clades I, II, and III, especially species possessing *epa* genes that are closely related to those of *E. faecalis*. In contrast, many clade IV species lack close homologs to genes for synthesis of

the core rhamnopolysaccharide stem of Epa,[47,51] and their Rha-CWPS/epa operon counterparts have received little study. Interestingly, *E. testudinis* was found to be susceptible to efagin despite lacking detectable sequence identity to core or variable *epa* genes of *E. faecalis* V583, suggesting the presence of a distinct cell wall polysaccharide structure that nevertheless possesses terminal decorations, or perhaps other cell wall structures, that serve as a binding target for efagin type C.

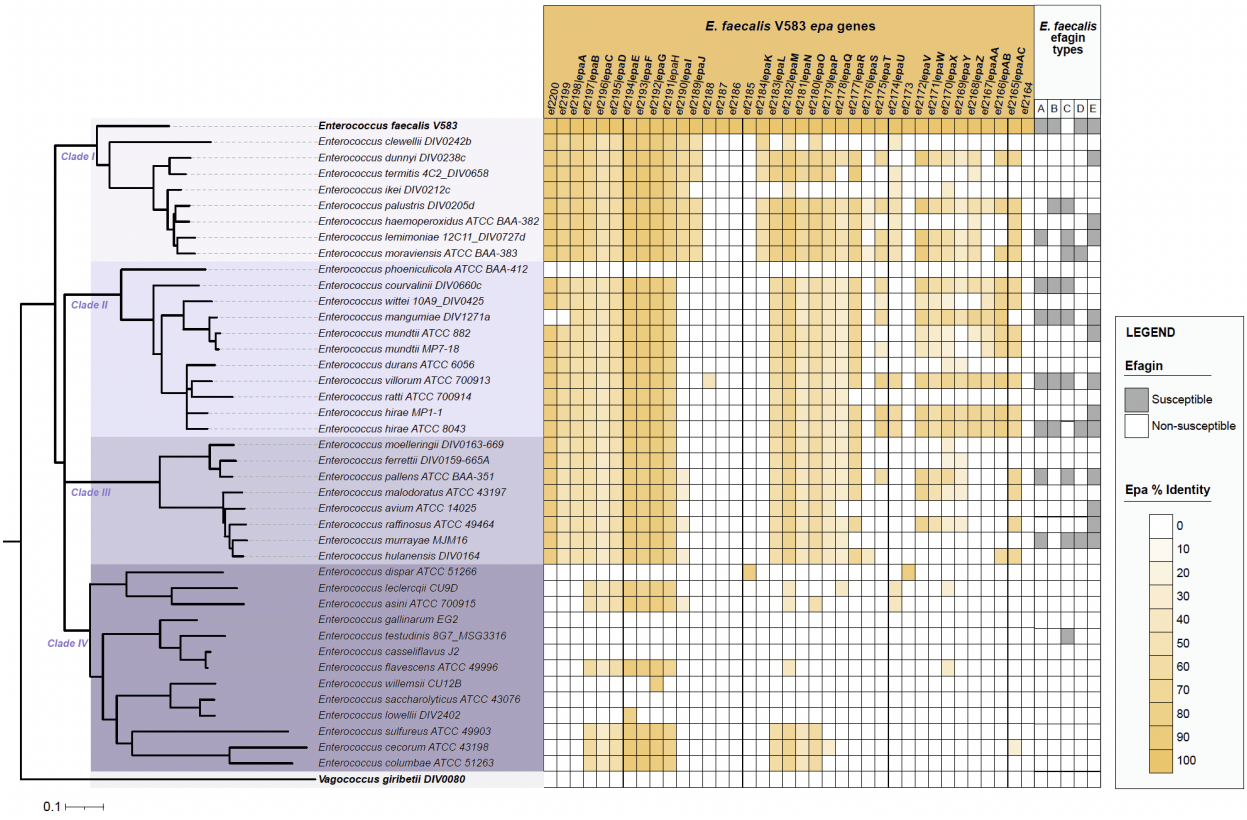


Figure 7. *Conservation of epa operon genes across Enterococcus species and correlation with efagin susceptibility*

Strains of *Enterococcus* clades are colored in graded shades of purple by membership in clades I – IV.[4] The extent of amino acid sequence identity with inferred *Epa* gene products, when compared to the prototypes of *E. faecalis* strain V583, is indicated by the gold heatmap. Efagin susceptibility is indicated with grey shading.

DISCUSSION

Here we identify efagins as a new class of independently evolved microbial inhibitors that are core to the biology of *E. faecalis*, the greatest generalist of the genus.[6] They exhibit target cell specificity that is determined by a phage tail fiber-like protein, and the binding target is the highly variable complex rhamnopolysaccharide homolog of wall teichoic acids, Epa, which occurs in the cell walls of most but not all enterococcal species. Whether functional targets occur on other non-rhamnose based carbohydrate structures on the surface of other types of microbes is currently unknown but appears likely based on the observed inhibition of *E. testudines,* which lacks homologs of the core genes for synthesis of the enterococcal rhamnopolysaccharide core. With activity against *E. faecium* strains, including some VRE clinical isolates, and the demonstrated ability to engineer specificity through discrete domain swaps, such agents hold promise for selectively controlling the gut ecology of hospitalized patients, limiting colonization by prevailing hospital adapted lineages of both VRE *E. faecium* and *E. faecalis*.

Aside from potential utility as therapeutics, the discovery of efagins and initial description of their characteristics also provides a new perspective on *E. faecalis* ecology. Shown to be the most commonly encountered species in a wide range of hosts, they are clearly highly effective generalists in the guts of

most if not all land animals,[6] as well as some marine species.[54,55] As enterococci are highly persistent when deposited in the environment,[56] animals throughout the food chain are exposed to *E. faecalis*, and efagins appear well suited to facilitate entry of this species into their gut consortia, potentially displacing other enterococcal species specifically adapted to various hosts. Similarly, as animals are consumed by others, *E. faecalis* may then deploy efagins to establish themselves in new hosts. In addition to implications for strain transmission, *E. faecalis* are also known to be facile gene exchangers,[57] which positions them well to mobilize naturally occurring adaptive genes such as antibiotic resistances in the guts of detritivores in the forest soil and transmit them to other ecologies. In these new environments, selection can confer an advantage: for example, *E. faecalis* strains are abundant in the guts of agricultural animals, including scratch feeders that consume soil-dwelling invertebrates,[58] animals that are also exposed to antibiotics. This would be consistent with an ecological model where such properties place *E. faecalis* centrally, possibly uniquely among microbes that commonly also colonize the human gut, as conduits of ancient resistance genes widespread in the soil ecosystem.[59]

The discovery of efagins provides yet another apparently independently derived example of what is often viewed as the co-option and redeployment of a phage component for the selective advantage of a bacterium.[60] However, as for other phage component-like bacterial factors, the lack of sequenced intermediates between efagins and their closest extant homologs makes it difficult to substantiate a precise evolutionary pathway. The repeated finding of elements as (or more) closely related to phages as they are to each other, [10,19,20,22,61] in contrast to a parallel series of convergent evolutionary events, would also be consistent with an evolutionary model where tail-like (or potentially also capsid-like) structures evolved as independent functional units and circulated as modules adapted independently for various cellular and viral uses. [61]

Aside from the variable C-terminal domain of the targeting protein encoded by *ef1291*, the high degree of conservation of efagin genes in all isolates of *E. faecalis,* and their uniqueness to that species, is striking. That they occur in all isolates of *E. faecalis* indicates that efagins were either gained at the time of the radiation of this species over 300 million years ago,[4] and transmitted vertically, or that this capability was later acquired and spread throughout the population in a sweep, but only in the species *E. faecalis*. The invariant occurrence of efagins in the genomes of isolates from a wide range of geographies and ecologies, and the lack of occurrence outside of the species *E. faecalis*, argue strongly for ancient acquisition with vertical transmission.

Cross selection and coevolution of efagins, phages and their Epa targets has likely contributed to *E. faecalis* genetic diversity. Here we observed that evolutionary dynamics of the efagin and epa differ from much of the rest of the core genome. Phylogenetic trees built with both efagin *ef1291* variants and *epa* are discordant from the bacterial core phylogeny. This phylogenetic discordance suggests that *ef1291* and *epa* genetic diversity likely stems from horizontal gene transfer and recombination, and efagin-mediated selection may represent a new important driver of that gene exchange. Examination of large and diverse genome sets of *E. faecalis* and *E. faecium* reveal the extent of diversity in operons for biosynthesis of the cell wall Epa rhamnopolysaccharide. Moreover, efagin susceptibility patterns across varying *epa* operon genotypes are likely to mirror phage susceptibility patterns, substantially informing their development for therapeutic purposes as well.

Limitations of the study

As has been shown for tailocins,[13,16,19,20,22,60,62] we show that efagins are capable of inhibiting the growth of targeted strains under the conditions tested here. However, much remains to be learned about the actual ecological role of efagins. The precise mechanism of this action for efagins, as for most non-contractile F-type tailocins, is unclear[16,19,20]. If competitor killing or inhibition is the objective, the nature of the selection that favors a 160 nm tail polymer is also unclear, when shorter variations spontaneously occur, as we show here, and if killing involves activities mainly of the baseplate[20]. Much remains to be learned about the contributions of efagins to the biology of *E. faecalis*, and their contribution to its exceptional success – both as a commensals of a wide range of terrestrial animal guts and as a persistent, multidrug resistant hospital pathogens.

Irrespective of the mechanism by which efagins contribute to the unique biology of the species *E. faecalis*, they clearly possess the ability to be engineered to target cells expressing a cognate cell wall rhamnopolysaccharide and related structures, which may be of considerable value in selectively depopulating the gastrointestinal tracts of hospitalized patients and limiting the outgrowth and transmission of multidrug resistant enterococci, including VRE, or possibly other medical or agricultural applications. In an era when few effective treatment options remain for those with multidrug-resistant enterococcal infections, the identification of a new agent that targets these recalcitrant organisms represents an important advance.

ACKNOWLEDGMENTS

We thank Nicki E. Watson and the Harvard Center for Nanoscale Systems for technical assistance with TEM. This work was supported by NIH grant AI083214 (Harvard-wide Program on Antibiotic Resistance) and was funded in part with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under grant numbers U19AI110818 (Broad Institute) and F32AI179151 (J.T.S). Work in the Kaplan laboratory at UChicago was supported by the National Institute of General Medical Sciences (R35GM157116 to M.K.) and the Searle Scholars Program (to M.K.). We gratefully acknowledge the expert technical assistance and support provided by Ryan Whipple and Jose T. Saavedra. The authors also thank members of the Gilmore and Earl Labs for helpful discussions during the development of this project and preparation of the manuscript.

AUTHOR CONTRIBUTIONS

M.S.G. conceived the study. M.S.G. and J.P. designed the overall experimental work and A.M.E. and A.L.M. supported project development and provided guidance on both the computational analyses and the biological aspects. A.L.M., J.T.S., G.R., and J.P. performed all the bioinformatic analyses. N.K. conducted the initial detection of efagin activity, and the efagin deletion mutant in OG1RF was constructed by N.K. and S.S.M. J.P. optimized efagin production and purification, performed activity assays, and generated efagin-resistant mutants and efagin chimeras. A.C. and M.K. performed the efagin particle measurements. J.P., A.L.M., J.T.S., A.M.E., and M.S.G. contributed to data interpretation and scientific discussion. J.P. and M.S.G. wrote the original draft, and J.P., A.L.M., J.T.S., and M.S.G. revised and edited the manuscript. A.M.E. and M.S.G. acquired funding and supervised the project. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

M.S.G., A.M.E., J.P., A.L.M., G.R., and N.K. are listed as co-inventors on a patent application related to efagin – WO2025054631 [https://patentscope.wipo.int/search/en/WO2025054631][65]. All other authors declare no competing interests.

RESOURCE AVAILABILITY

All Illumina sequencing reads and assembled genomes have been deposited in GenBank under accession numbers BioProjects PRJNA313452/PRJNA663900,[63] and PRJNA1371623[64]. Strains and corresponding accession numbers are provided in **Tables S1 and S28**.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact (Michael S. Gilmore; michael_gilmore@meei.harvard.edu) upon request.

METHODS

Collection of bacterial strains

The bacterial collection used in the experimental work is described in **Tables S23, S24**, and **S25.**

516 *Enterococcus* species were propagated in Brain Heart Infusion (BHI) broth or on 1.5% BHI agar at 34 °C
517 or 37 °C, depending on temperature preferences for each species[6], while *Vagococcus giribetii* was
518 incubated at 34 °C. *Escherichia coli* was maintained on 2x Yeast Tryptone medium (2xYT) at 37 °C.
519
520 **Induction and recovery of efagin particles**
521 To obtain efagin particles, bacterial cultures were grown overnight in BHI broth and incubated at 37 °C for
522 18 h. For strains not expressing efagin constitutively, overnight cultures were diluted 1:100 in BHI broth
523 and incubated for 4 h, followed by induction with mitomycin C (final concentration 2.5 µg/ml) and
524 incubation at 37 °C overnight. The cell debris was removed by centrifugation at 11,000 × *g* for 15 minutes
525 at 4 °C. Then the supernatant was collected and filtered through a 0.2 um membrane (Pall Life Sciences).
526 Efagin particles collected in the supernatants were added along with 1 M NaCl (5.84 g/100 ml) and 10%
527 (w/vol) polyethylene glycol 8000 (PEG), and the mixture was incubated for 1h at 4 °C. The PEG-
528 precipitated material was then collected by centrifugation at 25,000 × *g,* 4 °C, for 1 h. The pellet was
529 resuspended in SM buffer (50mM Tris-Cl pH 7.5, 100 mM NaCl, 8mM $MgSO_4 \cdot 7H_2O$,) and another
530 addition of 2.5 M NaCl and 20% (w/v) polyethylene glycol PEG-8000 was performed for 30 min at room
531 temperature. Efagin particles were collected by centrifugation at 17,000 × g, 4 °C, for 1 h, and the pellets
532 were resuspended in SM buffer.

533 **Efagin purification and concentration**
534 Efagin particles were purified and concentrated using centrifugal filter units: Macrosep 100 K MWCO
535 advance centrifugal devices (20 ml; Pall Life Sciences), and Amicon Ultra 100K; (0.5 ml; Millipore)
536 according to the manufacturer's instructions. Briefly, efagin suspensions in SM buffer were loaded into
537 Macrosep 100 K and centrifuged at 4,000 × g for 1.5 h. The collection tubes were then emptied, refilled
538 with 10 mL SM buffer, and subjected to a second centrifugation step. Retained particles were
539 resuspended in 2 mL SM buffer and transferred to Amicon Ultra 100 K filters. These were centrifuged at
540 5,000 × g for 30 min, and efagin was recovered by reverse spinning at 1,000 × g for 2 min. The resulting
541 concentrated preparations were used for activity assays of efagin and for subsequent analyses.

542 **Transmission Electron Microscopic (TEM)**
543 Purified efagin particles were diluted in SM buffer to a concentration of 10^7 efagin-forming units/ml.
544 Freshly ionized carbon coated grids were floated on a 10 ul drop of sample for 1 minute. Grids were
545 washed with 5 drops of 2% uranyl acetate (UA), and the excess of UA was drawn off with grade 50
546 Whatman filter paper. Grids were allowed to dry and imaged with a Hitachi 7800 at 100KV.

547 **Grids preparation and imaging conditions for single particle analysis**
548 Purified efagin particles in a volume of 3 µl were applied onto the glow discharged Quantifoil R 1.2/1.3,
549 300 Mesh, Au grid and plunge frozen using the Vitrobot Mark 4 (Thermo Fisher Scientific, Waltham, MA,
550 USA) in 100% humidity and 4°C. The grids were blotted for 3s and automatically plunged into liquid
551 ethane. Vitrified samples were transferred to storage boxes and stored in liquid nitrogen until use. Grids
552 were clipped in AutoGrid Rings and loaded into a 300 kV Titan Krios electron microscope (Thermo Fisher
553 Scientific, Waltham, MA, USA) with Gatan K3 Camera and BioQuantum Energy Filter set to a 20 eV slit
554 width. Micrographs were collected with a nominal 81.000x magnification, with a corresponding pixel size
555 of 1.065Å, a total dose equal 60e-/Å, 50 frames and defocus range of -2.1 to -0.9 µm.

556 Micrographs were imported to the CryoSPARC v4.6 software,[67] followed by Patch Motion Correction and
557 Patch CTF Correction. Three hundred random efagin particles were manually measured, which allowed
558 us to prepare the length distribution plot using RStudio v2025.09.2 software.[68]

559 **Detection of phage activity and titration**
560 Efagin particles from OG1RF *E. faecalis* were tested on an indicator strain for plaque formation.[87] Briefly,
561 the indicator strain was cultured overnight and then diluted to equal the turbidity of a 0.5 McFarland
562 standard (~1.5 x 10^8 colony forming units (CFU)/ml). One hundred microliters of the bacterial suspension
563 were mixed with 1 ul of diluted efagin particles (prepared 10 to 10^9-fold serial dilutions) and 5 ml of 0.5%
564 BHI agar, which was then poured onto a 1% BHI agar plate. Once the overlay solidified, the plates were

565 air-dried and incubated overnight at 37 °C. Plaque formation was visually detected and enumerated by
566 plaque assay using JM110 *Escherichia coli* (indicator strain) and M13 bacteriophages as a positive
567 control.

568 **Efagin locus knockout in *E. faecalis* OG1RF**
569 An efagin knockout strain of *E. faecalis* OG1RF was generated through allelic replacement. Genomic
570 regions flanking the efagin locus (878 bp upstream and 860 bp downstream) were amplified by PCR
571 using primer pairs KO-F1_XbaI up-forward/KO-R1_Acc65I up-reverse and KO-F2_KpnI down-
572 forward/KO-R2_SbfI down-reverse (**Table S27**), then fused via overlap extension PCR to create a 1,738
573 bp deletion construct. This fragment was digested with *XbaI*/ *SbfI* and cloned into the temperature-
574 sensitive plasmid pLT06 pre-digested with the same enzymes. The recombinant plasmid was propagated
575 in *E. coli* EC1000 and verified by sequencing. Following PCR confirmation with primers OriF and
576 KS05seqR, the construct was introduced into *E. faecalis* OG1RF by electroporation. Allelic exchange was
577 carried out using temperature shift and p-chlorophenylalanine counterselection, as described previously.[45]
578 Candidate mutants were screened by PCR using internal primers (efagin-internal-F and efagin-internal-R;
579 **Table S27**), and successful deletion was confirmed by whole genome sequencing.

580 **Construction of efagin chimeras in *E. faecalis* OG1RF**
581 Efagin chimeras of *E. faecalis* OG1RF were generated by allelic exchange. A 1,324 bp DNA fragment
582 spanning the region between the C-terminal end of *ef*1291 and the start of the EF1292 efagin gene was
583 amplified by PCR using primer pair 817_F_BamHI up-forward/2150_R_PstI up-reverse (**Table S27**). The
584 resulting product was digested with BamHI and PstI and ligated into the temperature-sensitive plasmid
585 pLT06, pre-digested with the same enzymes. Recombinant plasmids were propagated in *E. coli* EC1000
586 in medium supplemented with chloramphenicol (15 µg/mL), verified by PCR using primers OriF and
587 KS05seqR, and confirmed by Sanger sequencing. Plasmids were subsequently extracted and introduced
588 into *E. faecalis* OG1RF by electroporation. Allelic exchange was carried out using temperature shift and
589 p-chlorophenylalanine counterselection, as previously described.[45] Candidate mutants were screened by
590 PCR using internal primers for efagin types A–E (internal-F and internal-R; **Table S27**) and further
591 evaluated by activity assays. Successful recombinants were validated by sequencing.

592 **Host range determination**
593 Efagin killing assays were conducted using spot tests and the double agar layer method.[88] The indicator
594 strains were cultured overnight and then diluted to equal the turbidity of a 0.5 McFarland standard (~1.5 x
595 10^8 CFU/ml). One hundred microliters of the bacterial suspension were applied to a 0.5% Brain Heart
596 Infusion (BHI) agar overlay, which was then poured onto a 1% BHI agar plate. Once the overlay solidified,
597 the plates were air-dried and incubated in the refrigerator for 1 hour. Following this, 4 µl of efagin
598 preparations were carefully spotted onto the plates and allowed to dry/adsorb and incubated at 4°C for
599 2h. After that, the plates were incubated at 37 °C overnight, and zones of clearing indicated positive
600 bactericidal activity.

601 **Isolation of efagin-resistant *E. faecalis* strains**
602 The efagin-resistant phenotypes were selected using spot tests and the double agar layer method as
603 described previously. Plates were let dry, and 4-20 µl of efagin particles were spotted and allowed to
604 dry/adsorb for 4 h. After, plates were incubated at 37 °C overnight. The presumptive efagin-resistant
605 colonies appeared in the zones of clearing. The efagin-resistant phenotypes were confirmed by spot
606 assays and whole genome sequencing (WGS).

607 **Genomic DNA preparation, high-Throughput sequencing, assembly, and annotation**
608 *Enterococcus* strains were cultured in BHI broth at 37 °C for 18 h. Genomic DNA was extracted using the
609 DNeasy Blood & Tissue Kit (QIAGEN, San Luis, MO, USA) with minor modifications to the manufacturer's
610 protocol. Specifically, 50 µL of lysozyme (50 mg/mL) and 10 µL of mutanolysin (2,500 U/mL; Sigma-
611 Aldrich, Germantown, MD, USA) were added and incubated for 30 min at 37 °C prior to treatment with 20
612 µL proteinase K (20 mg/mL). DNA concentrations were determined using the Qubit dsDNA High-
613 Sensitivity Assay Kit (Thermo Fisher Scientific, MA, USA). Sequencing libraries were prepared with the

614 Nextera XT DNA kit and Illumina index primers. Paired-end reads (2 × 250 bp) were generated on an
615 Illumina MiSeq platform with the MiSeq Reagent Kit v2. Reads were trimmed to remove adapters, quality-
616 filtered, and screened for PhiX contamination before de novo assembly in CLC Genomics Workbench
617 v8.0.3. Ten additional strains (efagin resistant-mutants and chimeras) were sequenced externally by CD
618 Genomics[89] and processed using the same pipeline.

619 All Illumina sequencing reads and assembled genomes have been deposited in GenBank under
620 BioProjects PRJNA313452 and PRJNA663900,[63] and PRJNA1371623[64]. Strains and corresponding
621 accession numbers are provided in **Table S1**.

622 **WGS variant calling analysis**
623 Whole-genome sequencing was performed on both efagin-resistant mutants and their corresponding
624 parental strains to identify single nucleotide polymorphisms (SNPs) and insertions/deletions (indels).
625 Sequencing reads were aligned to the finished reference genome using the BWA-MEM algorithm
626 (v0.7.4).[85] Variant calling was then carried out using Pilon **(**v1.23).[90] Variants with a mapping quality score
627 below 10 were excluded from the final dataset to ensure accuracy.

628 **Searching for efagin sequence in Enterococcaceae and other prokaryotic bacteria**
629 We downloaded all 16,907 Enterococcaceae assemblies (Taxid 18,015) from NCBI in May 2021, to
630 analyze together with additional Enterococcaceae genomes from the Broad Institute's collection (n=
631 1108). Quality controls were performed to remove contaminated or incomplete genomes, using
632 Phylosift[69] and CheckM.[70] FastANI[71] was used to identify species in our dataset by comparing to
633 reference genomes, including *E. faecalis* V583, with the presence of >95% average nucleotide identity to
634 a known reference genome confirming the species designation. Our final dataset contained 15,854
635 confirmed Enterococcaceae assemblies for analysis, including 3,068 *E. faecalis*.

636 BLAST[23] was used to search our enterococcal assembly database for hits matching the efagin sequence
637 from *E. faecalis* reference strain V583. We calculated the total coverage of efagin for hits within each
638 target sequence, accounting for overlapping hits. Efagin was defined as present in a strain if there were
639 Blast hits with at least 90% coverage of efagin (E-value <10), including at least one hit of length > 2000
640 bp. Another, more comprehensive BLAST search was performed against all draft prokaryotic genomes
641 available on NCBI[23]. All sequencing submissions matching the query "wgs_master[prop] AND
642 (bacteria[orgn] OR archaea[orgn]) NOT refseq[filter]" on the NCBI Nucleotide website were included in
643 this dataset[23], or approximately 1.2 million sequencing submissions in Oct 2022.

644 **Selection of representative *E. faecalis, L. monocytogenes, P. aeruginosa and C. difficile* genomes**
645 **to cover the breadth of diversity within the species**
646 In order to remove biases within our dataset due to over-representation of commonly sequenced clinical
647 subtypes, and to ensure computational tractability, we selected representative genomes for each species,
648 chosen to represent the diversity present across the species. Briefly, all genomes were clustered using
649 the StrainGR (v1.3) software package[91], using its utility for creating representative reference genome
650 databases. We clustered genomes with a jaccard distance of 0.9, and a representative was chosen for
651 each cluster.

652 For *E. faecalis*, we started with our full set of 3,068 genomes. After clustering, we chose the
653 representatives with the smallest number of contigs, that also contained a full efagin on a single contig,
654 resulting in a set of 441 *E. faecalis* genomes. Additionally, we added lab-tested strains important to this
655 study, with a final total of 493 genomes representing the diversity of *E. faecalis*.

656 For analyses of *Listeria monocytogenes*, *Pseudomonas aeruginosa*, and *Clostridium difficile*, we started
657 with 414, 883 and 189 genomes from the complete refseq database, respectively. After clustering, the
658 resulting datasets contained 74, 343, and 32 genomes, respectively, highlighting different levels of
659 diversity for each species in the database.

660 **Searching for partial phages**

661 To detect partial phages from additional enterococcal species and closely related *Melissococcus*
662 *plutonius*, *Tetragenococcus halophilus*, and *Vagococcus fluvialis*, we used all isolates for each species
663 from the refined dataset of 15,854 Enterococcaceae assemblies described above. These were then used
664 as input into the provirus identification tool within the geNomad software package.[24] This tool identifies
665 and extracts proviruses from assembled genomes by detecting regions enriched in viral hallmark genes
666 using genomad's neural network classification model. We filtered for proviruses present in the majority of
667 isolates in each species and explored whether they lack typical prophage proteins using geNomad[24].
668 Incomplete annotations were confirmed with Phastest.[25]

669 **Identification of the closest known ancestor of efagin and protein identity**
670 All V583 efagin gene sequences were analyzed using the phage prediction tool PHASTER[25] to identify
671 the closest known phage ancestor. The resulting phage protein sequences were then compared with
672 efagin proteins using BLAST alignments (default E-value thresholds) to determine protein-level identity.

673 **Monocin similarity analysis**
674 Using nucleotide sequences from all V583 efagin genes, we searched against the monocin gene
675 sequences and obtained no hits (default E-value thresholds). Using sequences from all V583 efagin
676 proteins, we searched against the monocin protein sequences from ATCC 35152 strain using a BLAST e-
677 value threshold of 1e-5.

678 **Functional annotation of efagin and genes in its genomic neighborhood**
679 Gene function information for efagin (*ef1276* to *ef1293*) and its surrounding region (20 genes upstream
680 and downstream) was obtained by combining information from Prokka,[72] Pfam,[73] Kegg,[74] Phyre 2.2,[75]
681 BioCyc,[76] Alphafold 3,[77] and PaperBlast.[78] The annotated genes immediately upstream and downstream
682 of *efagin* via this pipeline are not present in the *E. faecalis* V583 genome annotation in GenBank[23] but
683 were found in our Prokka[72] annotations. For reference in our analysis, we assigned them arbitrary *ef* IDs:
684 *ef3441* (upstream) and *ef3442* (downstream).

685 **Orthogroup clustering and characterization of *epa***
686 To obtain efagin particles, bacterial cultures were grown. Among all isolates containing a complete efagin
687 sequence, 1,972 also had a complete epa operon on one contig. Since *epa* consists of a conserved
688 region followed by a variable region (containing 7 to 27 genes), a conserved gene downstream of *epa*,
689 *ef2162*, was used to define the downstream boundary of the operon. The genes between *epa*A and
690 *ef2162*, not including *ef2162*, were considered part of the *epa* operon and were included in our
691 orthogroup clustering analysis. SynerClust[92] was run on these genomes, classifying all *epa* genes into
692 137 orthogroups. To further correct for possible overclustering artifacts, we split orthogroups with within-
693 cluster average nucleotide identity <90%.

694 To explore the presence of *epa* in *E. faecium*, we downloaded 58 diverse *E. faecium* genomes from a
695 previously published study[52] and repeated the same orthogroup clustering process described above. In
696 total, 95 epa orthogroups were present in *E. faecium*.

697 In order to understand epa diversity across the *Enterococcus* genus, we used previously published
698 results[51] that used BLAST to identify which epa genes are present in other enterococcal species,
699 calculating the percent identity between genes from different enterococci and the *E. faecalis* V583 epa
700 genes. For isolates not included in that study,[51] we also used the same method to detect epa genes,
701 using the same E-value threshold of 1E-10 for inclusion as a positive hit.

702 **Analysis of correlation between efagin SNPs and *epa* genes**
703 For each gene in the variable region of the epa, we created a vector with a binary value corresponding to
704 the presence or absence of this gene within each *E. faecalis* representative genome selected as
705 described above. Correspondingly, for each 150bp region within efagin, we created a vector containing
706 the number of SNPs in that region for each representative genome within our database. A position in a
707 genome was considered a SNP if it differed from the nucleotide at that position in V583 or if it was an

inserted or deleted relative to V583. Pearson correlations were calculated between the vectors for each epa gene and each efagin region.

Protein structural modeling

Alphafold 3[77] was used to model the full-length protein structures of all five efagin *ef*1291 types via alphafoldserver.com. As gp15 is known to form a homotrimeric structure, we included three copies of *ef*1291 as input. We generated visualizations using Pymol.[79]

Defining Epa and efagin types

We defined epa types using hierarchical clustering, based on the presence or absence of *epa* genes in our orthogroup clustering based on 1,972 *E. faecalis* genomes. This was performed using the AgglomerativeClustering package from scikit-learn.[81] We chose an optimal threshold by comparing the clusters with patterns of efagin killing activity, achieving optimal agreement with the distance_threshold parameter value of 9. For the 58 *E. faecium* genomes, we applied the same approach, using a distance_threshold parameter value of 6.

Efagin types were defined for the complete dataset using hierarchical clustering of an alignment of amino acid residues representing the variable region of *ef*1291 (V583 efagin coordinates 12,539-12,967). We used the AgglomerativeClustering package,[81] and an optimal threshold was chosen by comparing the clusters with patterns of efagin killing activity, achieving optimal agreement with a distance_threshold value of 24.

Construction of phylogenies

The whole-genome phylogeny of *E. faecali*s representatives was generated with Parsnp v1.1.2[82] using default parameters and V583 as the reference genome. To construct the single-copy core phylogenetic tree of Enterobacteriaceae, individual representative genomes for each species were used as input to Orthofinder[83] to find all genes present in >=95% of isolates. These were aligned with MAFFT[84] and then phylogenetically reconstructed using IQTree[86] in modelfinder mode with 1000 ultrafast bootstraps.

SUPPLEMENTAL INFORMATION

Supplemental Files (pdf): Figures S1–S14, Tables S18, S26, S27, and S28, and Supplemental References

Supplemental Information:

Table S1. BLAST results of efagin against 3,068 high-quality *E. faecalis* genomes
Table S2. Summary of Genomad-identified phage-derived elements in *Melissococcus plutonius*
Table S3. Summary of Phastest-identified phage-derived elements in *Melissococcus plutonius*
Table S4. Results of BLAST analysis comparing efagin and monocin
Table S5. Summary of monocin loci identified in 74 *Listeria monocytogenes* genomes
Table S6. Monocin deletion identified across 13 *Listeria monocytogenes* genomes
Table S7. Summary of R-type pyocin loci detected across 343 *Pseudomonas aeruginosa* genomes
Table S8. R-pyocin variant region identified across 187 *Pseudomonas aeruginosa* genomes
Table S9. Summary of F-type pyocin loci detected across 343 *Pseudomonas aeruginosa* genomes
Table S10. F-pyocin variant region identified across 55 *Pseudomonas aeruginosa* genomes
Table S11. Summary of R-type diffocin loci detected across 32 *Clostridium difficile* genomes
Table S12. R-diffocin variant region identified across 18 *Clostridium difficile* genomes
Table S13. BLAST results for the *E. faecalis* phage most closely related to the efagin
Table S14. V583 efagin gene function prediction performed by Prokka, Pfam, and KEGG
Table S15. List of efagin and epa types assigned to 1,972 *E. faecalis* genomes included in the study
Table S16. BLAST results of type E efagin C-terminal end
Table S17. Whole-genome sequencing data of spontaneous efagin-resistant *E. faecalis* mutants
Table S19. EPA orthogroup dataset of 1,972 *E. faecalis* genomes
Table S20. EPA orthogroup dataset of 58 *E. faecium* genomes

Table S21. Amino acid identity between epa core genes of *E. faecalis* V583 and *E. faecium* Aus0004 based on BLASTP alignments
Table S22. Amino acid sequence identity (%) between genes from the epa cluster of *E. faecalis* V583, and their top hit in 40 Enterococcus species
Table S23. List of *E. faecalis* strains used in the experimental work
Table S24. List of *E. faecium* strains used in the experimental work
Table S25. List of *Enterococcus* species used in the experimental work

REFERENCES

1. Centers for Disease Control and Prevention (CDC). Antibiotic Resistance Threats in the United States. https://www.cdc.gov/antimicrobial-resistance/data-research/threats/index.html.

2. Ikuta, K.S., Swetschinski, L.R., Robles Aguilar, G., Sharara, F., Mestrovic, T., Gray, A.P., Davis Weaver, N., Wool, E.E., Han, C., Gershberg Hayoon, A., et al. (2022). Global mortality associated with 33 bacterial pathogens in 2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet *400*, 2221–2248. https://doi.org/10.1016/S0140-6736(22)02185-7.

3. European Centre for Disease Prevention and Control - ECDC (2024). Antimicrobial resistance in the EU/EEA (EARS-Net)—Annual epidemiological report 2023. November 18, 2024. https://www.ecdc.europa.eu/en/publications-data/antimicrobial-resistance-eueea-ears-net-annual-epidemiological-report-2023.

4. 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. https://doi.org/10.1016/j.cell.2017.04.027.

5. Engel, P., and Moran, N.A. (2013). The gut microbiota of insects – diversity in structure and function. FEMS Microbiol. Rev. *37*, 699–735. https://doi.org/10.1111/1574-6976.12025.

6. Schwartzman, J.A., Lebreton, F., Salamzade, R., Shea, T., Martin, M.J., Schaufler, K., Urhan, A., Abeel, T., Camargo, I.L.B.C., Sgardioli, B.F., et al. (2024). Global diversity of enterococci and description of 18 previously unknown species. Proc. Natl. Acad. Sci. *121*, e2310852121. https://doi.org/10.1073/pnas.2310852121.

7. Paulsen, I.T. (2003). Role of Mobile DNA in the Evolution of Vancomycin-Resistant *Enterococcus* faecalis. Science *299*, 2071–2074. https://doi.org/10.1126/science.1080613.

8. Matos, R.C., Lapaque, N., Rigottier-Gois, L., Debarbieux, L., Meylheuc, T., Gonzalez-Zorn, B., Repoila, F., Lopes, M.D.F., and Serror, P. (2013). *Enterococcus faecalis* prophage dynamics and contributions to pathogenic traits. PLoS Genet. *9*, e1003539. https://doi.org/10.1371/journal.pgen.1003539.

9. Scholl, D. (2017). Phage Tail–Like Bacteriocins. Annu. Rev. Virol. *4*, 453–467. https://doi.org/10.1146/annurev-virology-101416-041632.

10. Backman, T., Burbano, H.A., and Karasov, T.L. (2024). Tradeoffs and constraints on the evolution of tailocins. Trends Microbiol. *32*, 1084–1095. https://doi.org/10.1016/j.tim.2024.04.001.

11. Kageyama, M., Ikeda, K., and Egami, F. (1964). Studies of a pyocin: III. Biological properties of the pyocin. J. Biochem. (Tokyo) *55*, 59–64. https://doi.org/10.1093/oxfordjournals.jbchem.a127841.

12. Heo, Y.-J., Chung, I.-Y., Choi, K.B., and Cho, Y.-H. (2007). R-type pyocin is required for competitive growth advantage between *Pseudomonas aeruginosa* strains. J. Microbiol. Biotechnol. *17*, 180–185.

797 13. Ge, P., Scholl, D., Prokhorov, N.S., Avaylon, J., Shneider, M.M., Browning, C., Buth, S.A., Plattner,
798 M., Chakraborty, U., Ding, K., et al. (2020). Action of a minimal contractile bactericidal nanomachine.
799 Nature *580*, 658–662. https://doi.org/10.1038/s41586-020-2186-z.

800 14. Carim, S., Azadeh, A.L., Kazakov, A.E., Price, M.N., Walian, P.J., Lui, L.M., Nielsen, T.N.,
801 Chakraborty, R., Deutschbauer, A.M., Mutalik, V.K., et al. (2021). Systematic discovery of
802 pseudomonad genetic factors involved in sensitivity to tailocins. ISME J. *15*, 2289–2305.
803 https://doi.org/10.1038/s41396-021-00921-1.

804 15. Blasco, L., De Aledo, M.G., Ortiz-Cartagena, C., Blériot, I., Pacios, O., López, M., Fernández-García,
805 L., Barrio-Pujante, A., Hernández-Garcia, M., Cantón, R., et al. (2023). Study of 32 new phage tail-like
806 bacteriocins (pyocins) from a clinical collection of Pseudomonas aeruginosa and of their potential use
807 as typing markers and antimicrobial agents. Sci. Rep. *13*, 117. https://doi.org/10.1038/s41598-022-
808 27341-1.

809 16. Saha, S., Ojobor, C.D., Li, A.S.C., Mackinnon, E., North, O.I., Bondy-Denomy, J., Lam, J.S.,
810 Ensminger, A.W., Maxwell, K.L., and Davidson, A.R. (2023). F-Type pyocins are diverse noncontractile
811 phage tail-like weapons for killing *Pseudomonas aeruginosa*. J. Bacteriol. *205*, e00029-23.
812 https://doi.org/10.1128/jb.00029-23.

813 17. Zink, R., Loessner, M.J., and Scherer, S. (1995). Charaterization of cryptic prophages (monocins) in
814 *Listeria* and sequence analysis of a holin/endolysin gene. Microbiology *141*, 2577–2584.
815 https://doi.org/10.1099/13500872-141-10-2577.

816 18. Lee, G., Chakraborty, U., Gebhart, D., Govoni, G.R., Zhou, Z.H., and Scholl, D. (2016). F-type
817 bacteriocins of *Listeria monocytogenes*: a new class of phage tail-like structures reveals broad parallel
818 coevolution between tailed bacteriophages and high-molecular-weight bacteriocins. J. Bacteriol. *198*,
819 2784–2793. https://doi.org/10.1128/JB.00489-16.

820 19. Sigal, N., Lichtenstein-Wolfheim, R., Schlussel, S., Azulay, G., Borovok, I., Holdengraber, V., Elad,
821 N., Wolf, S.G., Zalk, R., Zarivach, R., et al. (2024). Specialized *Listeria monocytogenes* produce
822 tailocins to provide a population-level competitive growth advantage. Nat. Microbiol. *9*, 2727–2737.
823 https://doi.org/10.1038/s41564-024-01793-9.

824 20. Gu, Z., Ge, X., and Wang, J. (2025). Structure of an F-type phage tail-like bacteriocin from *Listeria*
825 *monocytogenes*. Nat. Commun. *16*, 1695. https://doi.org/10.1038/s41467-025-57075-3.

826 21. Gebhart, D., Williams, S.R., Bishop-Lilly, K.A., Govoni, G.R., Willner, K.M., Butani, A., Sozhamannan,
827 S., Martin, D., Fortier, L.-C., and Scholl, D. (2012). Novel high-molecular-weight, R-type bacteriocins of
828 *Clostridium difficile*. J. Bacteriol. *194*, 6240–6247. https://doi.org/10.1128/JB.01272-12.

829 22. Cai, X., He, Y., Yu, I., Imani, A., Scholl, D., Miller, J.F., and Zhou, Z.H. (2024). Atomic structures of a
830 bacteriocin targeting Gram-positive bacteria. Nat. Commun. *15*, 7057. https://doi.org/10.1038/s41467-
831 024-51038-w.

832 23. National Center for Biotechnology Information (NCBI) https://www.ncbi.nlm.nih.gov
833 https://www.ncbi.nlm.nih.gov.

834 24. Camargo, A.P., Roux, S., Schulz, F., Babinski, M., Xu, Y., Hu, B., Chain, P.S.G., Nayfach, S., and
835 Kyrpides, N.C. (2024). Identification of mobile genetic elements with geNomad. Nat. Biotechnol. *42*,
836 1303–1312. https://doi.org/10.1038/s41587-023-01953-y.

837 25. Wishart, D.S., Han, S., Saha, S., Oler, E., Peters, H., Grant, J.R., Stothard, P., and Gautam, V.
838 (2023). PHASTEST: faster than PHASTER, better than PHAST. Nucleic Acids Res. *51*, W443–W450.
839 https://doi.org/10.1093/nar/gkad382.

840 26. Bobay, L.-M., Touchon, M., and Rocha, E.P.C. (2014). Pervasive domestication of defective
841 prophages by bacteria. Proc. Natl. Acad. Sci. *111*, 12127–12132.
842 https://doi.org/10.1073/pnas.1405336111.

843 27. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput.
844 Nucleic Acids Res. *32*, 1792–1797. https://doi.org/10.1093/nar/gkh340.

845 28. McBride, S.M., Fischetti, V.A., LeBlanc, D.J., Moellering, R.C., and Gilmore, M.S. (2007). Genetic
846 diversity among *Enterococcus faecalis*. PLoS ONE *2*, e582.
847 https://doi.org/10.1371/journal.pone.0000582.

848 29. Cheng, S., Xing, S., Zhang, X., Pei, G., An, X., Mi, Z., Huang, Y., and Tong, Y. (2016). Complete
849 genome sequence of a new *Enterococcus faecalis* bacteriophage, vB_EfaS_IME197. Genome
850 Announc. *4*, e00827-16. https://doi.org/10.1128/genomeA.00827-16.

851 30. Afonina, I., Tien, B., Nair, Z., Matysik, A., Lam, L.N., Veleba, M., Jie, A.K.J., Rashid, R., Cazenave-
852 Gassiot, A., Wenk, M., et al. (2021). The composition and function of *Enterococcus faecalis* membrane
853 vesicles. microLife *2*, uqab002. https://doi.org/10.1093/femsml/uqab002.

854 31. Jacob, A.E., and Hobbs, S.J. (1974). Conjugal transfer of plasmid-borne multiple antibiotic resistance
855 in *Streptococcus faecalis* var. *zymogenes*. J. Bacteriol. *117*, 360–372.
856 https://doi.org/10.1128/jb.117.2.360-372.1974.

857 32. Dunny, G.M., Brown, B.L., and Clewell, D.B. (1978). Induced cell aggregation and mating in
858 *Streptococcus faecalis*: evidence for a bacterial sex pheromone. Proc. Natl. Acad. Sci. *75*, 3479–3483.
859 https://doi.org/10.1073/pnas.75.7.3479.

860 33. Bourgogne, A., Garsin, D.A., Qin, X., Singh, K.V., Sillanpaa, J., Yerrapragada, S., Ding, Y., Dugan-
861 Rocha, S., Buhay, C., Shen, H., et al. (2008). Large scale variation in *Enterococcus faecalis* illustrated
862 by the genome analysis of strain OG1RF. Genome Biol. *9*, R110. https://doi.org/10.1186/gb-2008-9-7-
863 r110.

864 34. Garsin, D.A. (2004). Peptide signals sense and destroy target cells. Science *306*, 2202–2203.
865 https://doi.org/10.1126/science.1107236.

866 35. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web
867 portal for protein modeling, prediction and analysis. Nat. Protoc. *10*, 845–858.
868 https://doi.org/10.1038/nprot.2015.053.

869 36. Dunne, M., Rupf, B., Tala, M., Qabrati, X., Ernst, P., Shen, Y., Sumrall, E., Heeb, L., Plückthun, A.,
870 Loessner, M.J., et al. (2019). Reprogramming Bacteriophage Host Range through Structure-Guided
871 Design of Chimeric Receptor Binding Proteins. Cell Rep. *29*, 1336-1350.e4.
872 https://doi.org/10.1016/j.celrep.2019.09.062.

873 37. Spinelli, S., Veesler, D., Bebeacua, C., and Cambillau, C. (2014). Structures and host-adhesion
874 mechanisms of lactococcal siphophages. Front. Microbiol. *5*. https://doi.org/10.3389/fmicb.2014.00003.

875 38. Ayala, R., Moiseenko, A.V., Chen, T.-H., Kulikov, E.E., Golomidova, A.K., Orekhov, P.S., Street,
876 M.A., Sokolova, O.S., Letarov, A.V., and Wolf, M. (2023). Nearly complete structure of bacteriophage
877 DT57C reveals architecture of head-to-tail interface and lateral tail fibers. Nat. Commun. *14*, 8205.
878 https://doi.org/10.1038/s41467-023-43824-9.

879 39. d'Acapito, A., Decombe, A., Arnaud, C.-A., and Breyton, C. (2025). Comparative anatomy of
880 siphophage tails before and after interaction with their receptor. Curr. Opin. Struct. Biol. *92*, 103045.
881 https://doi.org/10.1016/j.sbi.2025.103045.

882 40. Xu, Y., Murray, B.E., and Weinstock, G.M. (1998). A cluster of genes involved in polysaccharide
883 biosynthesis from *Enterococcus faecalis* OG1RF. Infect. Immun. *66*, 4313–4323.
884 https://doi.org/10.1128/IAI.66.9.4313-4323.1998.

885 41. Teng, F., Singh, K.V., Bourgogne, A., Zeng, J., and Murray, B.E. (2009). Further characterization of
886 the *epa* gene cluster and Epa polysaccharides of *Enterococcus faecalis*. Infect. Immun. *77*, 3759–3767.
887 https://doi.org/10.1128/IAI.00149-09.

888 42. Palmer, K.L., Godfrey, P., Griggs, A., Kos, V.N., Zucker, J., Desjardins, C., Cerqueira, G., Gevers, D.,
889 Walker, S., Wortman, J., et al. (2012). Comparative genomics of enterococci: variation in *Enterococcus*
890 *faecalis*, clade structure in *E. faecium*, and defining characteristics of *E. gallinarum* and *E. casseliflavus*.
891 mBio *3*, e00318-11. https://doi.org/10.1128/mBio.00318-11.

892 43. Clewell, D.B., Tomich, P.K., Gawron-Burke, M.C., Franke, A.E., Yagi, Y., and An, F.Y. (1982).
893 Mapping of *Streptococcus faecalis* plasmids pAD1 and pAD2 and studies relating to transposition of
894 Tn917. J. Bacteriol. *152*, 1220–1230. https://doi.org/10.1128/jb.152.3.1220-1230.1982.

895 44. Sharpe, M.E. (1964). Serological types of *Streptococcus faecalis* and its varieties and their cell wall
896 type antigen. J. Gen. Microbiol. *36*, 151–160. https://doi.org/10.1099/00221287-36-1-151.

897 45. Thurlow, L.R., Thomas, V.C., and Hancock, L.E. (2009). Capsular polysaccharide production in
898 *Enterococcus faecalis* and contribution of CpsF to capsule serospecificity. J. Bacteriol. *191*, 6203–6210.
899 https://doi.org/10.1128/JB.00592-09.

900 46. Maekawa, S., Yoshioka, M., and Kumamoto, Y. (1992). Proposal of a new scheme for the serological
901 typing of *Enterococcus faecalis* strains. Microbiol. Immunol. *36*, 671–681.
902 https://doi.org/10.1111/j.1348-0421.1992.tb02070.x.

903 47. Guerardel, Y., Sadovskaya, I., Maes, E., Furlan, S., Chapot-Chartier, M.-P., Mesnage, S., Rigottier-
904 Gois, L., and Serror, P. (2020). Complete structure of the Enterococcal Polysaccharide Antigen (EPA)
905 of vancomycin-resistant *Enterococcus faecalis* V583 reveals that EPA decorations are teichoic acids
906 covalently linked to a rhamnopolysaccharide backbone. mBio *11*, e00277-20.
907 https://doi.org/10.1128/mBio.00277-20.

908 48. Davis, J.L., Norwood, J.S., Smith, R.E., O'Dea, F., Chellappa, K., Rowe, M.L., Williamson, M.P.,
909 Stafford, G.P., Vinogradov, E., Maes, E., et al. (2025). Dissecting the Enterococcal Polysaccharide
910 Antigen (EPA) structure to explore innate immune evasion and phage specificity. Carbohydr. Polym.
911 *347*, 122686. https://doi.org/10.1016/j.carbpol.2024.122686.

912 49. Gilmore, M.S., and Haas, W. (2005). The selective advantage of microbial fratricide. Proc. Natl. Acad.
913 Sci. *102*, 8401–8402. https://doi.org/10.1073/pnas.0503828102.

914 50. Chatterjee, A., Johnson, C.N., Luong, P., Hullahalli, K., McBride, S.W., Schubert, A.M., Palmer, K.L.,
915 Carlson, P.E., and Duerkop, B.A. (2019). Bacteriophage resistance alters antibiotic-mediated intestinal
916 expansion of enterococci. Infect. Immun. *87*, e00085-19. https://doi.org/10.1128/IAI.00085-19.

917 51. Salamzade, R., Tran, P.Q., Martin, C., Manson, A.L., Gilmore, M.S., Earl, A.M., Anantharaman, K.,
918 and Kalan, L.R. (2025). zol and fai: large-scale targeted detection and evolutionary investigation of
919 gene clusters. Nucleic Acids Res. *53*, gkaf045. https://doi.org/10.1093/nar/gkaf045.

920 52. Lebreton, F., van Schaik, W., Manson McGuire, A., Godfrey, P., Griggs, A., Mazumdar, V., Corander,
921 J., Cheng, L., Saif, S., Young, S., et al. (2013). Emergence of epidemic multidrug-resistant
922 *Enterococcus faecium* from animal and commensal strains. mBio *4*, e00534-13.
923 https://doi.org/10.1128/mBio.00534-13.

53. Schloissnig, S., Arumugam, M., Sunagawa, S., Mitreva, M., Tap, J., Zhu, A., Waller, A., Mende, D.R., Kultima, J.R., Martin, J., et al. (2013). Genomic variation landscape of the human gut microbiome. Nature *493*, 45–50. https://doi.org/10.1038/nature11711.

54. Prichula, J., Van Tyne, D., Schwartzman, J., Sant'Anna, F.H., Pereira, R.I., da Cunha, G.R., Tavares, M., Lebreton, F., Frazzon, J., d'Azevedo, P.A., et al. (2020). Enterococci from wild Magellanic penguins (*Spheniscus magellanicus*) as an Indicator of marine ecosystem health and human impact. Appl. Environ. Microbiol. *86*, e01662-20, /aem/86/19/AEM.01662-20.atom. https://doi.org/10.1128/AEM.01662-20.

55. Prichula, J., Primon-Barros, M., Luz, R.C.Z., Castro, Í.M.S., Paim, T.G.S., Tavares, M., Ligabue-Braun, R., d'Azevedo, P.A., Frazzon, J., Frazzon, A.P.G., et al. (2021). Genome mining for antimicrobial compounds in wild marine animals-associated enterococci. Mar. Drugs *19*, 328. https://doi.org/10.3390/md19060328.

56. Boehm, A.B., and Sassoubre, L.M. (2014). Enterococci as indicators of environmental fecal contamination. in enterococci: from commensals to leading causes of drug resistant infection, M. S. Gilmore, D. B. Clewell, Y. Ike, and N. Shankar, eds. (Massachusetts Eye and Ear Infirmary).

57. Clewell, D.B., Weaver, K.E., Dunny, G.M., Coque, T.M., Francia, M.V., 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, M. S. Gilmore, D. B. Clewell, Y. Ike, and N. Shankar, eds. (Massachusetts Eye and Ear Infirmary).

58. Hayes, J.R., English, L.L., Carr, L.E., Wagner, D.D., and Joseph, S.W. (2004). Multiple-antibiotic resistance of *Enterococcus* spp. isolated from commercial poultry production environments. Appl. Environ. Microbiol. *70*, 6005–6011. https://doi.org/10.1128/AEM.70.10.6005-6011.2004.

59. Surette, M.D., and Wright, G.D. (2017). Lessons from the environmental antibiotic resistome. Annu. Rev. Microbiol. *71*, 309–329. https://doi.org/10.1146/annurev-micro-090816-093420.

60. Ghequire, M.G.K., and De Mot, R. (2015). The tailocin tale: peeling off phage tails. Trends Microbiol. *23*, 587–590. https://doi.org/10.1016/j.tim.2015.07.011.

61. Krupovic, M., and Koonin, E.V. (2017). Multiple origins of viral capsid proteins from cellular ancestors. Proc. Natl. Acad. Sci. *114*. https://doi.org/10.1073/pnas.1621061114.

62. Sarris, P.F., Ladoukakis, E.D., Panopoulos, N.J., and Scoulica, E.V. (2014). A phage tail-derived element with wide distribution among both prokaryotic domains: a comparative genomic and phylogenetic study. Genome Biol. Evol. *6*, 1739–1747. https://doi.org/10.1093/gbe/evu136.

63. Broad Institute. NCBI Genbank Broad Institute. Data from Bioproject PRJNA313452 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA313452) and PRJNA663900 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA663900).

64. Mass Eye and Ear/ Harvard Medical School. NCBI Genbank. Data from Bioproject PRJNA1371623. https://dataview.ncbi.nlm.nih.gov/object/PRJNA1371623.

65. Gilmore, S. M., Earl, A., Prichula, J., Manson, A, Richards, G., Karumidze, N., Lebreton, F. Compositions and methods for control of specific bacterial populations. US Patent WO2025054631, filed September 10, 2023. https://patentscope.wipo.int/search/en/WO2025054631.

66. OpenAI (2025). ChatGPT (GPT-5.1). https://chat.openai.com

67. Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods *14*, 290–296. https://doi.org/10.1038/nmeth.4169.

68. Posit team (2025). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. https://posit.co.

69. Darling, A.E., Jospin, G., Lowe, E., Matsen, F.A., Bik, H.M., and Eisen, J.A. (2014). PhyloSift: phylogenetic analysis of genomes and metagenomes. PeerJ *2*, e243. https://doi.org/10.7717/peerj.243.

70. Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. *25*, 1043–1055. https://doi.org/10.1101/gr.186072.114.

71. Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. *9*, 5114. https://doi.org/10.1038/s41467-018-07641-9.

72. Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics *30*, 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

73. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: The protein families database in 2021. Nucleic Acids Res. *49*, D412–D419. https://doi.org/10.1093/nar/gkaa913.

74. Kanehisa, M. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. *28*, 27–30. https://doi.org/10.1093/nar/28.1.27.

75. Powell, H.R., Islam, S.A., David, A., and Sternberg, M.J.E. (2025). Phyre2.2: a community resource for template-based protein structure prediction. J. Mol. Biol. *437*, 168960. https://doi.org/10.1016/j.jmb.2025.168960.

76. Karp, P.D., Billington, R., Caspi, R., Fulcher, C.A., Latendresse, M., Kothari, A., Keseler, I.M., Krummenacker, M., Midford, P.E., Ong, Q., et al. (2019). The BioCyc collection of microbial genomes and metabolic pathways. Brief. Bioinform. *20*, 1085–1093. https://doi.org/10.1093/bib/bbx085.

77. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature *630*, 493–500. https://doi.org/10.1038/s41586-024-07487-w.

78. Price, M.N., and Arkin, A.P. (2017). PaperBLAST: Text Mining Papers for Information about Homologs. mSystems *2*, e00039-17. https://doi.org/10.1128/mSystems.00039-17.

79. Schrödinger, L., and DeLano, W. (2020). PyMOL. http://www.pymol.org/pymol

80. Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. *47*, W256–W259. https://doi.org/10.1093/nar/gkz239.

81. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: machine learning in Python. Journal of Machine Learning Research *12*, 2825–2830.

82. Treangen, T.J., Ondov, B.D., Koren, S., and Phillippy, A.M. (2014). The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol. *15*, 524. https://doi.org/10.1186/s13059-014-0524-x.

1005 83. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative
1006 genomics. Genome Biol. *20*, 238. https://doi.org/10.1186/s13059-019-1832-y.

1007 84. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7:
1008 improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.
1009 https://doi.org/10.1093/molbev/mst010.

1010 85. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler
1011 transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

1012 86. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A., and
1013 Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the
1014 genomic era. Mol. Biol. Evol. *37*, 1530–1534. https://doi.org/10.1093/molbev/msaa015.

1015 87. Kutter, E. (2009). Phage host range and efficiency of plating. In bacteriophages methods in molecular
1016 biology., M. R. J. Clokie and A. M. Kropinski, eds. (Humana Press), pp. 141–149.
1017 https://doi.org/10.1007/978-1-60327-164-6_14.

1018 88. Carlson, K. Working with bacteriophages: common techniques and methodological approaches, vol
1019 1. CRC Press, Boca Raton, FL.

1020 89. CD Genomics. https://www.cd-genomics.com

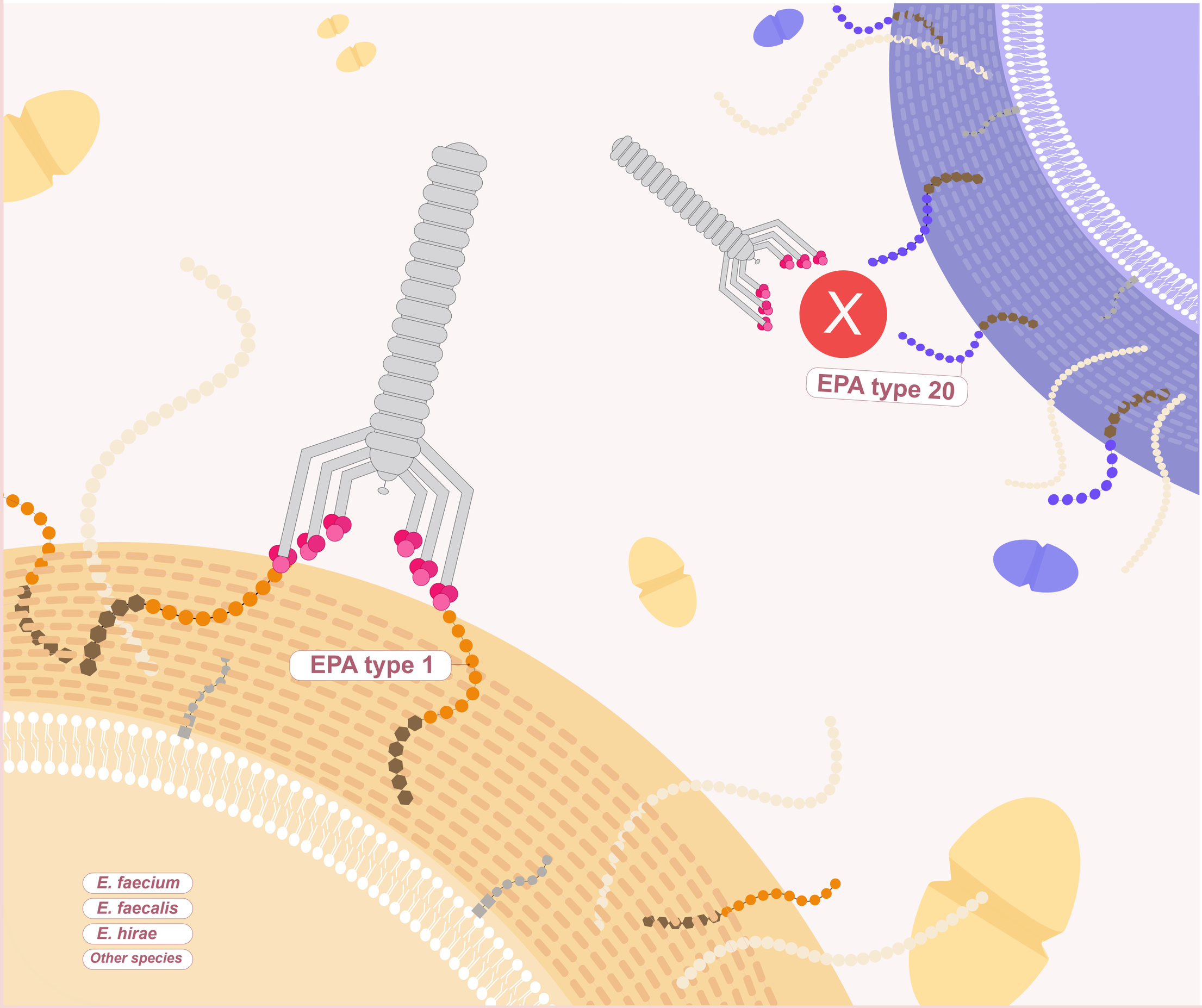
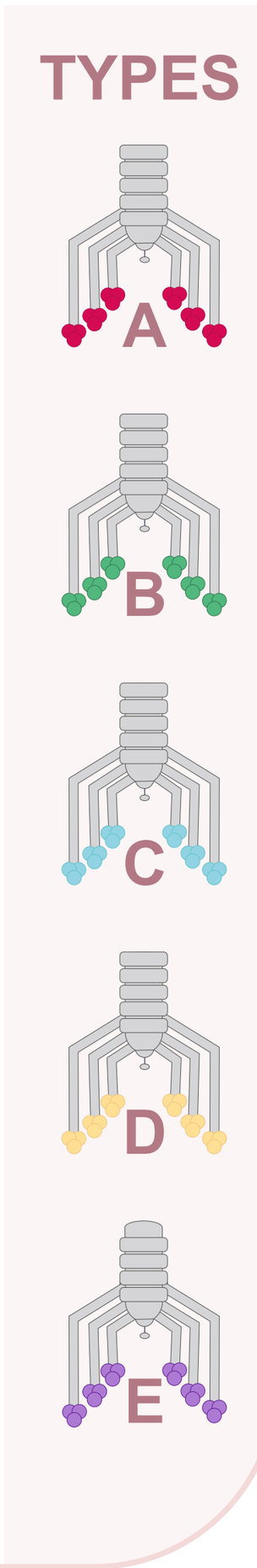
1021 90. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q.,
1022 Wortman, J., Young, S.K., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant
1023 detection and genome assembly improvement. PLoS ONE *9*, e112963.
1024 https://doi.org/10.1371/journal.pone.0112963.

1025 91. Van Dijk, L.R., Walker, B.J., Straub, T.J., Worby, C.J., Grote, A., Schreiber, H.L., Anyansi, C.,
1026 Pickering, A.J., Hultgren, S.J., Manson, A.L., et al. (2022). StrainGE: a toolkit to track and characterize
1027 low-abundance strains in complex microbial communities. Genome Biol. *23*, 74.
1028 https://doi.org/10.1186/s13059-022-02630-0.

1029 92. Georgescu, C.H., Manson, A.L., Griggs, A.D., Desjardins, C.A., Pironti, A., Wapinski, I., Abeel, T.,
1030 Haas, B.J., and Earl, A.M. (2018). SynerClust: a highly scalable, synteny-aware orthologue clustering
1031 tool. Microb. Genomics *4*. https://doi.org/10.1099/mgen.0.000231.

1032 93. Adobe Illustrator (Version 2026). https://www.adobe.com/products/illustrator.html.

EFAGINS

E. faecalis phage-related inhibitors

- Conserved *E. faecalis* genomes
- Target include VRE
- Five natural variants identified
- Programmable to target EPA variations