Extreme genetic diversity in the type VII secretion system of *Listeria monocytogenes* suggests a role in bacterial antagonism

Kieran Bowran and Tracy Palmer*

**Abstract**

The type VII protein secretion system (T7SS) has been characterized in members of the phyla Actinobacteria and Firmicutes. In mycobacteria the T7SS is intimately linked with pathogenesis and intracellular survival, while in Firmicutes there is mounting evidence that the system plays a key role in interbacterial competition. A conserved membrane-bound ATPase protein, termed EssC in *Staphylococcus aureus*, is a critical component of the T7SS and is the primary receptor for substrate proteins. Genetic diversity in the *essC* gene of *S. aureus* has previously been reported, resulting in four protein variants that are linked to specific subsets of substrates. Here we have analysed the genetic diversity of the T7SS-encoding genes and substrate proteins across *Listeria monocytogenes* genome sequences. We find that there are seven EssC variants across the species that differ in their C-terminal region; each variant is correlated with a distinct subset of genes for likely substrate and accessory proteins. EssC1 is most common and is exclusively linked with polymorphic toxins harbouring a YeeF domain, whereas EssC5, EssC6 and EssC7 variants all code for an LXG domain protein adjacent to *essC*. Some *essC1* variant strains encode an additional, truncated *essC* at their T7 gene cluster. The truncated EssC, comprising only the C-terminal half of the protein, matches the sequence of either EssC2, EssC3 or EssC4. In each case the truncated gene directly precedes a cluster of substrate/accessory protein genes acquired from the corresponding strain. Across *L. monocytogenes* strains we identified 40 LXG domain proteins, most of which are encoded at conserved genomic loci. These loci also harbour genes encoding immunity proteins and sometimes additional toxin fragments. Collectively our findings strongly suggest that the T7SS plays an important role in bacterial antagonism in this species.

**INTRODUCTION**

Protein secretion systems are found in almost all bacteria and are critical for processes such as nutrient capture and niche adaptation. To date, 10 distinct protein secretion systems (named types I–X) have been identified [1–3]. While most of these are exclusive to Gram-negative bacteria, the type VII secretion system (T7SS) is encoded by many Gram-positive Actinobacteria and Firmicutes [4, 5]. The T7SS has been heavily studied in pathogenic mycobacteria, where it is essential for virulence [6–8]. It is also linked to virulence in *Staphylococcus aureus*, supporting abscess formation during persistent infection [9, 10].

The T7SS comprises two subtypes – the T7a found in Actinobacteria and T7b in Firmicutes, that although related, have distinct differences [11, 12]. A hexameric membrane-bound ATPase, termed EccC in Actinobacteria and EssC in Firmicutes, is common to both systems and is a critical component of the T7SS [13, 14]. It lies at the centre of a multi-subunit membrane complex that mediates protein transport across the cytoplasmic membrane [15–17]. The protein has four ATP-binding domains in its cytoplasmic C-terminus, each of which is essential for secretion activity [13, 14, 16–19], and the most C-terminal of these domains also plays a role in substrate recognition [13, 20, 21]. Substrates are recognized through a short C-terminal signal sequence, and secretion appears to be post-translational because at least some substrates are exported in a folded form [22–24].

In Firmicutes three additional membrane proteins, EsaA, EssA and EssB, and the cytoplasmic protein EsaB function alongside EssC to mediate T7 secretion [9, 25–29] (Fig. 1). Characterized substrates of the T7b system include proteins from the WXG and LXG families [9, 24, 30, 31]. In all T7b
Fig. 1. The Firmicutes T7b secretion system. (a) Loci encoding the functional components of the T7b secretion system in *Staphylococcus aureus*, *Listeria monocytogenes* and *Bacillus subtilis*. The gene numbers for *L. monocytogenes* relate to the reference strain EGDe. (b) Subcellular location and topology of the T7b components. The two forkhead-associated domains (FHA1/FHA2) and the four ATPase domains (D0, D1, D2 and D3) of EssC are indicated. (c) The ess (T7) gene cluster of *S. aureus* strain NCTC8325. Genes encoding functional components are shaded in blue, secreted substrates in yellow, a chaperone protein in white and EsaG family immunity proteins in turquoise. (d) Schematic representation of *L. monocytogenes* EssC. Amino acid positions of the predicted domains are shown and the region that is variable across strains is indicated. Amino acid 771, shown boxed in red, is the position at which the truncated EssC proteins encoded in some *L. monocytogenes* essC1 strains start to align with the full-length protein.
systems a conserved WXG protein, EssA, is present. In *S. aureus* EssA is essential for T7 secretion activity and may act as a structural/functional component of the secretion system [9, 25]. The genes encoding these essential T7 components are arranged contiguously (Fig. 1), and are often found in gene clusters encoding secreted substrates and accessory proteins [9, 29, 32].

The *S. aureus* T7SS is encoded at the *ess* locus. Although part of the core genome, it was noted that there is significant genetic diversity at the *ess* region across *S. aureus* strains [33]. Sequence divergence occurs part way through *essC*, giving rise to four *essC* variants that differ in their C-terminal domains. Each *essC* variant is associated with a specific set of genes encoding T7 substrate proteins, and the variable C-terminal domain of *EssC* has been implicated in the recognition of strain-specific substrates [21, 33].

The most commonly studied *S. aureus* strains encode the *EssC*1 variant (Fig. 1c). These strains produce a T7-secreted nuclease, EssD, a member of the YeeF protein family, and are protected from self-intoxication by an immunity protein, EsaG [34, 35]. A chaperone protein, EssE, interacts with both EssD and EsaC and likely targets the nuclease to the secretion machinery [34, 36]. Strains harbouring *essC1* account for approximately 50% of sequenced *S. aureus* [33]. Analysis of EssD sequences across these strains shows that the nuclease domain is polymorphic, suggestive of an anti-bacterial toxin [34]. All *S. aureus* strains, including *essC2*, *essC3* and *essC4* strains that do not encode EssD, have multiple genes coding for EsaG homologues at their *ess* loci, suggesting that EssD targets staphylococci (Fig. 1c) [33, 34]. An anti-staphylococcal role for EssD has been confirmed in competition experiments, where strains were sensitized to killing by the nuclease when *esaG* genes were deleted [34, 37].

Since the identification of EssD, a growing number of T7-secreted antibacterial toxins have been described. TspA, a second polymorphic toxin found in *S. aureus*, has membrane-depolarizing activity that is neutralized by the Tsal immunity protein [37]. *Streptococcus intermedius* produces at least three antibacterial toxins, of which TelB is an NADase and TelC a lipid II phosphatase [30]. Moreover, recent reports indicate that the T7b systems of *Bacillus subtilis* and *Enterococcus faecalis* also have antibacterial activity [38, 39], indicating that bacterial antagonism is a common feature of several Firmicutes T7SSs.

*Listeria monocytogenes* is a Gram-positive firmicute and human foodborne pathogen that is able to invade and replicate intracellularly [40]. A T7SS is encoded in the genome and has been partially characterized in strains EGDe and 10403s, where it was shown to be dispensable for virulence and host cell invasion [32, 41]. Here we have analysed the genetic diversity at the T7 gene clusters of sequenced *L. monocytogenes* strains. We show that seven variants of *EssC*, which differ in their C-terminal region, are encoded across *L. monocytogenes*. Each variant is associated with specific genes encoding likely substrate proteins. We identify polymorphic toxins that fall into the YeeF (EssD) and LXG families and clusters of immunity proteins for self and non-self toxins. Our observations point to a major role for the T7SS in interbacterial competition.

**METHODS**

*L. monocytogenes* EGDe serovar 1/2a (NC_003210.1) was used as the reference strain in this analysis. EssC, encoded by *lmo0061*, was used as the query for Basic Local Alignment Search Tool Protein (BLASTp) analysis using the National Center for Biotechnology Information’s (NCBI’s) Reference Sequence (RefSeq) database of non-redundant protein sequences [42]. All identified *L. monocytogenes* EssC homologues were assigned a percentage identity to Lmo0061 using the BLOSUM62 scoring matrix and an expect threshold of 10. EssC sequence variants were aligned using the clustalw algorithm of the European Bioinformatic Institute (EBI) via the Clustal omega software (v.1.2.4) [43], and protein sequence alignments were viewed in Jalview [v.2.11.1.0] [44]. EssC sequences were further analysed through the construction of maximum-likelihood phylogenetic trees using the MEGAX software package (v.10.1.8) [45, 46]. The Pasteur Institute’s Listeromics tool [47] was used to acquire sequences of EssCs across 60 annotated *L. monocytogenes* genomes with information on lineage, clonal clusters and serotypes. The presence of sigma 70 promoters and terminator regions within the T7b locus was evaluated using BPROM and FindTerm, respectively [48]. Candidate Rho-independent terminator regions were analysed using ARNold [49].

Protein accession numbers were subjected to flanking genes (FlaGs) analysis [50] to establish the conservation of T7b genes across multiple *L. monocytogenes* strains. Gene products were analysed using BLASTp analysis, and the presence of transmembrane regions was predicted using TMHMM [51]. Homology modelling was achieved using the Protein Homology/analogy Recognition Engine (PHYRE [52]).

**RESULTS**

The T7SS gene cluster of *L. monocytogenes*

The functional components of the *L. monocytogenes* T7SS are encoded contiguously from *lmo0056* (*esxA*) through to *lmo0061* (*essC*; Fig. 1a; all gene numbers relate to type strain EGDe), with the gene order mirroring that of the *S. aureus* *ess* locus. In *S. aureus*, a peptidoglycan hydrolase, EssH, is encoded divergently to *esxA* and is essential for T7 substrate secretion across the cell wall [53]. Analysis of the genomic region immediately upstream of *esxA* in *L. monocytogenes* did not identify a homologue of *essH* at this locus, or elsewhere on the chromosome. A stem–loop structure is present in the *S. aureus* *esxA–esaA* intergenic region [25]. A similar stem–loop with an estimated Gibbs free energy of −13.6 kcal mol⁻¹ and sharing 62% identity with that of *S. aureus*, is conserved across the sequenced *L. monocytogenes* genomes. In *S. aureus* the stem–loop plays a role in regulating the expression level of the genes immediately downstream of *esxA*, which are transcribed 10–100 fold lower than *esxA* [25, 54]. It likely
has a similar role in *L. monocytogenes*, as it has been reported that *essC* and *essB* are much more weakly expressed than *esxA* in strain EGDe [41].

**Seven distinct EssC variants are encoded across *L. monocytogenes* strains**

Analysis of the amino acid sequence of the T7 core components indicates that five of them (EsxA, EsaA, EsaB, EssA and EssB) have high sequence conservation (approximately 97% identity across strains present in RefSeq). By contrast, EssC sequences are much more variable, with less than 80% identity between strains. To further analyse the sequence variability of EssC, all *L. monocytogenes* full-length EssC sequences (274 in total) were extracted from NCBI RefSeq and aligned. It was observed that the proteins fall into seven different groups, here named EssC1–7. The sequences of EssC1 and EssC3 diverge from the other EssC groups at around the start of domain D2 (Figs 1a and S1, available in the online version of this article). The five other EssC variants are almost invariant in their proximal D2 sequences, starting to diverge from one another approximately 180 amino acids into this domain. As shown in Table 1, EssC1 is the most common variant, being encoded in 26% of sequenced *L. monocytogenes* strains, and EssC6 the least common, with only 3% occurrence.

To assess the distribution of *essC* variants across clonal complexes and evolutionary lineages [55, 56], the 65 complete *L. monocytogenes* EssC sequences were extracted from the Pasteur Institute's publicly available annotated collection and used to construct a maximum-likelihood phylogenetic tree (Fig. 2). While essC6 strains are notably absent from the phylogenetic tree, the six other EssC variants are represented. The tree divides into two main branches, with the first comprising EssC2, EssC4, EssC5 and EssC7, and the second solely comprising EssC1. As with the NCBI RefSeq database, the majority of the Pasteur Institute sequences encode EssC1 or EssC2, which both show distribution across evolutionary lineages I and II. The remaining essC variants are confined to a single lineage, with the exception of *L. monocytogenes* FSL J1-208, which is the only essC7 strain to belong to lineage IV out of those analysed. The four essC3 strains belonging to CC69 of lineage III cluster outside of the main two branches away from the other variants.

### Table 1. Seven EssC variants identified encoded across *L. monocytogenes* genome sequences present in RefSeq

| Variant | Sequence length (aa) | C-terminal amino acid sequence | No. of sequences | Percentage frequency |
|---------|----------------------|--------------------------------|------------------|---------------------|
| EssC1   | 1498                 | IAPNQ                          | 71               | 26%                 |
| EssC2   | 1498                 | YEEEV                          | 56               | 23%                 |
| EssC3   | 1496                 | KIPKG                          | 56               | 16%                 |
| EssC4   | 1501                 | RMPKH                          | 36               | 14%                 |
| EssC5   | 1483                 | IKLPN                          | 33               | 13%                 |
| EssC6   | 1500                 | VKEEE                          | 7                | 3%                  |
| EssC7   | 1490                 | VLSKS                          | 12               | 5%                  |

**Two variable genomic regions are found at the T7 gene cluster of *L. monocytogenes***

It was noted for *S. aureus* that genetic diversity at the ess locus starts within the essC gene and continues downstream [33]. Inspection of the genomic region 3′ of *L. monocytogenes* essC reveals that it is exceptionally diverse across different strains, in agreement with prior studies that identified this region as a hypervariable hotspot [57, 58]. We identified two closely spaced regions of high variability (Figs 3a and S2). The first of these variable regions is bounded by essC at the 5′ side and at the 3′ side by a cluster of housekeeping genes (*lmot0075-lmot011*; encoding a predicted phosphoenolpyruvate mutase, 6-O-methylguanine DNA methyltransferase, a protein of the YjbI superfamily, a predicted p-isomer-specific 2-hydroxyacid dehydrogenase and a tRNA, respectively). The second variable region lies between *lmot1* at the 5′ end and *lm00082* (encoding a hypothetical membrane protein) at the 3′ end (Fig. 3a).

**YeeF family toxins are encoded in variable region 1 of *L. monocytogenes* essC1 strains**

Although variable region 1 is highly diverse across *L. monocytogenes*, strains with the same essC variant show some common features. For example, a shared set of genes are encoded immediately downstream of essC in all essC1 strains (Figs 3b and S2). The first three of these genes, *lm00065*, *lm00066* and *lm00064*, are syntenous with the three genes that are adjacent to essC1 in *S. aureus* strains (Fig. 1c), and encode homologues of the substrate proteins EssC (Lmo0062) and EssB (Lmo0063), and the chaperone protein EssE (Lmo0064). Lmo0065 is also common to all essC1 strains and encodes a small helical protein with some limited similarity to *S. aureus* EssD.

The essC1 sequences start to diverge at *lm00066*. In each essC1 strain this gene codes for a protein of the YeeF superfamily, to which the *S. aureus* toxin EssD also belongs. However, we note that the encoded proteins have a highly variable C-terminal toxin domain across variant 1 strains. In total, 10 different toxin domains could be identified (Fig. S3, Table 2). In each case, the N-terminal 400 amino acids (the ‘YeeF domain’) are highly conserved and this region is predicted to be almost completely α-helical. In *S. aureus*
Fig. 2. Phylogenetic distribution of essC variants across strains of *L. monocytogenes*. Maximum-likelihood phylogenetic tree showing the presence of multiple essC variants across the four evolutionary lineages (I–IV) of *L. monocytogenes*. CC, clonal complex; ST, sequence type. The scale bar represents the number of substitutions per site.
Fig. 3. Two genetically variable regions can be identified downstream of essC. (a) Schematic representation of the T7-encoding region on the chromosome of *L. monocytogenes*, highlighting the positions of two genetically variable regions. Genes shown in blue encode functional components of the T7SS and genes in black are invariant. The gene shown in dashed lines encodes a tRNA. (b, c) Three genetic arrangements of variable region 1 from (b) *essC1* and (c) *essC2* strains. Genes encoding candidate toxins are shaded yellow with the cognate immunity gene in green. Probable orphan immunity proteins are shown in peach (with likely immunity indicated where identified) and genes encoding toxin fragments in hatched grey and yellow shading. Genes numbered 21–28 are conserved in the same position across all *essC2* strains. Gene 24 in this cluster is predicted to encode an immunity protein and is also found in the EGD-e *essC1* strain. Genes numbered 44–47 are common to *essC4* strains. Gene sizes and intergenic regions not to scale.
variant 1 strains, the EsaE chaperone recognizes the YeeF domain of EsaD, and it is likely that this highly conserved domain in *L. monocytogenes essC1* toxins is also recognized by EsaE. The toxin domains are predicted to have functions such as ADP ribosyltransferase, ribonuclease or NADase activity (Table 2), and each toxin is always paired with the same candidate immunity protein encoded directly downstream. Some *essC1* strains encode two YeeF domain proteins within this region, and some also encode a protein with a tuberculosis necrotizing toxin (TNT) domain, although this lacks a detectable YeeF domain (Figs 3b and S2). Other genes found within this region of *essC1* strains encode toxin fragments and orphan immunity proteins.

### A conserved set of genes are found downstream of *essC2*

All *essC2* strains have a common set of genes immediately downstream of *essC* (Fig. 3c; numbered 21–28 for variant 2 genes 1–8). Most of these genes encode small proteins (>150 amino acids) of unknown function. The largest protein encoded in this cluster, represented by B0647_RS15290 (gene 26 in Fig. 3c) is approximately 320 amino acids in size and is strongly predicted to be a α-isomer-specific 2-hydroxyacid dehydrogenase family protein. Curiously, this is the same predicted function as that for the gene encoded by *lmo0078* in the invariant housekeeping cluster, although the two proteins have no detectable sequence identity. Two of the genes in the *essC2* variable region 1 cluster encode related proteins (24 in Fig. 3c). This duplication is seen in all strains, and there is always an approximate 500 bp intergenic region between the last two genes in the cluster.

Genes 24–28 in this cluster seem to form a module of six genes (always in the order 24–25–26–27–28–24; with an extended intergenic region between genes 28 and 24). This module is often but not always found in variable region 1 of other *essC* variants, and sometimes additional copies of gene 24 also flank this cluster (Fig. S2). Within this module genes 24 and 25 may form a pair, as they are occasionally co-occur away from the other genes of the module. Many orphan copies of 24 are also found in the T7 gene clusters of most *essC* variants, suggesting that it may encode an immunity protein.

Downstream of these nine conserved genes the *essC2* strains are much more variable, but many strains encode multiple predicted antitoxins, including to the YeeF and LXG domain toxins present in other *L. monocytogenes* strain variants.

### *essC3* strains may encode a novel toxin

Many *essC3* strains analysed have a very short variable region 1, comprising two genes of unknown function (Fig. 4a). The second gene of the pair encodes a protein of approximately 660 amino acids, which is a similar size to other T7b toxins, although it is unrelated to either the LXG or YeeF domain proteins. It therefore may represent a novel toxin. A few *essC3* strains have larger gene clusters in variable region 1, with a few carrying the six-gene module [24, 25] that is present in all *essC2* strains. Orphan immunity proteins to other T7b toxins are also found in this region in some strains.

### Seven conserved genes are found downstream of *essC4*

The first seven genes downstream of *essC4* are conserved in all *essC4* strains (41–47 in Fig. 4b). Most of the encoded proteins are <200 amino acids; the largest protein encoded by the cluster, at 433 amino acids, is the product of gene 44. Most of the proteins lack any identifiable domains, with the exception of 41, which is a member of the SMC_prok_A

| Toxin | Sequence length (aa) | Predicted function | Probable immunity |
|-------|---------------------|--------------------|-------------------|
| Toxic 1 | 604 | VIP2 superfamily: ADP ribosyltransferase | Im1: ADP ribosyl glycohydrolase |
| Toxic 2 | 557 | Unknown | Im2: unknown |
| Toxic 3 | 577 | Predicted ribonuclease | Im3: SUKH-1 family immunity protein |
| Toxic 4 | 565 | Predicted ribonuclease | Im4: SUKH-1 family immunity protein |
| Toxic 5 | 564 | Structurally related to Tne2: NADase? | Im5: bipartite? haemagglutinin-related protein/Ssp translocon-related protein |
| Toxic 6 | 598 | Ntox15 superfamily: predicted ribonuclease | Im6: DUF1851 domain-containing protein: structurally related to type VI immunity protein Atu4351 from *Agrobacterium tumefaciens* |
| Toxic 7 | 563 | Structurally related to Tne2: NADase? | Im7: unknown |
| Toxic 8 | 609 | Structurally related to *Streptococcus pyogenes* Spn: NAD+ glycohydrolase? | Im8: unknown |
| Toxic 9 | 556 | Unknown | Im9: unknown |
| Toxic 10 | 578 | Unknown | Im 10: unknown |
superfamily, and the C-terminal 65 amino acids of 45, which is an Mfa2 superfamily domain.

The last four genes in this cluster form a module – they are found in variable region 1 of other essC variant strains (e.g. Figs 3c and S2). Interestingly, in this module gene 44 is shorter and encodes a protein of only 335 amino acids. The shorter product aligns with 100% sequence identity to the longer 44 protein from amino acid 99 onwards, but lacks the N-terminal

---

Fig. 4. Genetic variability of region 1 in essC3 and essC4 strains. (a, b) Three genetic arrangements of variable region 1 from (a) essC3 and (b) essC4 strains. Genes encoding candidate toxins are shaded yellow. Probable orphan immunity proteins are shown in peach (with likely immunity indicated where identified) and genes encoding toxin fragments in hatched grey and yellow shading. Genes numbered 41–47 are common across all essC4 strains. Genes numbered 24–28 are common to essC2 strains, and genes numbered 61 and 62 are common to essC6 strains. Gene sizes and intergenic regions not to scale.
region. Several essC4 strains also encode a copy of the essC2 six-gene module at this locus (Fig. 4b).

**essC5, essC6 and essC7 strains encode an LXG toxin in variable region 1**

Variable region 1 of essC5, essC6 and essC7 strains shares some superficial similarity (Fig. 5). An LXG family toxin is always encoded by the third gene downstream of essC, followed by a probable immunity protein. The three toxins (toxin A in essC5 strains, toxin B in essC6 strains and toxin C in essC7 strains) share some limited sequence identity (approximately 15–20%; Table 3, Fig. S4) in the first ~270 amino acids, but are less similar in the C-terminal toxin domains. Nonetheless, all three proteins can be structurally modelled with

---

**Fig. 5.** An LXG toxin is encoded in variable region 1 of essC5, essC6 and essC7 strains. (a, c) Two genetic arrangements of variable region 1 from (a) essC5 and (c) essC7. (b) A single genetic arrangement is seen for essC6 strains. Genes encoding candidate toxins are shaded yellow with the cognate immunity gene in green. Probable orphan immunity proteins are shown in peach (with likely immunity indicated where identified). Genes numbered 61–64 are common to essC6 strains. Genes numbered 24–28 are common to essC2 strains and those numbered 44–47 are common to essC4 strains. Gene sizes and intergenic regions not to scale.
| LXG toxin | Sequence length (aa) | Predicted function | Probable immunity: function | Occurrence/location |
|-----------|----------------------|--------------------|-----------------------------|---------------------|
| Toxin A   | 438                  | Structurally related to colicin Ia: membrane depolarizing? | ImA: polytopic membrane protein | essC5 strains variable region 1 |
| Toxin B   | 423                  | Structurally related to colicin Ia: membrane depolarizing? | ImB: DUF5082 domain protein | essC6 strains variable region 1 |
| Toxin C   | 537                  | Structurally related to colicin Ia: membrane depolarizing? | ImC: unknown | essC7 strains variable region 1 |
| Toxin D   | 492                  | Unknown             | ImD: unknown | essC1, essC2, essC3, essC4, essC5, essC6, essC7 strains variable region 2 |
| Toxin E   | 493                  | Unknown             | ImE: unknown | essC1, essC2, essC4, essC5 strains variable region 2 |
| Toxin F   | 507                  | Structurally related to Tne2: NADase? | ImF: unknown | essC1, essC2, essC3, essC7 strains variable region 2 |
| Toxin G   | 479                  | Structurally related to colicin D/E5 nuclease domain | ImG: unknown | essC1, essC2, essC3, essC5, essC7 strains variable region 2 |
| Toxin H   | 462                  | Unknown             | ImH: DUF1871 family protein | Cluster of ABC transporter-encoding genes |
| Toxin I   | 466                  | Unknown             | ImI: unknown | Cluster of ABC transporter-encoding genes |
| Toxin J   | 435                  | Structurally related to colicin D/E5 nuclease domain | ImJ: unknown | Cluster of ABC transporter-encoding genes |
| Toxin K   | 436                  | Unknown             | ImK: unknown | Cluster of ABC transporter-encoding genes |
| Toxin L   | 463                  | Structurally related to colicin D/E5 nuclease domain | ImL: unknown | Cluster of ABC transporter-encoding genes |
| Toxin M   | 468                  | Structurally related to NAD+ glycohydrolase | ImM: immunity protein 74; pfam15603 | Cluster of ABC transporter-encoding genes |
| Toxin N   | 478                  | EndoU-nuclease      | ImN: unknown | Cluster of ABC transporter-encoding genes |
| Toxin O   | 465                  | Unknown             | ImO: unknown | Cluster of ABC transporter-encoding genes |
| Toxin P   | 477                  | EndoU-nuclease      | ImP: unknown | Cluster of ABC transporter-encoding genes |
| Toxin Q   | 475                  | EndoU-nuclease      | ImQ: unknown | Cluster of ABC transporter-encoding genes |
| Toxin R   | 490                  | EndoU-nuclease      | ImR: unknown | Cluster of ABC transporter-encoding genes |
| Toxin S   | 491                  | Unknown             | ImS: unknown | Insertion between DNA-directed RNA polymerase and transketolase-encoding genes |
| Toxin T   | 517                  | Unknown             | ImT: unknown | Insertion between DNA-directed RNA polymerase and transketolase-encoding genes |
| Toxin U   | 509                  | Unknown             | ImU: unknown | Insertion between DNA-directed RNA polymerase and transketolase-encoding genes |
| Toxin V   | 482                  | Unknown             | ImV: unknown | Insertion between DNA-directed RNA polymerase and transketolase-encoding genes |
| Toxin W   | 469                  | Structurally related to colicin D/E5 nuclease domain | ImW: cysteine-rich CPCC superfamily protein | Insertion between DNA-directed RNA polymerase and transketolase-encoding genes |
| LXG toxin | Sequence length (aa) | Predicted function | Probable immunity: function | Occurrence/location |
|-----------|----------------------|--------------------|-----------------------------|--------------------|
| Toxin X   | 478                  | Unknown            | ImX: unknown                | Insertion between DNA-directed RNA polymerase and transketolase-encoding genes |
| Toxin Y   | 551                  | Structurally related to *Legionella* effector SdeA | ImY: immunity 70 family protein | Insertion between DNA-directed RNA polymerase and transketolase-encoding genes |
| Toxin Z   | 458                  | Unknown            | ImZ: unknown                | Insertion between leucine-rich repeat surface protein family BspA and octanocin A resistance protein OhrB |
| Toxin α   | 470                  | Unknown            | Imα: unknown                | Insertion between leucine-rich repeat surface protein family BspA and octanocin A resistance protein OhrB |
| Toxin β   | 491                  | Structurally related to the C-terminal Tip2 of CdiA from *Pseudomonas aeruginosa* – haemagglutinin | Imβ: member of the prominin superfamily of membrane proteins | Insertion between nitrate monoxygenase and cation-translocating P-type ATPase-encoding genes |
| Toxin γ   | 433                  | Unknown            | Imγ: unknown                | Insertion between haem oxygenase IsdG and NUDIX hydrolase-encoding genes |
| Toxin δ   | 457                  | Unknown            | Imδ: unknown                | Insertion between haem oxygenase IsdG and NUDIX hydrolase-encoding genes |
| Toxin ε   | 492                  | EndoU-nuclease     | Imε: unknown                | Insertion between haem oxygenase IsdG and NUDIX hydrolase-encoding genes |
| Toxin ζ   | 502                  | Structurally related to TelC – probable lipid II phosphatase | Imζ: unknown | Insertion between effector binding domain-containing protein and DeoR/GlpR transcription factor encoding genes |
| Toxin η   | 512                  | Unknown            | Imη: unknown                | Insertion between effector binding domain-containing protein and DeoR/GlpR transcription factor encoding genes |
| Toxin θ   | 540                  | Unknown            | Imθ: polytopic membrane protein | Insertion between effector binding domain-containing protein and DeoR/GlpR transcription factor encoding genes |
| Toxin ι   | 466                  | Unknown            | Imι: unknown                | Insertion between glutamine-hydrolysing GMP synthase and α/β hydrolase-encoding genes* |
| Toxin μ   | 481                  | Unknown            | Imμ: unknown                | Insertion between glutamine-hydrolysing GMP synthase and α/β hydrolase-encoding genes* |
| Toxin ν   | 462                  | Unknown            | Imν: unknown                | Insertion between glutamine-hydrolysing GMP synthase and α/β hydrolase-encoding genes* |
| Toxin ρ   | 514                  | Structurally related to *Legionella* effector SdeA | Imρ: unknown | Insertion between chitinase ChiB and ABC transporter ATP-binding protein-encoding genes |
| Toxin σ   | 491                  | Fic superfamily    | Imσ: unknown                | Insertion between lytic polysaccharide monoxygenase and TeiR/AcrR family regulator-encoding genes |
| Toxin ϕ   | 485                  | Unknown            | Imϕ: unknown                | Insertion between histidine phosphatase family protein and ATP-dependent RNA helicase DhpB-encoding genes |

*Two LXG toxins are often encoded at this locus in a head-to-head arrangement.*
99% confidence using colicin Ia as a template (PHYRE [52]). Recently it has been shown that TspA, a T7-secreted protein from *S. aureus* that can also be modelled on the colicin Ia, is a membrane-depolarizing toxin [37], suggesting that these three *L. monocytogenes* toxins may also have a similar mode of action. The candidate immunity proteins for these toxins fall into three unrelated protein families (Figs 5 and S2, Table 3).

The essC5 strains always encode a pair of small proteins of unknown function in-between essC5 and the toxin A gene. Downstream of the candidate immunity gene the essC5 strains start to diverge (Fig. 5a), with some strains encoding the six-gene module found in essC2 strains, while others have the five-gene module found in essC4.

All essC6 strains sequenced to date have an identical set of genes in variable region 1 (Fig. 5b). The first gene in this region encodes a 367 amino acid product that appears to be structurally related to a metallopeptidase (match to the crystal structure of the m16b metallopeptidase subunit from *Sphingomonas* sp. a1 with 96% confidence; HYRE [52]). The second gene encodes a 93 amino acid peptide that is a member of the DUF3130 family. Following the toxin and immunity gene are a cluster of five genes encoding proteins of between 136 and 151 amino acids. The final four genes in this cluster form a module (61–64) that is also found in variable region 1 of some essC5 and essC7 strains (Figs 5c and S2). The first two genes of the module are also found in some essC3 strains (Fig. 4).

The majority of essC7 strains have a relatively short variable region 1, comprising genes encoding two small proteins (115 and 113 amino acids, respectively) alongside the toxin- and immunity-encoding genes (Fig. 5c). A few strains also have the essC6 four gene module at this region (Fig. 5c).

### Some essC variant 1 strains encode truncated EssC2, EssC3 and EssC4 variants

During our analysis we observed that some *essC1* strains encode the conserved sets of genes found in variable region 1 of other essC variant strains. For example, strain N11-1255 encodes the complete eight-gene module [21–28] found in essC2 strains, strain FSL L7-0763 encodes the candidate toxin found in essC3 strains and strain CFSAN026581 encodes the full seven-gene module from essC4 strains (Fig. 6a). In each of these cases, a truncated EssC protein comprising the C-terminal ~728 amino acids is also encoded immediately preceding these ‘foreign’ genes. The truncated EssCs (which are annotated as pseudogenes on the FlaGs output in Fig. S2) are much shorter than the full-length EssC1 protein encoded by these strains; each one aligns to the canonical EssC sequence at residue 771 (Fig. 1d) and has a potential initiator methionine a few codons away. Strikingly, these truncated EssCs do not have the EssC1 variable region, instead the variable region is from the same essC variant as the foreign genes they harbour (i.e. *essC2* for N11-1255, *essC3* for FSL L7-0763 and *essC4* for CFSAN026581).

We also noted that some *essC1* strains only encode part of the *essC2* and *essC4* conserved modules (Fig. 6b). For example, strain PNUSAL000361 harbours genes 24–28 of the *essC2* module but lacks genes 21–23, and strain 25B09 encodes the final four genes [44–47] of the *essC4* conserved cluster, including the shorter version of gene 44 (Fig. S2). In these instances, no truncated EssC is encoded at the gene cluster. These findings suggest that the first three genes at these conserved clusters may be involved in mediating interaction with the variable region of the cognate EssC, potentially to target substrates or to promote functional assembly of the secretion system.

### An LXG toxin is encoded at variable region 2 of many strains

The second variable region lies between the tRNA gene *lmot01* and *lm0082*, which encodes a small membrane protein of unknown function (Fig. 3a). We noted that while approximately 25% of strains have no ‘insertion’ in this region, many other strains have a small locus of two or three genes relating to the T7SS (Figs 7 and S2). Commonly the locus comprises a gene encoding an LXG toxin, along with one or two genes encoding probable immunity proteins (Fig. 7). Four different toxins were identified in variable region 2 across strains (Table 3; toxins D–G; Fig. S5). All four toxins have a highly conserved N-terminal region – the first 330 or so amino acids are almost invariant, but the sequences diverge substantially in the toxin domain (Fig. S5).

A small number of strains (approximately 4% of those we analysed) have two genes at this locus encoding a protein of 262 amino acids with three predicted transmembrane domains at its N-terminus and a one of 207 amino acids with an N-terminal Tir domain (Fig. S2). We also observed that approximately 5% of strains had undergone a recombination event next to the tRNA gene to introduce a recombinase gene at this locus (Fig. S2).

### LXG proteins are encoded at other loci on the *L. monocytogenes* chromosome

To determine whether there may be LXG proteins encoded outside of the *L. monocytogenes* T7 gene cluster, we selected 14 strains (2 of each *essC* variant) from those listed in Fig. S2 and searched the genome annotations for the term ‘LXG’. This yielded 44 protein sequences containing this term, 17 of which we had not previously identified. We next entered the multiple protein accessions collectively into a single round of analysis using the FlaGs tool to see whether any of the encoding genes shared a common genomic location. The output (shown in Fig. S6 and summarized in Table 3) indicates that there are common genomic regions where toxins are likely to be encoded.

To ensure we had covered the diversity of LXG proteins in *L. monocytogenes*, we next searched RefSeq under the term ‘Listeria monocytogenes LXG protein’ which returned 569 protein accession numbers. Following manual inspection to see whether they differed from previously identified toxins, we found a further 10 unique protein sequences.
To identify the chromosomal loci where these additional LXG proteins were encoded, we used two accessions for each toxin that we identified that is encoded outside of the T7 locus and collectively submitted them for flanking gene analysis (shown in Fig. S7 and summarized in Table 3).

It is clear that there are preferred chromosomal locations where LXG proteins are encoded, shown on a map in Fig. 8. We noted that a cluster of ABC transporter genes (lm0135/lmo0152) is a hotspot for LXG protein-encoding genes, and across L. monocytogenes strains we detected...
11 different candidate toxins (toxins H–R in Table 3) encoded at that region. Seven unique LXG protein-encoding sequences were also identified at an insertion site between DNA-directed RNA polymerase- and transketolase (lmo0334/lmo0342)-encoding genes (toxins S–Y in Table 3). As far as we could determine, each individual LXG protein is encoded at the same chromosomal locus across the strains. Other hotspots were also identified, although these generally encode fewer unique sequences.

The loci share a number of common features; insertions generally encode a single LXG protein, with the exception of the insertion between the glutamine-hydrolysing GMP synthase and α/β hydrolase-encoding genes, where two LXG protein-encoding genes are arranged in a head-to-head organization (Fig. S7). Also encoded at the LXG loci are the likely cognate immunity proteins and often additional orphan immunities and toxin fragments. Interestingly, in almost all instances the LXG-encoding gene is preceded by a pair of genes encoding the same two small proteins (numbered 4 and 3 in Figs S6 and 1 and 2 in Fig. S7). Although neither of these is classified as being in the WXG100 superfamily, structural prediction suggests that they share the same WXG100 protein fold, and gene number 4 in Fig. S6/2 in Fig. S7 encodes a protein of the DUF3130/SACOL2603 family that has been linked to the T7SS by phylogenetic profiling.

It is striking to note that all of the LXG proteins encoded next to this gene pair share a similar LXG domain sequence (Fig. S8), although proteins encoded at the same loci are clearly more closely related. Whitney et al. [30] have previously shown that WXG proteins bind to their cognate LXG proteins to promote export by the T7SS. It is very likely that one or both of these small proteins plays a similar role, presumably interacting with the highly conserved N-terminal region. We also observed that three LXG proteins, toxins θ, μ and ϕ, are quite divergent in their N-terminal sequence (Fig. S8). In the case of toxins θ and ϕ, they are preceded by a different pair of small genes (346/349 and 319/324, respectively, in Fig. S7) that may encode cognate export partners. No candidate small genes immediately precede toxins σ and υ, and they therefore may rely on partner proteins encoded elsewhere in the genome.

Fig. 7. LXG toxins are often encoded in variable region 2. Three genetic arrangements of variable region 2. Genes encoding candidate toxins are shaded yellow with the cognate immunity gene in green. Probable orphan immunity proteins are shown in peach (with likely immunity indicated where identified). Gene sizes and intergenic regions not to scale.
Finally, when we analysed the \textit{bspA}/\textit{oxrB} (\textit{lmo0460}/\textit{lmo0476}) locus in the genome of EGDe where we had observed LXG protein-encoding genes in other strains, we observed that this strain encodes an unusual truncated LXG domain protein at this locus (Fig. S7). The protein, Lmo0473, lacks the first 130 amino acids of a canonical LXG domain protein and is preceded by a gene encoding a protein with a bacteriophage abortive infection AbiH domain. A second AbiH domain protein is also encoded nearby, and it appears that EGDe and a few other strains may have undergone a genomic rearrangement or insertion event at the locus. A manual search using the sequence of Lmo0473 revealed that several \textit{L. monocytogenes} strains encode an otherwise sequence identical protein at this locus, but it is longer and is a bona fide LXG protein (toxin $\alpha$ on the sequence alignment shown in Fig. S8). Curiously, these longer proteins are not found in the RefSeq database, which is used as input for the FlaGs program (and are therefore missing from Fig. S7), although there are several entries in the NCBI’s non-redundant protein sequence database. It is not clear whether the truncated variant encoded by EGDe and other strains would be a substrate for the T7SS or if it may use an alternative secretion pathway.

**DISCUSSION**

In this study we have undertaken a genomic survey of the T7SS of \textit{L. monocytogenes}. Our results show that there is exceptional diversity across strains in both the identity of the core component EssC and the likely secreted substrates.

We identified seven variants of EssC based on the sequence of the C-terminal 500–600 amino acids. The most common variant is EssC1, which is encoded in just over a quarter of sequenced \textit{L. monocytogenes} strains, including the reference strain EGDe. The genetic organization of \textit{essC1} to some extent mirrors that of \textit{S. aureus} \textit{essC1} strains, with genes encoding the secreted substrates EsxC and EsxB and the chaperone protein EsaE common between the two, and sharing conserved positioning. An toxin with an N-terminal YeeF domain is found in all \textit{L. monocytogenes} \textit{essC1} strains. In \textit{S. aureus} EsaE interacts with the YeeF domain of EsaD and with EssC1 and appears to target the toxin to the secretion machinery [34, 36]. It is likely that the YeeF domain toxins are similarly targeted in \textit{L. monocytogenes}. Since we failed to identify an EsaE-like protein encoded in any of the other \textit{essC} variants, we speculate that other substrate proteins, such as the LXG domain proteins, do not require an EsaE-like protein for targeting to the T7SS.

Strains encoding the EssC5, EssC6 and EssC7 variants are the least common. These strains share some similarity at the \textit{essC} locus; all three strains encode an LXG domain protein and likely immunity protein downstream of \textit{essC}. In each case the LXG protein-encoding gene is separated from \textit{essC} by two intervening genes. Based on the positioning of these genes and the fact that one or both encode small proteins, we speculate that they may be LXG partners required to mediate export.
For the other three essC variants the identity of substrate protein(s) encoded immediately downstream of essC is not obvious. Strains harbouring essC3 variants always encode the same two proteins downstream, the larger of which may represent a novel secretion substrate. The essC2 and essC4 variants have clusters of genes in the vicinity of essC that mainly encode proteins of unknown function. Some or all of these genes are frequently found at the essC locus of other variant strains. Often a smaller module of five genes (essC2) or four genes (essC4) is found across other strain variants, including some essC1 strains. Curiously, however, some other essC1 strains contain the full complement of ‘foreign’ genes – in this case the strains always adjacent encode an ‘orphan’ truncated EssC covering the C-terminal 728 amino acids, which shares the same sequence as the essC variant from where the genes were acquired. It is not clear whether these shorter EssC proteins (which lack transmembrane domains) are functional and can integrate into the T7SS. Alternatively, they may serve as a site of recombination with the full-length essC gene, which would splice out the intervening genes and convert an essC1 variant into an essC2/essC4 strain.

We identified a second variable locus at the T7 gene cluster that is present in most, but not all, strains. One of four different proteins, each sharing a highly conserved LXG domain, was found to be encoded at this region. Outside of the T7 cluster we also identified additional loci where LXG proteins are encoded. In total, across all strains we identified 40 different LXG proteins. Although none of these LXG proteins have yet been characterized, some of them could be structurally modelled on nuclease, NAD+ glycohydrolase and membrane-depolarising toxins. Taken together, our findings imply an important role for the T7SS secretion system in bacterial antagonism and kin discrimination in L. monocytogenes. We note that prior studies failed to demonstrate a function for the T7SS in virulence and host cell invasion, if anything finding that it had a negative impact on L. monocytogenes infection. Intriguingly, genes encoding some of the T7SS components and the YeeF domain toxin are upregulated when strain EGDe is used to infect germ-free mice that have been colonized with Lactobacillus [59], which would be consistent with a role in interspecies competition. Our findings provide a framework for the functional analysis of the T7SS across the species.

Funding information
This study was supported by the Wellcome Trust (through Investigator Award 10183/2/15/Z to T.P.).

Acknowledgements
We are grateful to Stephen Garrett (Newcastle University), Chayan Kumar Saha and Dr Gemma Atkinson (Umea University) for helpful advice with the FlaGs programme. We thank Dr Felicity Alcock, Dr Lisa Bowman and Stephen Garrett for critical reading of the manuscript.

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

References
1. Costa TRD, Felisberto-Rodrigues C, Meir A, Prevost MS, Redzej A et al. Secretion systems in gram-negative bacteria: structural and mechanistic insights. Nat Rev Microbiol 2015;13:343–359.
2. Gorasia DG, Veith PD, Reynolds EC. The type IX secretion system: advances in structure, function and organisation. Microorganisms 2020;8:1173.
3. Palmer T, Finney AJ, Saha CK, Atkinson GC, Sargent F. A holin/peptidoglycan hydrolase-dependent protein secretion system. Mol Microbiol 2021;in press.
4. Bunduc CM, Bitter W, Houben ENG. Structure and function of the mycobacterial type VII secretion systems. Annu Rev Microbiol 2020;74:315–335.
5. Klein TA, Ahmad S, Whitney JC. Contact-dependent interbacterial antagonism mediated by protein secretion machines. Trends Microbiol 2020;28:387–400.
6. Hsu T, Hingley-Wilson SM, Chen B, Chen M, Dai AZ et al. The primary mechanism of attenuation of Bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. Proc Natl Acad Sci U S A 2003;100:12420–12425.
7. Pym AS, Brodin P, Majiessi L, Brosch R, Demangel C et al. Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. Nat Med 2003;9:533–539.
8. Stanley SA, Raghavan S, Hwang WW, Cox JS. Acute infection and macrophage subversion by Mycobacterium tuberculosis require a specialized secretion system. Proc Natl Acad Sci U S A 2003;100:13001–13006.
9. Burts ML, Williams WA, DeBord K, Missiakas DM. EssXa and EssXb are secreted by an ESAT-6-like system that is required for the pathogenesis of Staphylococcus aureus infections. Proc Natl Acad Sci U S A 2005;102:1169–1174.
10. Burts ML, DeDent AC, Missiakas DM. Essa substrate for the ESAT-6 secretion pathway and its role in persistent infections of Staphylococcus aureus. Mol Microbiol 2008;69:736–746.
11. Abdallah AM, Gey van Pittius NC, Champion PAD, Cox J, Luirink J et al. Type VII secretion—mycobacteria show the way. Nat Rev Microbiol 2007;5:883–891.
12. Unnikrishnan M, Constantinidou C, Palmer T, Pallen MJ. The enigmatic Exs proteins: looking beyond mycobacteria. Trends Microbiol 2017;25:192–204.
13. Rosenberg OS, Dovala D, Li X, Connolly L, Bendebury A et al. Substrates control multimerization and activation of the multi-domain ATPase motor of type VII secretion. Cell 2015;161:501–512.
14. Zoltner M, Ng WMMA, Money JJ, Fyfe PK, Kneuper H et al. EssC: domain structures inform on the elusive translocation channel in the Type VII secretion system. Biochem J 2016;473:1941–1952.
15. Beckham FI, Czudnochowski N, Abdel-Aty H, Le L, Zoltner M et al. Structure of the mycobacterial ESX-5 type VII secretion system membrane complex by single-particle analysis. Nat Microbiol 2017;2:17047.
16. Famelis N, Rivera-Calzada A, Degliesposti G, Wingender M, Mietrach et al. Architecture of the mycobacterial type VII secretion system. Nature 2019;576:321–325.
17. Poweleit T, Czudnochowski N, Nakagawa R, Trinidad DD, Murphy KC et al. The structure of the endogenous ESX-3 secretion system. Elife 2019;8:e52983.
18. Ramsdell TL, Humpert LA, Sysoeva TA, Fortune SM, Burton BM. Linked domain architectures allow for specialization of function in the FtsK/SpoIIE ATPases of Ess secretion systems. J Mol Biol 2015;427:1119–1132.
19. Mietrach et al. From where the genes were acquired. It is not clear whether Strains harbouring essC variants always encode the same two proteins downstream, the larger of which may represent a novel secretion substrate. The essC2 and essC4 variants have clusters of genes in the vicinity of essC that mainly encode proteins of unknown function. Some or all of these genes are frequently found at the essC locus of other variant strains. Often a smaller module of five genes (essC2) or four genes (essC4) is found across other strain variants, including some essC1 strains. Curiously, however, some other essC1 strains contain the full complement of ‘foreign’ genes – in this case the strains always adjacent encode an ‘orphan’ truncated EssC covering the C-terminal 728 amino acids, which shares the same sequence as the essC variant from where the genes were acquired. It is not clear whether these shorter EssC proteins (which lack transmembrane domains) are functional and can integrate into the T7SS. Alternatively, they may serve as a site of recombination with the full-length essC gene, which would splice out the intervening genes and convert an essC1 variant into an essC2/essC4 strain.

We identified a second variable locus at the T7 gene cluster that is present in most, but not all, strains. One of four different proteins, each sharing a highly conserved LXG domain, was found to be encoded at this region. Outside of the T7 cluster we also identified additional loci where LXG proteins are encoded. In total, across all strains we identified 40 different LXG proteins. Although none of these LXG proteins have yet been characterized, some of them could be structurally modelled on nuclease, NAD+ glycohydrolase and membrane-depolarising toxins. Taken together, our findings imply an important role for the T7SS secretion system in bacterial antagonism and kin discrimination in L. monocytogenes. We note that prior studies failed to demonstrate a function for the T7SS in virulence and host cell invasion, if anything finding that it had a negative impact on L. monocytogenes infection. Intriguingly, genes encoding some of the T7SS components and the YeeF domain toxin are upregulated when strain EGDe is used to infect germ-free mice that have been colonized with Lactobacillus [59], which would be consistent with a role in interspecies competition. Our findings provide a framework for the functional analysis of the T7SS across the species.

Funding information
This study was supported by the Wellcome Trust (through Investigator Award 10183/2/15/Z to T.P.).

Acknowledgements
We are grateful to Stephen Garrett (Newcastle University), Chayan Kumar Saha and Dr Gemma Atkinson (Umea University) for helpful advice with the FlaGs programme. We thank Dr Felicity Alcock, Dr Lisa Bowman and Stephen Garrett for critical reading of the manuscript.

Conflicts of interest
The authors declare that there are no conflicts of interest.
21. Jäger F, Kneuper H, Palmer T. EssC is a specificity determinant for *Staphylococcus aureus* type VII secretion. *Microbiology* 2018;164:816–820.

22. Champion PAD, Stanley SA, Champion MM, Brown EJ, Cox JS. C-terminal signal sequence promotes virulence factor secretion in *Mycobacterium tuberculosis*. *Science* 2006;313:1632–1636.

23. Anderson M, Aly KA, Chen Y-H, Missiakas D. Secretion of atypical protein substrates by the ESAT-6 secretion system of *Staphylococcus aureus*. *Mol Microbiol* 2013;90:734–743.

24. Sysoeva TA, Zepeda-Rivera MA, Huppert LA, Burton BM. Dimer recognition and secretion by the ESX secretion system in *Bacillus subtilis*. *Proc Natl Acad Sci U S A* 2014;111:7653–7658.

25. Kneuper H, Cao ZP, Twomey KB, Zottl M, Jäger F et al. Heterogeneity in eis transcriptional organization and variable contribution of the Ess/Type VII protein secretion system to virulence across closely related *Staphylococcus aureus* strains. *Mol Microbiol* 2014;93:928–943.

26. Jäger F, Zottl M, Kneuper H, Hunter WN, Palmer T. Membrane interactions and self-association of components of the Ess/Type VII secretion system of *Staphylococcus aureus*. *FEBS Lett* 2016;590:349–357.

27. Casabona MG, Buchanan G, Zottl M, Harkins CP, Holden MTG et al. Functional analysis of the EssB component of the *Staphylococcus aureus* type VII secretion system. *Microbiology* 2017;163:1851–1863.

28. Huppert LA, Ramsdell TL, Chase MR, Sarracino DA, Fortune SM et al. The ESX system in *Bacillus subtilis* mediates protein secretion. *PLoS One* 2014;9:e96267.

29. Baptista C, Barreto HC, Sá-José C. High levels of DegU-P activate an Esat-6-like secretion system in *Bacillus subtilis*. *PLoS One* 2013;8:e67840.

30. Whitney JC, Peterson SB, Kim J, Pazos M, Verster AJ et al. A broadly distributed toxin family mediates contact-dependent antagonism between gram-positive bacteria. *Elife* 2017;6:e26938.

31. Dai Y, Wang Y, Liu Q, Gao Q, Lu H et al. A novel ESAT-6 secretion system-secreted protein EsxH of community-associated *Staphylococcus aureus* lineage ST398 contributes to immune evasion and virulence. *Front Microbiol* 2017;8:819.

32. Way SS, Wilson CB. The *Mycobacterium tuberculosis* ESAT-6 homologue in *Listeria monocytogenes* is dispensable for growth in vitro and in vivo. * Infect Immun* 2005;73:6151–6153.

33. Warne B, Harkins CP, Harris SR, Vatsiou A, Stanley-Wall N et al. The Ess/Type VII secretion system of *Staphylococcus aureus* shows unexpected genetic diversity. *BMC Genomics* 2016;17:22.

34. Cao Z, Casabona MG, Kneuper H, Chalmers JD, Palmer T. The type VII secretion system of *Staphylococcus aureus* secretes a nuclease toxin that targets competitor bacteria. *Nat Microbiol* 2016;2:1618.

35. Ohr RJ, Anderson M, Shi M, Schnewind O, Missiakas D. EssD, a nuclease effector of the *Staphylococcus aureus* ESS pathway. *J Bacteriol* 2017;199:e00528–16.

36. Anderson M, Ohr RJ, Aly KA, Nocadello S, Kim HK et al. EssE Promotes *Staphylococcus aureus* ESS-dependent protein secretion to modify host immune responses during infection. *J Bacteriol* 2017;199:e00527–16.

37. Ulhuq FR, Gomes MC, Duggan GM, Guo M, Mendonca C et al. A membrane-depolarizing toxin substrate of the *Staphylococcus aureus* type VII secretion system mediates interspecies competition. *Proc Natl Acad Sci U S A* 2020;117:20836–20847.

38. Chatterjee A, Willett JLE, Dunny GM, Duerkop BA. Phage infection and sub-lethal antibiotic exposure mediate *Enterococcus faecalis* type VII secretion system dependent inhibition of bystander bacteria. *PLoS Genet* 2021;17:e1009204.

39. Tassinari M, Doan T, Bellinzoni M, Chabalier M, Ben-Assaya M. Central role and structure of the membrane pseudokinase YykC in the antibacterial *Bacillus subtilis* type VIII secretion system. *bioRxiv* 2020:05.09.085852.

40. Radoshevich L, Cossart P. *Listeria monocytogenes*: towards a complete picture of its physiology and pathogenesis. *Nat Rev Microbiol* 2018;16:32–46.

41. Pinheiro J, Reis O, Vieira A, Moura IM, Zanolli Moreno L et al. *Listeria monocytogenes* encodes a functional ESX-1 secretion system whose expression is detrimental to in vivo infection. *Virulence* 2017;8:993–1004.

42. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402.

43. Madeira F, Park YM, Lee J, Buso N, Gur T et al. The EMBL-EBI search and sequence analysis tools Apis in 2019. *Nucleic Acids Res* 2019;47:W636–W641.

44. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009;25:1189–1191.

45. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 2018;35:1547–1549.

46. Stecher G, Tamura K, Kumar S. Molecular evolutionary genetics analysis (MEGA) for macOS. *Mol Biol Evol* 2020;37:1237–1239.

47. Bécavin C, Koutero M, Tichtchek N, Cerutti F, Lechat P et al. Listeriomics: an interactive web platform for systems biology of *Listeria*. *mSystems* 2017;2:e00186–16.

48. Solovyev V, Salamov A. Automatic annotation of microbial genomes and metagenomic sequences. In: RW Li (editor). *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies*. Nova Science Publishers; 2011. pp. 61–78.

49. Macke TJ, Ecker DJ, Gutell RR, Gautheret D, Case DA et al. RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Res* 2001;29:4724–4735.

50. Saha CK, Pires RS, Brolin H, Delannoy M, Atkinson GC. Flags and webFlaGs: discovering novel biology through the analysis of gene neighbourhood conservation. *Bioinformatics* 2020:in press.

51. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 2001;305:567–580.

52. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 2015;10:845–858.

53. Bobrovskyy M, Willing SE, Schneewind O, Missiakas D. EssH peptidoglycan hydrolase enables *Staphylococcus aureus* Type VII secretion across the bacterial cell wall envelope. *J Bacteriol* 2018;200:e00268–18.

54. Cao Z. Characterisation of the Type VII secretion system of *Staphylococcus aureus*: PhD thesis. University of Dundee 2017.

55. Orsi RH, den Bakker HC, Wiedmann M. *Listeria monocytogenes* lineages: genomics, evolution, ecology, and phenotypic characteristics. *Int J Med Microbiol* 2011;301:79–96.

56. Ragon M, Wirth T, Hollandt F, Lavenir R, Lecuit M et al. A new perspective on *Listeria monocytogenes* evolution. *PLoS Pathog* 2008;4:e1000146.

57. Kuenne C, Billon A, Mrheil MA, Strittmatter A, Daniel R et al. Reassessment of the *Listeria monocytogenes* pan-genome reveals dynamic integration hotspots and mobile genetic elements as major components of the accessory genome. *BMC Genomics* 2013;14:47.

58. Rychli K, Wagner EM, Ciocaciu LC, Zaiser A, Tasara T et al. Comparative genomics of human and non-human *Listeria monocytogenes* sequence type 121 strains. *PLoS One* 2017;12:e0176857.

59. Archambaud C, Nahori M-A, Soubigou G, Bécavin C, Laval L et al. Impact of lactobacilli on orally acquired listeriosis. *Proc Natl Acad Sci U S A* 2012;109:16684–16689.

Edited by: R. Manganelli and P. Brodin