Genome-directed study uncovers a family of phospholipases and highlights the diversity of *Salmonella* T6SS effectors.

Gianlucca G. Nicastro[2,3,¥], Stephanie Sibinelli-Sousa[2,¥], Julia Takuno Hespanhol[1], Thomas W. C. Santos[2], Joseph P. Munoz[1], Rosangela S. Santos[4], André L. Araujo-Silva[2], Blanca M. Perez-Sepulveda[5], Sayuri Miyamoto[4], L. Aravind[3], Robson F. de Souza[2,*], Ethel Bayer-Santos[1,2,6*].

1. Department of Molecular Biosciences, College of Natural Sciences, The University of Texas at Austin, Austin, USA.

2. Departamento de Microbiologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, Brazil.

3. Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, USA.

[4] Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil.

[5] Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, UK.

[6] LaMontagne Center for Infectious Disease, The University of Texas at Austin, Austin, Texas, USA.

[¥]These authors contributed equally to this work.

*Corresponding authors: ebayersantos@austin.utexas.edu; rfsouza@usp.br

Keywords: *Salmonella*, type VI secretion system, toxins, effectors, phospholipase, acyltransferase.

Abstract

Bacteria live in complex communities and use antagonistic strategies to compete for resources. Gram-negatives encode a type VI secretion system (T6SS) that delivers toxins into eukaryotic and prokaryotic cells. Previous studies highlighted the importance of *Salmonella* T6SSs in bacterial antagonism and host cell infection, but few antibacterial effectors have been described. Here, we used bioinformatic strategies to identify the T6SS effectors in 10,000 *Salmonella* genomes. As a result, we identified and manually curated 128 candidates. Specific *Salmonella* serovars harbor unique effectors that are target-specific (antibacterial or anti-eukaryotic), depending on the T6SS subtype (e.g. i1 or i3). In addition, 45 previously uncharacterized toxin domains were identified and designated STox (*Salmonella* toxins). STox15 was among the most frequent candidates and was selected for in-depth characterization. STox15 is an antibacterial effector belonging to the NlpC/P60 papain-like fold superfamily with a permuted catalytic core typical of lipid-targeting versions rather than peptidases or amidases. Biochemical analysis with recombinant protein and lipidomics of intoxicated *Escherichia coli* revealed that STox15 displays phospholipase activity cleaving off acyl groups from phosphatidylglycerol and phosphatidylethanolamine. This work improves our understanding of the *Salmonella* T6SS effector repertoire and provides the first direct characterization of a lipid-targeting NlpC/P60 domain in biological conflicts.

Introduction

Competition is a fundamental biological process in nature, occurring both within and between species that share a common environment. Bacteria actively participate in these ecological battles employing a potent arsenal of toxins as their weaponry[1]. A prominent mechanism for toxin delivery among Gram-negative bacteria is via the type 6 secretion system (T6SS)[2]. The T6SS is a contractile apparatus composed of 13 structural components (TssA-TssM), which form three main complexes: membrane, baseplate, and tail[3,4]. Phylogenetic analysis of T6SS components showed that there are four types of T6SSs (T6SS^{i-iv})[5-7], with the canonical T6SS^{i} present in *Proteobacteria* being further classified into six subtypes (i1, i2, i3, i4a, i4b and i5)[6,8,9]. The T6SS functions in a contact-dependent manner and rely on the biochemical properties of secreted effector toxins for its function[10]. Effectors are loaded onto a spear-like structure formed by hexameric rings of Hcp proteins capped by a spike comprising a trimer of VgrG sharpened by a PAAR protein[11]. Effectors are translocated into target cells either fused at the C-terminus of Hcp, VgrG, and PAAR proteins (named specialized effectors), or associated with these proteins via adaptors (cargo effectors)[12]. Antibacterial toxins are often paired with immunity proteins that prevent self-intoxication, thus forming gene pairs that are frequently located near structural components of the T6SS[10,13].

The protective role of the endogenous microbiota against *Salmonella* infection has been recognized for years[14]; however, only recently have studies started to reveal the mechanism by which the microbiota maintains gut homeostasis and promotes colonization resistance[15]. Despite this understanding, there is a relative scarcity of information regarding the direct antimicrobial mechanisms employed by commensals and pathogens during these disputes for niche control[16]. T6SSs clusters are conserved across several *Salmonella* spp., highlighting their importance to the fitness of these bacteria[17,18].

Salmonella spp. encode T6SSs in different pathogenicity islands (SPI): SPI-6, SPI-19, SPI-20, SPI-21, and SPI-22, which were acquired by distinct horizontal transfer events[17,18]. *S.* Typhimurium encodes a T6SS subtype i3 within SPI-6, which is important for interbacterial competition and gut colonization in mice[19,20]. *S.* Gallinarum encodes a T6SS subtype i1 within SPI-19, which is important for survival in macrophages [17,21,22] and colonization of birds[23]. *S. bongori* encodes a T6SS subtype i4a with antibacterial activity within SPI-22[24]. Despite this information, little is known about the identity or mechanism of the T6SS effector repertoire that contributes to these phenotypes.

Here, we set out to identify the repertoire of T6SS effectors in a dataset of 10,000 *Salmonella* isolates utilizing a computational screening, followed by careful manual curation. Employing a blend of sensitive sequence and structure searches alongside comparative genomics, we identified 128 putative effectors, including 45 divergent or new toxin domains which were not recognized in publicly available domain databases. This comprehensive analysis strongly indicates that T6SSs subtypes i3 and i1 are associated with antibacterial and anti-eukaryotic effectors, respectively. Furthermore, our findings reveal that each of the 149 *Salmonella* serovars contained in the 10KSG dataset harbors a unique combination of effectors, likely required for their specific ecological interactions. A detailed examination of a selected candidate encoding a newly identified domain (STox15) showed that it is an antibacterial effector belonging to the NlpC/P60 superfamily with a permuted catalytic core. STox15 displays phospholipase activity *in vitro* and its ectopic expression in target cells modifies the membrane composition. This study offers a comprehensive characterization of the toxin arsenal linked to T6SSs in *Salmonella*, identifying novel effector families and providing an in-depth analysis of a new family that specifically targets lipids.

Results

Identification of the T6SS effector repertoire within 10,000 *Salmonella* genomes

T6SS effectors are often encoded in close genomic proximity to structural components of the system[17,24,25]. To identify new T6SS effectors, we employed a "guilt-by-association" approach, which relies on the conservation of genomic context and is a powerful strategy for pinpointing new effectors[26]. We applied this methodology to a well-known dataset of *Salmonella* genomes[27] (Fig. 1A). This dataset comprises over 10,000 sequenced genomes, from which 6,165 assemblies passed quality filters and were included in our analyses. The genome annotations provided by the 10KSG project were used to identify all coding sequences and hidden Markov models (HMMs) were used to detect T6SS-related proteins. HMM search results were compiled and organized into a tabular format, resulting in a comprehensive dataset with more than 27,896,579 protein entries. The HMM collection used in our search includes those specifically designed for secretion systems[28] and toxin detection[29], as well as broader-spectrum models from the Pfam and CDD databases[30,31]. Redundancy among proteins was eliminated via clustering with MMseqs[32], with stringent parameters of 100% identity and coverage. This process yielded 544,319 non-redundant proteins, which underwent additional clustering with a coverage threshold of 80% and a maximum e-value of 1e-3, thus generating 49,080 groups.

Next, we collected 10 genes upstream and downstream of each locus encoding proteins annotated as T6SS components, constituting what we referred to as genomic sites (Fig. 1A). Overlapping genomic sites featuring multiple genes encoding T6SS components were merged. The majority of genomic sites displayed a consistent pattern, with 10 genes before and after the first and last genes recovered by the T6SS models in

each site. Exceptions occurred when sites were near the edge of a contig, which hindered the identification of the complete set of genes downstream and/or upstream. Our analysis identified a total of 42,560 genomic sites housing at least one T6SS component. Subsequently, we applied a graph theoretic strategy leveraging the Jaccard index for network construction[33], followed by the Louvain algorithm[34] for automatically categorize genomic sites. Next, a network was constructed with genomic site identifiers as nodes and the Jaccard index values as edges (indexes lower than 0.33 were disregarded). The resulting network was subjected to analysis by the Louvain algorithm, which identified 46 distinct communities (Fig. 1A). Each of these 46 communities was manually inspected to assess the frequency and phylogenetic information of each T6SS component, resulting in the classification of all the genomic sites into five distinct neighborhood categories: T6SS subtype i1, subtype i2, subtype i3, subtype i4b, and orphan (Fig. 1A).

Our analysis revealed that a notable portion of the genomes sequenced in the 10KSG harbor the T6SS subtype i3 (Fig. 1B, C). This system is commonly situated within the *Salmonella* Pathogenicity Island-6 (SPI-6) and has been shown to play a role in interbacterial competition[19,20,25]. The second most abundant category was T6SS subtype i1, typically located within the SPI-19 (Fig. 1B, C). This system has been associated with infection of eukaryotic cells[21,22]. Conversely, the subtype i2 is primarily associated with SPI-21[18], and is less frequently encountered within the 10KSG dataset (Fig. 1B, C). In addition, our analysis uncovered a substantial number of orphan genes characterized by isolated components of the T6SS inner spear, such as Hcp, VgrG and PAAR (Fig. 1A, 1B, 1C).

Another interesting observation is the mobility of the distinct phylogenetic subtypes of T6SSs among genomic sites and/or pathogenicity islands, depending on the strain and/or serovar. Using genes encoding tRNAs as markers to identify the T6SS

insertion sites in *Salmonella*[35], we confirmed previous reports that the majority of subtype i3 (90%, 4929/5461 genomic sites) are located within SPI-6 and are flanked by the gene encoding aspartate tRNA ($tRNA^{Asp}$)[17]. Contrary to previous reports, the genomic location of subtype i1 is not well conserved. The system can be found in several places, including within the SPI-19 associated with $tRNA^{Ala}$ (12%, 66/532 genomic sites) as previously described[17], and within the SPI-6 (3%, 18/532 genomes) flanked by $tRNA^{Asp}$. However, most subtype i1 clusters could not be assigned to a specific position as their contigs do not contain a tRNA gene to be used as reference (82%, 437/532 genomic sites). Overall, these results indicate that distinct T6SSs subtypes are inserted into different genomic positions that are hot spots for horizontal transfer events, and the combination between the insertion site and the introduced T6SS subtype varies according to the strain and serovar under study.

Next, we focused on identifying effectors using three strategies (Fig. 1D). First, we located proteins containing N-terminal PAAR, VgrG and Hcp domains and additional C-terminal domains with more than 50 amino acids. These C-terminal regions were isolated and grouped (80% coverage and 1e-3 e-value). Second, we used the genomic sites containing T6SS components and analyzed up to five genes upstream and downstream via the software BastionX[36]. Third, we used amino acid sequences, HMMs and PSSMs (position-specific scoring matrix) from SecReT6[37], Pfam[30], and Zhang, et al. [29] to search the 10KSG dataset for previously described effectors. For candidates not recognized by previously described models, we collected homologs from NCBI using 4 iterations of PSI-BLAST[38] and generated multiple sequence alignments and HMMs. We then used a series of strategies to classify and annotate the function of these candidates, comprising both sequence and structural approaches. We used profile-profile comparison methods such as HHsearch[39] to detect distant homologs; and created structural models

using Alphafold2[40] from multiple sequence alignments of each candidate to perform searches in FoldSeek[41] and DALI[42] against the PDB[43] and AlphaFoldDB[44]. The structure-structure comparison algorithm from FoldSeek was used to cluster groups of candidates. All the sequence and structure information were manually examined to establish the final domain identification (Fig. 1D). Candidates that displayed sequence or structural similarity to known toxin domains or proteins of unknown function that displayed conserved genomic organization and/or adjacent conserved putative immunity proteins across several species were maintained. The remaining candidates were considered false positives or putative T6SS components (e.g. adaptors) and were removed from the analysis.

In total, we identified 128 groups of effectors in the 10KSG dataset (Table S1 and Fig. 1E). Eighty-three groups were already described in public databases (e.g. Ntox47)[29,30], or represent individual effectors that were experimentally characterized but for which HMMs have not been produced and made available (e.g. TreTu)[24,45-60]. For the latter, the newly created HMMs received a ".st" in their names (e.g. TreTu.st). In addition to the 83 effectors, 45 groups comprised new toxin domains or highly divergent variations of known toxin domains that were not detected by previously published HMMs. This justified them being distinct groups requiring the design of new HMMs. These groups of effectors were named STox (*Salmonella* toxin) followed by a number (e.g. STox_1) (Table S1 and Fig. 1E) - the numbering is not sequential due to the renaming of certain STox models to match the name of effectors already described in the literature. In addition, while this manuscript was under preparation a few STox models described in this study were also detected in recently published surveys of T6SS effectors and polymorphic toxins[61-64] - this information is depicted in Table S1. The amino acid

sequence alignments and AlphaFold structural prediction of each group of effectors are available at https://leepbioinfo.github.io/10ksgt6ss/.

To test whether the toxin HMMs detected in this study hit any of the previously published models, we used those models as queries in HMMSEARCH runs to collect protein sequences from a large non-redundant protein database. Then we annotated those collected proteins with public models from Pfam or Zhang, et al. [29] (Fig. 1F). The model that retrieved more hits was used as a reference for further analysis of the 10KSG dataset. This comparison revealed that 23 STox models recovered populations of proteins that were not recognized by any of the existing reference models (Fig. 1F, yellow circles), confirming that they are indeed new toxin domains or highly divergent versions of known superfamilies. Analysis of hits recovered by STox models and also recognized by reference models revealed that most of the STox models targeted considerably divergent versions of previously existing families (with less than 25% of members being recognized by reference models), thus warranting a new nomenclature. Several of the STox models sharing between 25-92% of their hits with preexisting reference models were still able to recover hundreds of new hits (Fig. 1F). Although such models do not represent novel toxin domains, some correspond to new families within known superfamilies and their use ensured higher recovery of effectors within the 10KSG dataset.

Inspection of genomic context across several bacterial species revealed that most of the 128 identified effectors exhibited a conserved adjacent gene coding for a predicted cognate immunity protein, thus suggesting that these effectors have antibacterial activity (83.6%, 107/128) (Table S1 and Fig. 1E). Some effectors, which lacked a conserved immunity protein, were predicted to display anti-eukaryotic activity (12.4%, 14/128). Additionally, a small proportion of effectors were predicted to target both cell types due to the sporadic presence of a putative immunity protein associated with it (4%, 4/128)

(Table S1 and Fig. 1E). Our analysis revealed a diverse array of cellular targets and biochemical activities among the 128 effector groups, including peptidases/proteases, glycosidases, phospholipases/acyltransferases, ART superfamily enzymes (ADP-ribosyltransferases and NADases), pore-forming toxins, adenylyl cyclases, RNases, DNases, and Zn-dependent nucleobase deaminases (Table S1 and Fig. 1E). Notably, the activities of a few STox effectors could not be predicted confidently and will require further analysis (Fig. 1E). Overall, these findings highlight the significant diversity of T6SS effectors encoded by *Salmonella* and reveal an array of putative new toxins used in interbacterial competition and as host virulence factors.

***Salmonella* serovars encode diverse sets of T6SS effectors with target-specific activities**

After classifying the different T6SS subtypes and carefully identifying the effector repertoire within the 10KSG dataset, we calculated the average number of effectors per genome. Our findings reveal a considerably higher number of genomes with only one T6SS cluster (4,334 genomes) compared to those with two clusters (240 genomes) within the dataset. Notably, genomes harboring multiple T6SS clusters tend to encode more effectors. We estimated that genomes with a single T6SS cluster - whether subtype i1, i2, or i3 – display an average of 3.57 ± 1.42 effectors per genome (Fig. 2A, blue). In contrast, genomes with two T6SS clusters - the most common combination was subtypes i1 and i3 - showed 5.35 ± 2.42 effectors per genome (Fig. 2A, orange). These data suggest an additive effect for the antibacterial subtype i3 and the anti-eukaryotic subtype i1 clusters, which together enhance the overall effector repertoire and contribute to bacterial fitness. It is worth highlighting that this estimated average is influenced by the characteristics of the 10KSG dataset used in the analysis, which is mainly composed of *S.* Typhimurium

(2,358 genomes) and *S.* Enteritidis (1,399 genomes)[27]. In addition, our manual curation was very stringent in maintaining only the candidates that were confidently described as effectors; thus, it is possible that these numbers are an underestimate.

The most frequent effectors detected within the 10KSG dataset were the peptidoglycan-targeting effectors Tlde1, a L,D-transpeptidase (4,988 genomes)[25], and Tae4, a papain-like amidase (3,987 genomes)[45] (Fig. 2B). These effector families were followed by Ntox47, a predicted RNase with the BECR fold (1,895 genomes)[65]; the metallopeptidase Tox-HopH1 (1,610 genomes); a second more divergent group of Ntox47 (Ntox47.st2) (1,552 genomes); the ADP-ribosyltransferases TreTu (1,224 genomes)[46] and STox_62 (906 genomes); and another peptidoglycan-targeting effector Tae2 (742 genomes)[45] (Fig. 2B). Together, these eight effectors were found in most genomes within the 10KSG and constitute the core T6SS effectors in *S. enterica*. Other effectors (ColE7, STox_15, Tox-WHH, STox_18, Tox-ART-HYD3, etc.) are encoded in less than 10% of genomes and may be considered accessory effectors. In addition, by looking at the combination of effectors in each genome, we found that 88% of the genomes within the 10KSG dataset were estimated to encode up to 4 effectors per genome (mode = 3), while 12% encode combinations that range between 5-18 effectors (mode = 5) (Fig. S1).

After identifying the most abundant effectors in the 10KSG dataset, we decided to determine the core 5 most frequent effectors detected in each of the 149 *Salmonella* serovars within the dataset (Fig. 1C and Fig. S2). Serovars that predominantly encode subtype i3, such as *S.* Typhimurium. *S.* Panama, *S.* Infantis, and *S.* Typhi, harbor effectors targeting peptidoglycan (e.g. Tae4, Tlde1, Tae2), nucleases (e.g. Ntox47, Tox-WHH, ColE7, Tox-AHH) and ART superfamily enzymes (e.g. TreTu, STox_62). *S.* Dublin, which contains both subtypes i1 and i3, displays a mixture of effectors targeting the

peptidoglycan (Tae2 and Tlde1), a nuclease (Ntox47), and a metallopeptidase (Tox-HopH1). *S.* Agona encodes only a T6SS subtype i1 but contains effectors with nuclease (ColE7 and Ntox19), pore-forming (STox_47), and undetermined activities (STox_18 and STox_53) (Fig. 2C). The core effectors of each serovar can be found in Fig. S2.

Additionally, we analyzed the combination of effectors most frequently encoded in genomes containing phylogenetically distinct T6SSs. Our findings revealed that the antibacterial subtype i3 predominantly displays a combination of peptidoglycan-targeting effectors (e.g. Tae4 and Tlde1); nucleases (e.g. Ntox47 and Tox-WHH); and ART family enzymes (e.g. TreTu and STox_62) (Fig. 2D). Conversely, genomes encoding the anti-eukaryotic subtype i1 lack peptidoglycan-targeting effectors, and instead show a pattern combining nucleases (e.g. ColE7) with pore-forming toxins (e.g. Ssp6) and effectors with undetermined activity (Fig. 2D). For genomes encoding both subtypes i1 and i3, there was a mixture of effectors with antibacterial (e.g. Tae2, Tlde1, Ntox47) and anti-eukaryotic activity (e.g. Tox-HopH1) (Fig. 2D). This data agrees with published literature[19-22,25] and supports the classification of *Salmonella* T6SS subtype i3 as antibacterial and subtype i1 as anti-eukaryotic. These results become even more evident when the effectors encoded by each T6SS subtype are grouped according to their functional categories (Fig. 2E).

Notably, we found that effectors encoded within or close to T6SS structural clusters, show a minimum overlap between subtypes i1, i2, and i3 (Fig. 2F). Of a total of 128 distinct effectors identified in the 10KSG dataset, 54 were found in subtype i3; 16 were detected within subtype i1; and 10 effectors were encoded within subtype i2. Finally, 48 effectors were exclusively found outside of the T6SS clusters (orphan) (Fig. 2F). Overall, this data shows that the effector subset is not only specific to the target, but it is also specific to the phylogenetic subtype of the T6SS (Fig. 2F). These data suggest

an evolutionary scenario where *Salmonella* independently acquired each T6SS cluster along with its associated set of effectors, and/or that a certain subset of effectors is preferably exchanged among bacteria harboring a particular T6SS subtype.

Previous analysis of T6SS subtype i3 identified the insertion of three variable regions (VR1-3) between the structural genes in which the effectors are encoded[17] (Fig. 2G). For subtype i3, VR1 and VR2 contain peptidoglycan-targeting and effectors with phospholipase/acyltransferase activity (e.g. Tae2, Tae4, Tlde1, STox_15). In contrast, VR3 is situated at the edge of the structural cluster and houses effectors with nuclease (e.g. Ntox47, Tox-WHH, STox_59) and ART (e.g. TreTu) activities, both typically associated with an N-terminal PAAR domain and Rhs (Rearrangement hotspot) repeats, which assist in the translocation across the inner bacterial membrane to reach the cytoplasmic targets[46,65-67] (Fig. 2G). Similarly, subtype i2 has three variable regions, with VR1 and VR2 displaying peptidoglycan-targeting effectors (e.g. STox_22 and STox_56), while VR3 contains nucleases (e.g. ColE7) (Fig. 2G). In contrast, subtype i1 holds just one variable region at the edge of the cluster, which contains nucleases (e.g. Tox-Rease-5.1, TseV), phospholipases (e.g. Tle1, an α/β hydrolase), peptidases (e.g. Tox-PeptidaseM23) and effectors with undetermined activities (e.g. STox_18) (Fig. 2G). The position of effectors at the edge of the T6SS cluster and the domain architecture containing Rhs repeats probably facilitate recombination events[65], leading to the greater diversity of effectors observed at the VR3 (Fig. S1).

STox15 is an antibacterial effector with a permuted NlpC/P60 domain and is evolutionarily related to enzymes involved in lipid biochemistry

Our comprehensive *in silico* analysis identified 45 groups of effector domains that were not recognized by any of the previously published HMMs. These domains were

classified as new or divergent families of toxins, termed STox (Table S1). Among these, STox15 emerged as one of the most abundant in the 10KSG dataset that was not previously studied (Fig. 2B). Interestingly, we observed that strains encoding STox15 typically lacked Tlde1 (Fig. 2D), and that both STox15 and Tlde1 are encoded within the VR2 of subtype i3 (Fig. 3A)[25]. Based on these observations, we chose STox15 for in-depth mechanistic characterization.

To analyze whether STox15 and its associated downstream gene (SImm15) comprise an effector and immunity pair, we cloned these genes in compatible vectors under the control of different promoters and evaluated the toxicity upon expression in *Escherichia coli*. This experiment revealed that STox15 is toxic in the periplasm but not in the cytoplasm of *E. coli*, and co-expression with SImm15 can neutralize its toxic effect (Fig. 3B). Next, we performed time-lapse microscopy to evaluate bacterial growth and morphology at the single cell level. *E. coli* carrying the plasmid with SP-STox15 grew normally in d-glucose (repressed conditions) (Movie S1). However, shortly after induction of SP-STox15 with l-arabinose, cells started lysing (Movie S2). It was curious that cells lysed without losing their rod shape, which suggests that the peptidoglycan structure was not being affected (Fig. 3C). Indeed, after cells lysis there were residual structures resembling the intact peptidoglycan sacculus (Fig. 3C), thus suggesting that although STox15 is toxic in the periplasm of *E. coli* the peptidoglycan is not the target.

Although the STox15 domain was not identified by any of the HMMs deployed in the initial steps of this study, subsequent HHpred analysis revealed a significant probability of homology with the DUF4105 and the T6SS effector TseH from *Vibrio cholerae*[51] (data not shown), both of which are members of the NlpC/P60 superfamily[68,69]. This superfamily was previously defined as comprising four families, which are divided into two higher-order groups (canonical and permuted)[68]. Members of

the canonical group (AcmB-like and P60-like) function as peptidases involved in peptidoglycan hydrolysis[70]. However, permuted members of this superfamily (YaeF-like and LRAT-like) exhibit a circular permutation in their catalytic core, which creates a hydrophobic binding pocket and provides specificity for lipid substrates[71,72]. Closer inspection of the sequence and structure of STox15 revealed a circular permutation of the catalytic domain indicating that it belongs to the permuted NlpC/P60 group (https://leepbioinfo.github.io/10ksgt6ss/)[68].

To more precisely predict the function of STox15, we sought to understand its evolutionary relationship by constructing a phylogenetic tree using the protein sequences from STox15, TseH and other previously defined permuted clades of the NlpC/P60 superfamily, such as LRAT and YiiX. The analysis revealed the formation of five major clades including YiiX, LRAT, TseH, STox15 and DUF4105 (Fig. 3D, Table S2 and Table S3). Genomic context analysis predicted that proteins of the STox15 clade are toxin domains of effectors deployed in biological conflicts (Fig. 3D and 3E). The TseH clade (Pfam DUF6695) similarly displays genomic contexts indicative of biological conflicts; however, their clear separation in the phylogenetic tree supports a distinction between STox15 and TseH (Fig. 3D and 3E). The clades LRAT and YiiX harbor groups of proteins known to be involved in lipid metabolism. LRAT (lecithin:retinol acyltransferase) is an enzyme present in mammals and involved in the transference of acyl groups from phosphatidylcholine to all-trans retinol to produce all-trans retinyl esters that are storage forms of vitamin A[73]. H-RAS-like suppressor (HRASLS) proteins are a group within the LRAT family that was shown to display both acyltransferase and phospholipase A1/2 activities[74]. YiiX-like family members from *Bacillus cereus* are also active against lipids[72]. In our tree, STox15 emerged as the sister clade of the so-called Pfam DUF4105 domain (Fig. 3D and 3E). While the biochemical function of DUF4105 has not yet been

characterized, recent studies have linked it to membrane protein and lipid processing[69]. Additionally, our comparative genomics analysis recovered a recurring evolutionary pattern: the DUF4105 domain-containing proteins are repeatedly displaced by apolipoprotein N-acyltransferases (Lnt) across three distinct genomic contexts (Fig. 3E). Additionally, our searches recovered a further clade of permuted papain-like domains related to DUF4105 from Gram-positive bacteria which was not detected by the said Pfam HMM (Fig. 3D). Notably, these Gram-positive bacteria lack Lnt homologs. Lnt plays a crucial role in the maturation of lipoproteins by catalyzing the phospholipid-dependent N-acylation of the N-terminal cysteine residue of apolipoproteins[75]. The repeated replacement suggests a similar function for these domains, indicating that DUF4105 and the related Gram-positive clade might act as an equivalent acyl transferase.

Importantly, we found that the cognate immunity protein for STox15, SImm15, contains a conserved domain with two transmembrane helices (Fig. S3). This suggests that the likely site of the neutralizing action of SImm15 is at the cell membrane, again consistent with a lipid-targeting role for STox15.

Multiple sequence alignments of each of the permuted clades including STox15 revealed the conserved catalytic His and Cys residues characteristic of the NlpC/P60 superfamily (Fig. 3F)[68]. Substitution of these residues for alanine ($STox15_{H43A}$ and $STox15_{C151A}$) abrogated STox15 toxicity in *E. coli* (Fig. 3G). These findings collectively confirm that the enzymatic function of the NlpC/P60 papain-like fold domain is crucial for toxicity. This finding, together with STox15 localization in the periplasm for its action, presence of a membrane-associated immunity protein and the fatty acyl linkage targeting activities common in the permuted group of the NlpC/P60 superfamily, strongly supports a function for STox15 in targeting cell-membrane lipids.

STox15 displays phospholipase activity and changes the membrane phospholipid composition of intoxicated cells

To analyze the enzymatic activity of STox15, we incubated purified recombinant protein (Fig. S4) with either purified phosphatidylglycerol (PG) 16:0-18:1 or phosphatidylethanolamine (PE) 16:0-18:1 and analyzed the reaction product by HPLC coupled to mass spectrometry (Fig. 4A and B). Results showed that STox15 has predominantly phospholipase A1 activity and cleaves both PG and PE (a preference for cleaving off the 16:0 acyl chain) as observed by the accumulation of lysophospholipids 18:1 (Fig. 4A and B).

Next, we set out to determine whether STox15 could cause changes in the composition of phospholipids when ectopically expressed by target cells. *E. coli* harboring the pBRA SP-STox15 plasmid were grown to OD_{600nm} of 1.0 in the presence of d-glucose (repressed), washed and resuspended in AB medium with l-arabinose to induce the expression of the toxin for 1h at 37°C. Total lipids were extracted and analyzed by UHPLC-MS, and spectra were manually identified. We observed a general decrease in intact phospholipid forms in $STox15_{WT}$, especially phosphatidylglycerol (PG), when compared with the catalytic mutant $STox15_{C151A}$, (Fig. 4A and 4B). In addition, an increase in lysophospholipids forms - either lysophosphatidylglycerol (LPG) or lysophosphatidylethanolamine (LPE) - and free fatty acids (FFA) was also detected in the wild-type (Fig. 4A and 4B). Lysophospholipids possess amphiphilic properties and have an inverted cone-shaped molecular structure that interacts with and modifies membrane properties (e.g. curvature) similarly to detergents[76]. The accumulation of lysophospholipids on the target cell membrane likely promotes the observed membrane disruption (Fig. 3C) due to its detergent-like properties[77]. Collectively, these results confirm that STox15 targets phospholipids similarly to other proteins possessing a

permuted NlpC/P60 domain, and that STox15 displays phospholipid-degrading activity. It remains to be determined if STox15 also possesses acyltransferase activity as reported for other permuted NlpC/P60 domains.

Discussion

A critical step for *Salmonella* to establish a successful infection is overcoming the colonization resistance imposed by the microbiota, which occupies gut niches and utilizes available resources. To displace these residents, *Salmonella* employs various strategies, including the injection of toxins through contact-dependent mechanisms that rely on T6SSs. In this study, we conducted a comprehensive analysis of *Salmonella* T6SS effectors utilizing a dataset of 10,000 genomes. Our bioinformatic pipeline identified 128 candidates, which were manually curated to elucidate their putative function and target. This rigorous curation process revealed a diverse arsenal employed by various *Salmonella* serovars and uncovered previously uncharacterized toxin domains, significantly expanding our understanding of the weaponry used in inter-species biological conflicts.

Previous strategies for discovering T6SS effectors have employed a variety of approaches, ranging from experimental screens[51,78,79] to *in silico* analysis[36,61-64,80-84]. Our approach integrated classical methods (e.g. SecRet6 and Bastion6) with the latest techniques (e.g. structural bioinformatics) to identify and classify T6SS effectors. Notably, our study was unique as it was the only large-scale *in silico* analysis to meticulously manually curate all putative candidates, thereby providing an additional level of confidence in our results.

In the context of *Salmonella* biology, the unprecedented diversity of T6SS effectors presents numerous opportunities for new studies. Our genomic analysis allowed us to estimate the number of effectors per genome. While we acknowledge that our

findings might be an underestimate, the use of multiple HMM libraries for searching homologs and the careful evaluation of all putative candidates makes it unlikely that the true number is significantly higher. Overall, our findings reveal the existence of a core repertoire of T6SS effectors for each serovar (Fig. S2), suggesting that *Salmonella* acquire and maintain T6SS effectors in response to specific environmental pressures and contexts rather than accumulating an increasingly larger array of effectors. Notably, we observed a higher number of effectors in serovars isolated from environmental sources compared to those from patients (Fig. S1), indicating that the number of effectors increases in more diverse environments where there are potentially more encounters with a variety of competitor species.

Furthermore, our analysis unveiled a remarkably diverse repertoire of effectors among the phylogenetically distinct subtypes of T6SS. In the case of *Salmonella*, effectors from the subtype i3 and i2 are predominantly antibacterial, targeting critical components such as peptidoglycan and nucleic acids. In contrast, subtype i1 is characterized by a diverse array of effectors with anti-eukaryotic activity, including lipases and pore-forming proteins. Remarkably, the functions of many effectors within subtype i1 remain undetermined. These findings highlight a potentially extensive array of virulence factors that have been overlooked over the years, thereby opening numerous avenues for future research.

As proof of concept, we conducted an in-depth characterization of one of the newly identified effector domains from this study. STox15 is unique as it is one of the first members of the permuted NlpC/P60 group associated with biological conflicts. Interestingly, unlike the canonical members of this group, STox15 does not exhibit amidase or peptidase activity but rather demonstrates phospholipase activity *in vitro* and *in vivo*. STox15 appears to preferentially cleave the acyl group of phosphatidylglycerol

in vivo, leading to the accumulation of lysophospholipid forms in the membranes of target cells, which act similarly to detergents to promote cell lysis. Notably, one of the homologs of STox15 is TseH, a T6SS effector of *V. cholerae*[51,85], which has been proposed to be an endopeptidase due to its membership in the NlpC/P60 superfamily and its similarity to Tse1, an amidase T6SS effector from *Pseudomonas aeruginosa*[86,87]. However, unlike Tse1, which belongs to the canonical group of NlpC/P60, TseH exhibits a permutation in its catalytic core[85]. This permutation, along with TseH's evolutionary relationship to STox15 and other permuted NlpC/P60 enzymes, suggests that TseH actual substrate might be an acyl group in phospholipids rather than a peptide/amide bond.

In conclusion, our comprehensive analysis of *Salmonella* T6SS effectors has significantly expanded the understanding of the diverse arsenal employed by various serovars. The identification and curation of 128 effector candidates revealed a core repertoire of T6SS effectors for each serovar, highlighting their role in response to serovar-specific environmental pressures. The discovery of previously uncharacterized toxin domains, such as STox15, underscores the potential for the discovery of novel biochemical activities among toxin domains. Additionally, exploring the evolutionary mechanisms driving the diversity of T6SS effectors will provide a deeper understanding of their adaptation and specialization. This study lays the groundwork for further investigations into the intricate dynamics of T6SS-mediated interactions and their implications for *Salmonella* ecology and pathogenesis.

Methods

Comparative genomic analysis

The .gff files from the genome assemblies retrieved from the 10KSG project[27] were organized and stored in tabular format using Python scripts, based on the

Biopython[88] and pandas[89] libraries. Iterative searches were conducted using Jackhmmer[90] with a 10e-6 e-value cutoff. Protein clustering was performed using MMseqs[32]. Redundancy was removed (80% coverage and 70% identity) and homologous groups were formed (80% coverage and e-value 10e-3). Multiple sequence alignments were generated using the local-pair algorithm in MAFFT[91], and phylogenetic trees were constructed using FastTree[92]. Domain identification and annotation was performed using HMMsearch and HMMscan[90,93] and models from the databases Pfam[30], TXSScan[28], and BastionHub[94]. Remote homology identification was performed using HHpred[39].

Bacterial strains

A list of bacterial strains used in this work can be found in Table S4. Strains were grown at 37 °C in Lysogeny Broth (10 g/L tryptone, 10 g/L NaCl, 5 g/L yeast extract) under agitation. AB medium was used for lipidomics: 0.2% (NH4)2SO2, 0.6% Na2HPO4, 0.3% KH2PO4, 0.3% NaCl, 0.1 mM CaCL2, 1 mM MgCl2, 3 µM FeCL3, supplemented with 0.2% sucrose, 0.2% casamino acids, 10 µg/mL thiamine, and 25 µg/mL uracil. Cultures were supplemented with antibiotics in the following concentration when necessary: 50 µg/mL kanamycin, 100 µg/mL ampicillin, and 50 µg/mL streptomycin.

Cloning and mutagenesis

All primers are listed in Table S4. STox15 and SImm15 were amplified by PCR and cloned into pBRA vector under the control of P_{BAD} promoter[95] with or without pelB signal peptide sequence from pET22b (Novagen)[96]. SImm15 was cloned into pEXT22 under the control of P_{TAC} promoter[97]. For protein expression and purification, STox15 was cloned into pET28a (Novagen), including a C-terminal Strep II tag. Point mutations

(STox15$_{H43A}$, STox15$_{C151A}$) were created using QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies) or by splicing by overlap extension (SOE) PCR. All constructs were confirmed by sequencing.

E. coli toxicity assay

Overnight cultures of *E. coli* DH5α co-expressing effectors for cytoplasmic (pBRA-STox15) or periplasmic (pBRA SP-STox15) localization and immunity protein (pEXT22-SImm15) were adjusted to OD_{600nm} 1, serially diluted in LB (1:4) and 5 µL were spotted onto LB agar (1.5%) containing either 0.2% d-glucose or 0.2% l-arabinose plus 200 µM IPTG, supplemented with streptomycin and kanamycin, and incubated at 37°C for 20 h.

Time-lapse microscopy

For time-lapse microscopy, LB agar (1.5%) pads were prepared by cutting a rectangular piece out of a double-sided adhesive tape, which was taped onto a microscopy slide as described previously[96]. *E. coli* DH5α harboring pBRA SP-STox15 was subcultured in LB (1:50) with 0.2% d-glucose until reaching OD_{600nm} 0.4–0.6 and adjusted to OD_{600nm} 1. Cultures were spotted onto LB agar pads supplemented either with 0.2% d-glucose or 0.2% l-arabinose plus antibiotics. Images were acquired every 15 min for 16 hr using a Leica DMi-8 epifluorescent microscope fitted with a DFC365 FX camera (Leica) and Plan-Apochromat ×63 oil objective (HC PL APO ×63/1.4 Oil ph3 objective Leica). Images were analyzed using FIJI software (Schindelin et al., 2012).

Protein expression and purification

E. coli BL21(DE3) carrying pET28a STox15$_{WT}$-Strep or STox15$_{C151A}$-Strep were grown in 4 L of LB supplemented with kanamycin (37 °C, 180 rpm) until OD_{600nm} 0.7. Expression was induced with 1 mM IPTG for 16 hr at 16 °C. Cells were harvested via centrifugation at 5000 *g* for 20 min, and pellets were resuspended in lysis buffer (50 mM Tris HCl pH 8.0, 350 mM NaCl, 45 mM β-mercaptoethanol, 5 mg/mL lysozyme, 10% glycerol) and lysed at 4°C using a sonicator. The lysate was centrifuged at 40,000 *g* for 45 min at 4°C. The supernatant was loaded onto a 1 ml StrepTrap HP column (Cytiva) equilibrated in buffer (50 mM Tris-HCl pH 8.0, 350 mM NaCl, 10% glycerol). The column was washed with 40 column volumes (CV) of wash buffer (50 mM Tris-HCl pH 8.0, 1 M NaCl, 10% glycerol, 45 mM β-mercaptoethanol), followed by a second wash with 12 CV (50 mM Tris-HCl pH 8.0, 1.5 M urea) to remove chaperonin GroEL[98]. The column was subjected to a third round of washes with 40 CV of wash buffer and eluted with 10 CV of elution buffer (50 mM Tris-HCl pH 8.0; 350 mM NaCl; 10% glycerol; 50 mM biotin). Fractions were buffer exchanged (25mM Tris-HCl pH 7.5, 100 mM NaCl, 5% glycerol) and concentrated using an Amicon of 30 kDa (Sigma). Protein aliquots were snap-frozen until use.

***In vitro* phospholipase assay**

For *in vitro* enzymatic assay, phospholipids 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (PG 16:0-18:1) and 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (PE 16:0-18:1) were purchased from Avanti Polar Lipids. Substrates were resuspended and diluted in methanol to adjust the concentration of aliquots. The methanol of each aliquot was dried under a nitrogen flow. A total of 1.2 mM of phospholipids (PG or PE) were resuspended in reaction buffer (25 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.5 mM $CaCl_2$, 0.5 mM $MgCl_2$, 180 mM sodium deoxycholate

and 0.5 mM DTT) and incubated with 0.8 mM of enzyme ($STox15_{WT}$ or $STox15_{C151A}$) in a total volume of 100 µL for 2 h at 37°C under agitation (350 rpm). Lipids were extracted by adding 830 µL of a mixture of MTBE/methanol/water (10:3:2.5, v/v/v), followed by incubation under agitation for 1 h at room temperature. Samples were centrifuged for 2 min at 220 *g* and 350 µL of the top fraction was transferred to a new tube, dried in a SpeedVac and stored at –80 °C until analysis.

For mass spectrometry analysis, samples were resuspended in 350 µL of isopropanol and analyzed in a Shimadzu 8060 Triple Quadrupole Liquid Chromatograph Mass Spectrometer. Samples (0.1-0.5 µL) were loaded into an Agilent column C18 ZORBAX Eclipse Plus (4.6 x 150 mm, 5 µm, 400 bar) with a flow rate of 0.5 mL/min and an oven temperature of 40 °C. HPLC gradients were as described below for lipidomic analysis. The phospholipids and lysophospholipids of interest were analyzed in the multiple reaction monitoring (MRM) mode using m/z transitions, collision energies (CE) and dwell times as shown in Table S5A. Data was acquired by Shimadzu LabSolutions and processed in LabSolutions Browser. Graphs were plotted using GraphPrism 5.

Lipidomics of STox15-intoxicated *E. coli*

E. coli MG1655 containing the plasmids pBRA SP-$STox15_{WT}$ or SP-$STox15_{C151A}$ were cultured in LB containing 0.2% d-glucose at 37 °C for 14 h. Cells were subcultured in LB 0.2% d-glucose until an OD_{600nm} of 1 and centrifuged (10 min, 2900 *g*, 30 °C). Cells were washed with 40 mL of preheated AB medium at 37 °C, centrifuged (10 min, 2900 *g*, 30 °C), and resuspended in 5 mL of AB supplemented with 0.2% l-arabinose to induce STox15 expression. Cells were incubated for 1 h at 37 °C with agitation (100 rpm). The cells were centrifuged (15 min, 2900 *g*, 4 °C) and washed once with 1 mL of PBS pH 7.4. PBS was removed by centrifugation and the cell pellet was stored at -80 °C until

lipid extraction. A cocktail of class specific internal standards was added to the cell mass prior to lipid extraction for subsequent quantification and normalization (Table S5B). Total lipid extraction was performed using an adapted version of the protocol described by Yoshida, et al. [99]. Briefly, cell pellets were resuspended with 500 µL of cold methanol and 1 mL of MilliQ water and transferred to glass tubes. A mixture of chloroform and ethyl acetate (4:1) was added, followed by agitation for 1 min at 25 °C. Samples were centrifuged (2 min, 1500 *g*, 4 °C) and the lipid-containing phase (lower phase) was extracted and transferred to a new glass tube that was dried under a nitrogen (N_2) flow until all solvent traces were evaporated. Samples were stored at -80 °C until analysis.

Lipid extracts were diluted in 100 µL of isopropanol and analyzed using ultra-high performance liquid chromatography (UHPLC Nexera, Shimadzu) coupled with an ESI-Q-TOF mass spectrometer (Triple TOF 6600, Sciex) (UHPLC-Q-TOF/MS). 2 µL of each sample were injected into the UHPLC-MS, and molecules were separated using a CORTECS column (C18, 1.6 µm, 2.1x100 mm, Waters) with a flow rate of 0.2 mL/min and temperature set to 35 °C [100]. The mobile phases consisted of (A) water/acetonitrile (60:40) and (B) isopropanol/acetonitrile/water (88:10:2). Ammonium acetate at a final concentration of 10 mM was incorporated in both mobile phases A and B for the negative ionization acquisition mode. The gradient elution used in the chromatography was from 40 to 100% (mobile phase) B over the first 10 min; 100% B from 10-12 min; 100 to 40% B for 12-13 and holding 40% B for 13-20 min. The negative mode was utilized for the examination of phospholipids and free fatty acids. MS and MS/MS data acquisition was performed using Analyst 1.7.1 software (Sciex). Mass spectrometry data was inspected using PeakView 2.0 software (Sciex), and lipid molecular species were manually identified with the help of an in-house manufactured Excel-based macro. Lipid species

were quantified using MultiQuant software (Sciex), where the precursor ions areas were normalized by the internal standards for each class (Table S5B).

Data availability

All data supporting the findings of this study are available within the paper and its Supplementary Information files or at https://leepbioinfo.github.io/10ksgt6ss/. The source data underlying Fig. 3B, Fig, 3G, Fig. 4A and Fig. 4B are provided as a Source Data file.

Code availability

All data and code used for sequence and genome context analyses are available on a GitHub repository at https://github.com/leepbioinfo/10ksgt6ss. The ROTIFER package can be downloaded from the GitHub repository https://github.com/leepbioinfo/rotifer. The contents of both repositories are made available under the Creative Commons Attribution 4.0 International or the GNU Lesser General Public License (LGPL) v. 2.0. We have also uploaded the GitHub data and code to Zenodo (DOI: 10.5281/zenodo.13845778).

Author contributions

G.G.N., R.F.S. and E.B.-S. conceived the study; G.G.N., S.S.-S., T.W.C.S. and R.F.S. conducted bioinformatic analysis; S.S.-S., J.T.H., J.P.M., R.S.S., A.L.A.-S. performed experiments and analyzed data; B.M.P.-S., S.M., L.A. analyzed data and contributed with discussions; G.G.N., R.F.S. and E.B.-S. wrote the manuscript with input from the other authors.

Acknowledgments

We are grateful to Jay Hinton for initially sharing the 10KSG dataset with us, to Ian Riddington for assistance with mass-spectrometry analysis of phospholipids, and Gerd Prehna and Larissa Diniz for discussions about enzymatic assay. Marcos Yoshinaga from PinguisLab for lipidomic data processing. Edgar Llontop and Diorge Paulo Souza for troubleshooting with protein purification. Cristiano G. Moreira for sharing strains. Alexandre Bruni for access to the microscope. Luize Nobrega Silva and Gustavo Chagas for technical support. This research was supported by Sao Paulo Research Foundation grant #2017-02178-2 to E.B-S, #2013/07937-8 to S.M, #2016/09047-8 to R.F.S and #2021/10577-0 to T.W.C.S. Startup funds from MBS and CNS to E.B-S. This research was supported in part by an appointment to the National Library of Medicine Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and the National Library of Medicine. ORISE is managed by ORAU under DOE contract number DE-SC0014664. All opinions expressed in this paper are the author's and do not necessarily reflect the policies and views of NIH, NLM, DOE, or ORAU/ORISE. This work is supported by the funds of the Intramural Research Program of National Library of Medicine at the National Institutes of Health, USA (to L.A. and G.G.N). This work utilized the NIH HPC Biowulf computer cluster.

Competing interests

The authors declare no competing interests.

Figure legends.

Fig. 1: The T6SS effector repertoire within 10k *Salmonella* genomes.

(a) Pipeline used for classification of genomic sites and T6SS subtypes. **(b)** Number of genomes containing the different T6SS subtypes within the 10KSG dataset. **(c)** Examples of the genomic organization of T6SS structural clusters from distinct phylogenetic subtypes. **(d)** *In silico* strategies used for the identification and classification of T6SS effectors. **(e)** Schematic representation of the functional classes of T6SS effector domains identified in the 10KSG distributed according to target cell type (see the complete list in Table S1). **(f)** Comparison of profile Hidden Markov Models (pHMMs) of STox to published models of related families. Each circle corresponds to a population of proteins detected by an STox model. The relative frequency of proteins detected by an STox model compared to a reference model is shown on the horizontal axis. The HMM divergence score is shown in the vertical axis. The blue-white-red color scale of each dot represents the values for the Spearman correlation index between STox and reference alignment scores for the same proteins. Yellow circles represent STox models that detect proteins that are not recognized by previously existing models. The radius of each circle is proportional to the total number of proteins detected only by STox models. The regression line does not include the data points for the models represented by yellow circles.

Fig. 2: The diversity of *Salmonella* T6SS effectors and target-specific subsets.

(a) Normal distribution and fitted curve showing the number of effectors per genome (single cluster in blue and >2 clusters in orange). **(b)** The most frequent effectors identified in the 10KSG dataset. Each bar represents the number of genomes encoding a specific effector. Colors represent different biochemical activities, with light colors representing orphan effectors while dark colors represent effectors encoded within the structural cluster. **(c)** The five most frequent effectors encoded in different *Salmonella* serovars (see Fig. S2 for all the 149 serovars). Colors indicate the effector activity as in

(b). **(d)** Schematic representation of the most common sets of effectors in genomes encoding different T6SS subtypes. The number of genomes is indicated on the right. Colors represent activity as shown in (b). **(e)** Pie chart illustrating the relative proportions of effectors by activities encoded within the T6SS subtypes i3, i1, and i2. **(f)** Venn diagram illustrating the proportion of overlap between effectors encoded within each T6SS structural cluster (blue: i1; purple: subtype i2; red: subtype i3; and green: orphan). **(g)** Schematic representation of the genetic organization of T6SSs showing the position of variable regions in which the effector and immunity proteins are encoded. Colors denote structural proteins forming the membrane complex (orange), sheath and inner tube (light blue), baseplate and spike components (green). Effectors are shown in red, and immunities in dark blue.

Fig. 3: STox15 has a permuted NlpC/P60 domain that is required for toxicity.

(a) Scheme of the genomic region encoding STox15 and SImm15 effector/immunity pair (*S.* Oranienburg, barcode FD01848827). **(b)** *E. coli* toxicity assay. Serial dilutions of *E. coli* carrying pBRA and pEXT22 constructs, as indicated, were spotted onto LB agar plates, and grown for 20 hr. Images are representative of three independent experiments. **(c)** Time-lapse microscopy of *E. coli* carrying pBRA SP-STox15 grown on LB agar pads containing either 0.2% d-glucose (repressed) or 0.2% l-arabinose (induced). Scale bar: 5 µm. Timestamps in hh:mm. **(d)** Maximum-likelihood phylogenetic tree of permuted NlpC/P60 members. Dots represent the number of PSI-BLAST iterations required to collect homologs and the red star marks the query: *S.* Oranienburg STox15 depicted in (a). **(e)** Genomic organization of representatives from clades TseH and STox15 showing the genes are encoded in the context of conflict systems, and DUF4105 showing context of lipid metabolism. **(f)** Sequence logo from the enzymatic core of permuted NlpC/P60

from all clades shown in (d). The arrows indicate conserved His and Cys residues that were mutated in (g). **(g)** *E. coli* toxicity assay. Serial dilution of *E. coli* containing pBRA and pEXT22 constructs, as indicated, spotted onto LB agar plates and grown for 20 hr. Images are representative of three independent experiments.

Fig. 4: STox15 has phospholipase A activity and changes the composition of target cell membranes.

(a) *In vitro* enzymatic assay with recombinant STox15 (red) or $STox15_{C151A}$ (blue) incubated with different phospholipids (both 16:0-18:1 PG and PE) for 2 hr at 37 °C. The amount of lysophospholipids produced was analyzed and quantified by HPLC-MS/MS. **(b)** Quantification of the peak area of lysophospholipids normalized by the intact substrate. Data corresponds to the mean ± SD. ***$p < 0.001$ and *$p < 0.01$, *ns* not significant (Student's t-test). **(c)** UHPLC-MS total ion chromatogram showing the profile of total lipids extracted from *E. coli* expressing STox15 (red) or $STox15_{C151A}$ (blue). **(d)** Heatmap plot of top 20 altered lipids of intoxicated *E. coli* lipidome. Results display four biological replicates of each condition (WT or C151A) with the quantification of lipids species (Tukey test; $p < 0.05$ FDR adjusted). Data were expressed in nM mg of protein^{-1} and normalized by log transformation (base 10) prior to analysis. Red (up) and blue (down) bars represent changes in lipid species concentration relative to the normalized mean. Letters differentiate between the isomers.

Fig. S1: T6SS effector repertoire in the 10KSG dataset. Each column indicates the presence or absence of a toxin as identified by the models developed in this study. Lines denote unique combinations of effectors. The histogram on the right shows the frequency of genomes in the 10KSG dataset containing each specific repertoire. The histogram at

the bottom illustrates the frequency of genomes with at least one protein identified by the above HMM model. Effector activities are color-coded as described in Fig. 2B.

Fig. S2: List of five most frequent T6SS effectors identified in each of the 149 Salmonella serovars contained in the 10KSG dataset. Colors indicate the effector activity as in Fig. 2B.

Fig. S3: SImm15 is a two-transmembrane helices protein. (a) Sequence alignment of SImm15 homologs with orange rectangles indicating predicted transmembrane helices. Sequences are colored according to the Clustal X color scheme[101]. Consensus abbreviations: h, hydrophobic (A, C, F, I, L, M, V, W, Y); l, aliphatic (L, I, V); a, aromatic (F, W, Y); b, large residues (L, I, Y, E, R, F, Q, K, M, W); s, small residues (A, G, S, V, C, D, N); u, tiny residues (G, A, S); p, polar residues (S, T, E, D, K, R, N, Q, H, C); c, charged residues (D, E, H, K, R); +, positively charged residues (H, K, R); −, negatively charged residues (D, E). (B) Transmembrane helix prediction for SImm15 using DeepTMHMM[102].

Fig. S4: Recombinant protein purification used in enzymatic assay. SDS–PAGE of recombinant proteins during purification steps to obtain purified protein for enzymatic assays. Affinity chromatography using Strep-Tactin sepharose to purify STox15 versions with C-terminus Strep-tag II: WT in (a) or C151A in (b). Recombinant proteins were purified from the soluble fraction. An additional step of washing with 1.5M urea was performed after the traditional washes to remove contamination with GroEL before elution with biotin. Additional bands were identified by mass-spectrometry to confirm identity. MK: marker; INS: insoluble; SOL: soluble; FT: flowthrough. GroEL:

chaperonin GroEL; Bccp: biotin carboxyl carrier protein; OmpF: outer membrane porin F.

Table S1. List of all T6SS toxin domains identified in this study in the 10K *Salmonella* genomes dataset.

Table S2. List of all homologs collected by JackHMMER searches and used to build the phylogenetic tree shown in Fig. 3D.

Table S3. Genomic context of members of each permuted NlpC/P60 members shown in Fig. 3D.

Table S4. List of strains, plasmids and primers used in the study.

Table S5. (a) MRM transition for LC-MS/MS method of lysophospholipids. (b) Internal standards used for lipidomics analysis in negative mode.

Movie S1. Time-lapse microscopy of *E. coli* harboring pBRA SP-STox15 growing in media supplemented with 0.2% d-glucose. Timestamp in hh:mm. Scale bar: 5 μm.

Movie S2. Time-lapse microscopy of *E. coli* harboring pBRA SP-STox15 growing in media supplemented with 0.2% l-arabinose. Timestamp in hh:mm. Scale bar: 5 μm.

References

1 Granato, E. T., Meiller-Legrand, T. A. & Foster, K. R. The Evolution and Ecology of Bacterial Warfare. *Curr Biol* **29**, R521-R537 (2019). https://doi.org/10.1016/j.cub.2019.04.024

2 Hood, R. D., Peterson, S. B. & Mougous, J. D. From Striking Out to Striking Gold: Discovering that Type VI Secretion Targets Bacteria. *Cell host & microbe* **21**, 286-289 (2017). https://doi.org/10.1016/j.chom.2017.02.001

3 Filloux, A., Hachani, A. & Bleves, S. The bacterial type VI secretion machine: yet another player for protein transport across membranes. *Microbiology (Reading)* **154**, 1570-1583 (2008). https://doi.org/10.1099/mic.0.2008/016840-0

4 Nguyen, V. S. *et al.* Towards a complete structural deciphering of Type VI secretion system. *Curr Opin Struct Biol* **49**, 77-84 (2018). https://doi.org/10.1016/j.sbi.2018.01.007

5 Russell, A. B. *et al.* A type VI secretion-related pathway in Bacteroidetes mediates interbacterial antagonism. *Cell host & microbe* **16**, 227-236 (2014). https://doi.org/10.1016/j.chom.2014.07.007

6 Li, J. *et al.* SecReT6: a web-based resource for type VI secretion systems found in bacteria. *Environ Microbiol* **17**, 2196-2202 (2015). https://doi.org/10.1111/1462-2920.12794

7 Bock, D. *et al.* In situ architecture, function, and evolution of a contractile injection system. *Science* **357**, 713-717 (2017). https://doi.org/10.1126/science.aan7904

8 Boyer, F., Fichant, G., Berthod, J., Vandenbrouck, Y. & Attree, I. Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learned from available microbial genomic resources? *BMC Genomics* **10**, 104 (2009). https://doi.org/10.1186/1471-2164-10-104

9 Barret, M., Egan, F., Fargier, E., Morrissey, J. P. & O'Gara, F. Genomic analysis of the type VI secretion systems in Pseudomonas spp.: novel clusters and putative effectors uncovered. *Microbiology (Reading)* **157**, 1726-1739 (2011). https://doi.org/10.1099/mic.0.048645-0

10 Hood, R. D. *et al.* A type VI secretion system of Pseudomonas aeruginosa targets a toxin to bacteria. *Cell host & microbe* **7**, 25-37 (2010). https://doi.org/10.1016/j.chom.2009.12.007

11 Cherrak, Y. *et al.* Biogenesis and structure of a type VI secretion baseplate. *Nat Microbiol* **3**, 1404-1416 (2018). https://doi.org/10.1038/s41564-018-0260-1

12 Jana, B. & Salomon, D. Type VI secretion system: a modular toolkit for bacterial dominance. *Future Microbiol* **14**, 1451-1463 (2019). https://doi.org/10.2217/fmb-2019-0194

13 Allsopp, L. P. & Bernal, P. Killing in the name of: T6SS structure and effector diversity. *Microbiology (Reading)* **169** (2023). https://doi.org/10.1099/mic.0.001367

14 Bohnhoff, M., Drake, B. L. & Miller, C. P. Effect of streptomycin on susceptibility of intestinal tract to experimental Salmonella infection. *Proc Soc Exp Biol Med* **86**, 132-137 (1954). https://doi.org/10.3181/00379727-86-21030

15 Rogers, A. W. L., Tsolis, R. M. & Baumler, A. J. Salmonella versus the Microbiome. *Microbiol Mol Biol Rev* **85** (2021). https://doi.org/10.1128/MMBR.00027-19

16 Sibinelli-Sousa, S., de Araujo-Silva, A. L., Hespanhol, J. T. & Bayer-Santos, E. Revisiting the steps of Salmonella gut infection with a focus on antagonistic interbacterial interactions. *FEBS J* (2021). https://doi.org/10.1111/febs.16211

17 Blondel, C. J., Jimenez, J. C., Contreras, I. & Santiviago, C. A. Comparative genomic analysis uncovers 3 novel loci encoding type six secretion systems differentially distributed in Salmonella serotypes. *BMC Genomics* **10**, 354 (2009). https://doi.org/10.1186/1471-2164-10-354

18 Bao, H. *et al.* Genetic diversity and evolutionary features of type VI secretion systems in Salmonella. *Future Microbiol* **14**, 139-154 (2019). https://doi.org/10.2217/fmb-2018-0260

19 Brunet, Y. R. *et al.* H-NS Silencing of the Salmonella Pathogenicity Island 6-Encoded Type VI Secretion System Limits Salmonella enterica Serovar Typhimurium Interbacterial Killing. *Infection and immunity* **83**, 2738-2750 (2015). https://doi.org/10.1128/IAI.00198-15

20 Sana, T. G. *et al.* Salmonella Typhimurium utilizes a T6SS-mediated antibacterial weapon to establish in the host gut. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E5044-5051 (2016). https://doi.org/10.1073/pnas.1608858113

21 Blondel, C. J. *et al.* The type VI secretion system encoded in Salmonella pathogenicity island 19 is required for Salmonella enterica serotype Gallinarum survival within infected macrophages. *Infection and immunity* **81**, 1207-1220 (2013). https://doi.org/10.1128/IAI.01165-12

22 Schroll, C. *et al.* The SPI-19 encoded type-six secretion-systems (T6SS) of Salmonella enterica serovars Gallinarum and Dublin play different roles during infection. *Vet Microbiol* **230**, 23-31 (2019). https://doi.org/10.1016/j.vetmic.2019.01.006

23 Blondel, C. J. *et al.* Contribution of the type VI secretion system encoded in SPI-19 to chicken colonization by Salmonella enterica serotypes Gallinarum and Enteritidis. *PloS one* **5**, e11724 (2010). https://doi.org/10.1371/journal.pone.0011724

24 Hespanhol, J. T. *et al.* Antibacterial T6SS effectors with a VRR-Nuc domain are structure-specific nucleases. *Elife* **11** (2022). https://doi.org/10.7554/eLife.82437

25 Sibinelli-Sousa, S. *et al.* A Family of T6SS Antibacterial Effectors Related to l,d-Transpeptidases Targets the Peptidoglycan. *Cell Rep* **31**, 107813 (2020). https://doi.org/10.1016/j.celrep.2020.107813

26 Aravind, L. Guilt by association: contextual information in genome analysis. *Genome Res* **10**, 1074-1077 (2000). https://doi.org/10.1101/gr.10.8.1074

27 Perez-Sepulveda, B. M. *et al.* An accessible, efficient and global approach for the large-scale sequencing of bacterial genomes. *Genome Biol* **22**, 349 (2021). https://doi.org/10.1186/s13059-021-02536-3

28 Abby, S. S. *et al.* Identification of protein secretion systems in bacterial genomes. *Sci Rep* **6**, 23080 (2016). https://doi.org/10.1038/srep23080

29 Zhang, D., de Souza, R. F., Anantharaman, V., Iyer, L. M. & Aravind, L. Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol Direct* **7**, 18 (2012). https://doi.org/10.1186/1745-6150-7-18

30 Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**, D412-D419 (2021). https://doi.org/10.1093/nar/gkaa913

31 Wang, J. *et al.* The conserved domain database in 2023. *Nucleic Acids Res* **51**, D384-D388 (2023). https://doi.org/10.1093/nar/gkac1096

32 Hauser, M., Steinegger, M. & Soding, J. MMseqs software suite for fast and deep clustering and searching of large protein sequence sets. *Bioinformatics* **32**, 1323-1330 (2016). https://doi.org/10.1093/bioinformatics/btw006

33 Jaccard, P. The Distribution of the Flora in the Alpine Zone. *New Phytologist* **11**, 37-50 (1912). https://doi.org/10.1111/J.1469-8137.1912.TB05611.X

34 De Meo, P. F., E.; Fiumara, G.; Provetti, A. Generalized Louvain method for community detection in large networks. *International Conference on Intelligent Systems Design and Applications* (2011). https://doi.org/10.1109/ISDA.2011.6121636

35 Hansen-Wester, I. & Hensel, M. Salmonella pathogenicity islands encoding type III secretion systems. *Microbes and infection / Institut Pasteur* **3**, 549-559 (2001). https://doi.org/10.1016/s1286-4579(01)01411-3

36 Wang, J. *et al.* Bastion6: a bioinformatics approach for accurate prediction of type VI secreted effectors. *Bioinformatics* **34**, 2546-2555 (2018). https://doi.org/10.1093/bioinformatics/bty155

37 Zhang, J. *et al.* SecReT6 update: a comprehensive resource of bacterial Type VI Secretion Systems. *Sci China Life Sci* (2022). https://doi.org/10.1007/s11427-022-2172-x

38 Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402 (1997). https://doi.org/10.1093/nar/25.17.3389

39 Soding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33**, W244-248 (2005). https://doi.org/10.1093/nar/gki408

40 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021). https://doi.org/10.1038/s41586-021-03819-2

41 van Kempen, M. *et al.* Fast and accurate protein structure search with Foldseek. *Nat Biotechnol* **42**, 243-246 (2024). https://doi.org/10.1038/s41587-023-01773-0

42 Holm, L. Using Dali for Protein Structure Comparison. *Methods Mol Biol* **2112**, 29-42 (2020). https://doi.org/10.1007/978-1-0716-0270-6_3

43 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235-242 (2000). https://doi.org/10.1093/nar/28.1.235

44 Varadi, M. *et al.* AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res* **52**, D368-D375 (2024). https://doi.org/10.1093/nar/gkad1011

45 Russell, A. B. *et al.* A widespread bacterial type VI secretion effector superfamily identified using a heuristic approach. *Cell host & microbe* **11**, 538-549 (2012). https://doi.org/10.1016/j.chom.2012.04.007

46 Jurenas, D. *et al.* Salmonella antibacterial Rhs polymorphic toxin inhibits translation through ADP-ribosylation of EF-Tu P-loop. *Nucleic Acids Res* **50**, 13114-13127 (2022). https://doi.org/10.1093/nar/gkac1162

47 Mariano, G. *et al.* A family of Type VI secretion system effector proteins that form ion-selective pores. *Nat Commun* **10**, 5484 (2019). https://doi.org/10.1038/s41467-019-13439-0

48 Hu, H. *et al.* Structure of the type VI secretion phospholipase effector Tle1 provides insight into its hydrolysis and membrane targeting. *Acta Crystallogr D Biol Crystallogr* **70**, 2175-2185 (2014). https://doi.org/10.1107/S1399004714012899

49 Bernal, P., Allsopp, L. P., Filloux, A. & Llamas, M. A. The Pseudomonas putida T6SS is a plant warden against phytopathogens. *ISME J* **11**, 972-987 (2017). https://doi.org/10.1038/ismej.2016.169

50 Miyata, S. T., Kitaoka, M., Brooks, T. M., McAuley, S. B. & Pukatzki, S. Vibrio cholerae requires the type VI secretion system virulence factor VasX to kill Dictyostelium discoideum. *Infection and immunity* **79**, 2941-2949 (2011). https://doi.org/10.1128/IAI.01266-10

51 Altindis, E., Dong, T., Catalano, C. & Mekalanos, J. Secretome analysis of Vibrio cholerae type VI secretion system reveals a new effector-immunity pair. *mBio* **6**, e00075 (2015). https://doi.org/10.1128/mBio.00075-15

52 Whitney, J. C. *et al.* Identification, structure, and function of a novel type VI secretion peptidoglycan glycoside hydrolase effector-immunity pair. *The Journal of biological chemistry* **288**, 26616-26624 (2013). https://doi.org/10.1074/jbc.M113.488320

53 Bullen, N. P. *et al.* An ADP-ribosyltransferase toxin kills bacterial cells by modifying structured non-coding RNAs. *Molecular cell* **82**, 3484-3498 e3411 (2022). https://doi.org/10.1016/j.molcel.2022.08.015

54 Li, D. Y. *et al.* Identification and Characterization of EvpQ, a Novel T6SS Effector Encoded on a Mobile Genetic Element in Edwardsiella piscicida. *Front Microbiol* **12**, 643498 (2021). https://doi.org/10.3389/fmicb.2021.643498

55 Berni, B., Soscia, C., Djermoun, S., Ize, B. & Bleves, S. A Type VI Secretion System Trans-Kingdom Effector Is Required for the Delivery of a Novel Antibacterial Toxin in Pseudomonas aeruginosa. *Front Microbiol* **10**, 1218 (2019). https://doi.org/10.3389/fmicb.2019.01218

56 Yahr, T. L., Vallis, A. J., Hancock, M. K., Barbieri, J. T. & Frank, D. W. ExoY, an adenylate cyclase secreted by the Pseudomonas aeruginosa type III system. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 13899-13904 (1998). https://doi.org/10.1073/pnas.95.23.13899

57 Poole, S. J. *et al.* Identification of functional toxin/immunity genes linked to contact-dependent growth inhibition (CDI) and rearrangement hotspot (Rhs) systems. *PLoS Genet* **7**, e1002217 (2011). https://doi.org/10.1371/journal.pgen.1002217

58 Russell, A. B. *et al.* Diverse type VI secretion phospholipases are functionally plastic antibacterial effectors. *Nature* **496**, 508-512 (2013). https://doi.org/10.1038/nature12074

59 Koskiniemi, S. *et al.* Rhs proteins from diverse bacteria mediate intercellular competition. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 7032-7037 (2013). https://doi.org/10.1073/pnas.1300627110

60 Jiang, F., Waterfield, N. R., Yang, J., Yang, G. & Jin, Q. A Pseudomonas aeruginosa type VI secretion phospholipase D effector targets both prokaryotic and eukaryotic cells. *Cell host & microbe* **15**, 600-610 (2014). https://doi.org/10.1016/j.chom.2014.04.010

61 Blondel, C. J., Amaya, F. A., Bustamante, P., Santiviago, C. A. & Pezoa, D. Identification and distribution of new candidate T6SS effectors encoded in Salmonella Pathogenicity Island 6. *Front Microbiol* **14**, 1252344 (2023). https://doi.org/10.3389/fmicb.2023.1252344

62 Geller, A. M., Shalom, M., Zlotkin, D., Blum, N. & Levy, A. Identification of type VI secretion system effector-immunity pairs using structural bioinformatics. *Mol Syst Biol* **20**, 702-718 (2024). https://doi.org/10.1038/s44320-024-00035-8

63 Kanarek, K., Fridman, C. M., Bosis, E. & Salomon, D. The RIX domain defines a class of polymorphic T6SS effectors and secreted adaptors. *Nat Commun* **14**, 4983 (2023). https://doi.org/10.1038/s41467-023-40659-2

64 Carobbi, A. *et al.* PIX is an N-terminal delivery domain that defines a class of polymorphic T6SS effectors in Enterobacterales. *Cell Rep* **43**, 114015 (2024). https://doi.org/10.1016/j.celrep.2024.114015

65 Koskiniemi, S. *et al.* Selection of orphan Rhs toxin expression in evolved Salmonella enterica serovar Typhimurium. *PLoS Genet* **10**, e1004255 (2014). https://doi.org/10.1371/journal.pgen.1004255

66 Quentin, D. *et al.* Mechanism of loading and translocation of type VI secretion system effector Tse6. *Nat Microbiol* **3**, 1142-1152 (2018). https://doi.org/10.1038/s41564-018-0238-z

67 Ahmad, S. *et al.* Structural basis for effector transmembrane domain recognition by type VI secretion system chaperones. *Elife* **9** (2020). https://doi.org/10.7554/eLife.62816

68 Anantharaman, V. & Aravind, L. Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. *Genome Biol* **4**, R11 (2003). https://doi.org/10.1186/gb-2003-4-2-r11

69 Ozhelvaci, F. & Steczkiewicz, K. Identification and classification of papain-like cysteine proteinases. *The Journal of biological chemistry* **299**, 104801 (2023). https://doi.org/10.1016/j.jbc.2023.104801

70 Griffin, M. E., Klupt, S., Espinosa, J. & Hang, H. C. Peptidoglycan NlpC/P60 peptidases in bacterial physiology and host interactions. *Cell Chem Biol* **30**, 436-456 (2023). https://doi.org/10.1016/j.chembiol.2022.11.001

71 Golczak, M. *et al.* Structural basis for the acyltransferase activity of lecithin:retinol acyltransferase-like proteins. *The Journal of biological chemistry* **287**, 23790-23807 (2012). https://doi.org/10.1074/jbc.M112.361550

72 Xu, Q. *et al.* Structural analysis of papain-like NlpC/P60 superfamily enzymes with a circularly permuted topology reveals potential lipid binding sites. *PloS one* **6**, e22013 (2011). https://doi.org/10.1371/journal.pone.0022013

73 Sears, A. E. & Palczewski, K. Lecithin:Retinol Acyltransferase: A Key Enzyme Involved in the Retinoid (visual) Cycle. *Biochemistry* **55**, 3082-3091 (2016). https://doi.org/10.1021/acs.biochem.6b00319

74 Mardian, E. B., Bradley, R. M. & Duncan, R. E. The HRASLS (PLA/AT) subfamily of enzymes. *J Biomed Sci* **22**, 99 (2015). https://doi.org/10.1186/s12929-015-0210-7

75 Wiktor, M. *et al.* Structural insights into the mechanism of the membrane integral N-acyltransferase step in bacterial lipoprotein synthesis. *Nat Commun* **8**, 15952 (2017). https://doi.org/10.1038/ncomms15952

76 Zheng, L., Lin, Y., Lu, S., Zhang, J. & Bogdanov, M. Biogenesis, transport and remodeling of lysophospholipids in Gram-negative bacteria. *Biochim Biophys Acta Mol Cell Biol Lipids* **1862**, 1404-1413 (2017). https://doi.org/10.1016/j.bbalip.2016.11.015

77 Henriksen, J. R., Andresen, T. L., Feldborg, L. N., Duelund, L. & Ipsen, J. H. Understanding detergent effects on lipid membranes: a model study of lysolipids. *Biophys J* **98**, 2199-2205 (2010). https://doi.org/10.1016/j.bpj.2010.01.037

78 Fritsch, M. J. *et al.* Proteomic identification of novel secreted antibacterial toxins of the Serratia marcescens type VI secretion system. *Mol Cell Proteomics* **12**, 2735-2749 (2013). https://doi.org/10.1074/mcp.M113.030502

79 Dong, T. G., Ho, B. T., Yoder-Himes, D. R. & Mekalanos, J. J. Identification of T6SS-dependent effector and immunity proteins by Tn-seq in Vibrio cholerae. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 2623-2628 (2013). https://doi.org/10.1073/pnas.1222783110

80 Salomon, D. *et al.* Marker for type VI secretion system effectors. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 9271-9276 (2014). https://doi.org/10.1073/pnas.1406110111

81 Jana, B., Fridman, C. M., Bosis, E. & Salomon, D. A modular effector with a DNase domain and a marker for T6SS substrates. *Nat Commun* **10**, 3595 (2019). https://doi.org/10.1038/s41467-019-11546-6

82 Fridman, C. M., Keppel, K., Gerlic, M., Bosis, E. & Salomon, D. A comparative genomics methodology reveals a widespread family of membrane-disrupting T6SS effectors. *Nat Commun* **11**, 1085 (2020). https://doi.org/10.1038/s41467-020-14951-4

83 Jana, B., Keppel, K., Fridman, C. M., Bosis, E. & Salomon, D. Multiple T6SSs, Mobile Auxiliary Modules, and Effectors Revealed in a Systematic Analysis of the Vibrio parahaemolyticus Pan-Genome. *mSystems* **7**, e0072322 (2022). https://doi.org/10.1128/msystems.00723-22

84 Tchelet, D., Keppel, K., Bosis, E. & Salomon, D. Vibrio parahaemolyticus T6SS2 effector repertoires. *Gut Microbes* **15**, 2178795 (2023). https://doi.org/10.1080/19490976.2023.2178795

85 Hersch, S. J. *et al.* Envelope stress responses defend against type six secretion system attacks independently of immunity proteins. *Nat Microbiol* **5**, 706-714 (2020). https://doi.org/10.1038/s41564-020-0672-6

86 Russell, A. B. *et al.* Type VI secretion delivers bacteriolytic effectors to target cells. *Nature* **475**, 343-347 (2011). https://doi.org/10.1038/nature10244

87 Chou, S. *et al.* Structure of a peptidoglycan amidase effector targeted to Gram-negative bacteria by the type VI secretion system. *Cell Rep* **1**, 656-664 (2012). https://doi.org/10.1016/j.celrep.2012.05.016

88 Cock, P. J. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422-1423 (2009). https://doi.org/10.1093/bioinformatics/btp163

89 McKinney, W. o. Data structures for statistical computing in python. . *Proceedings of the 9th Python in Science Conference* **445**, 51-56 (2010).

90 Johnson, L. S., Eddy, S. R. & Portugaly, E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11**, 431 (2010). https://doi.org/10.1186/1471-2105-11-431

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

92 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PloS one* **5**, e9490 (2010). https://doi.org/10.1371/journal.pone.0009490

93 Eddy, S. R. Accelerated Profile HMM Searches. *PLoS computational biology* **7**, e1002195 (2011). https://doi.org/10.1371/journal.pcbi.1002195

94 Wang, J. *et al.* BastionHub: a universal platform for integrating and analyzing substrates secreted by Gram-negative bacteria. *Nucleic Acids Res* **49**, D651-D659 (2021). https://doi.org/10.1093/nar/gkaa899

95 Souza, D. P. *et al.* Bacterial killing via a type IV secretion system. *Nat Commun* **6**, 6453 (2015). https://doi.org/10.1038/ncomms7453

96 Bayer-Santos, E. *et al.* The opportunistic pathogen Stenotrophomonas maltophilia utilizes a type IV secretion system for interbacterial killing. *PLoS pathogens* **15**, e1007651 (2019). https://doi.org/10.1371/journal.ppat.1007651

97 Dykxhoorn, D. M., St Pierre, R. & Linn, T. A set of compatible tac promoter expression vectors. *Gene* **177**, 133-136 (1996). https://doi.org/10.1016/0378-1119(96)00289-2

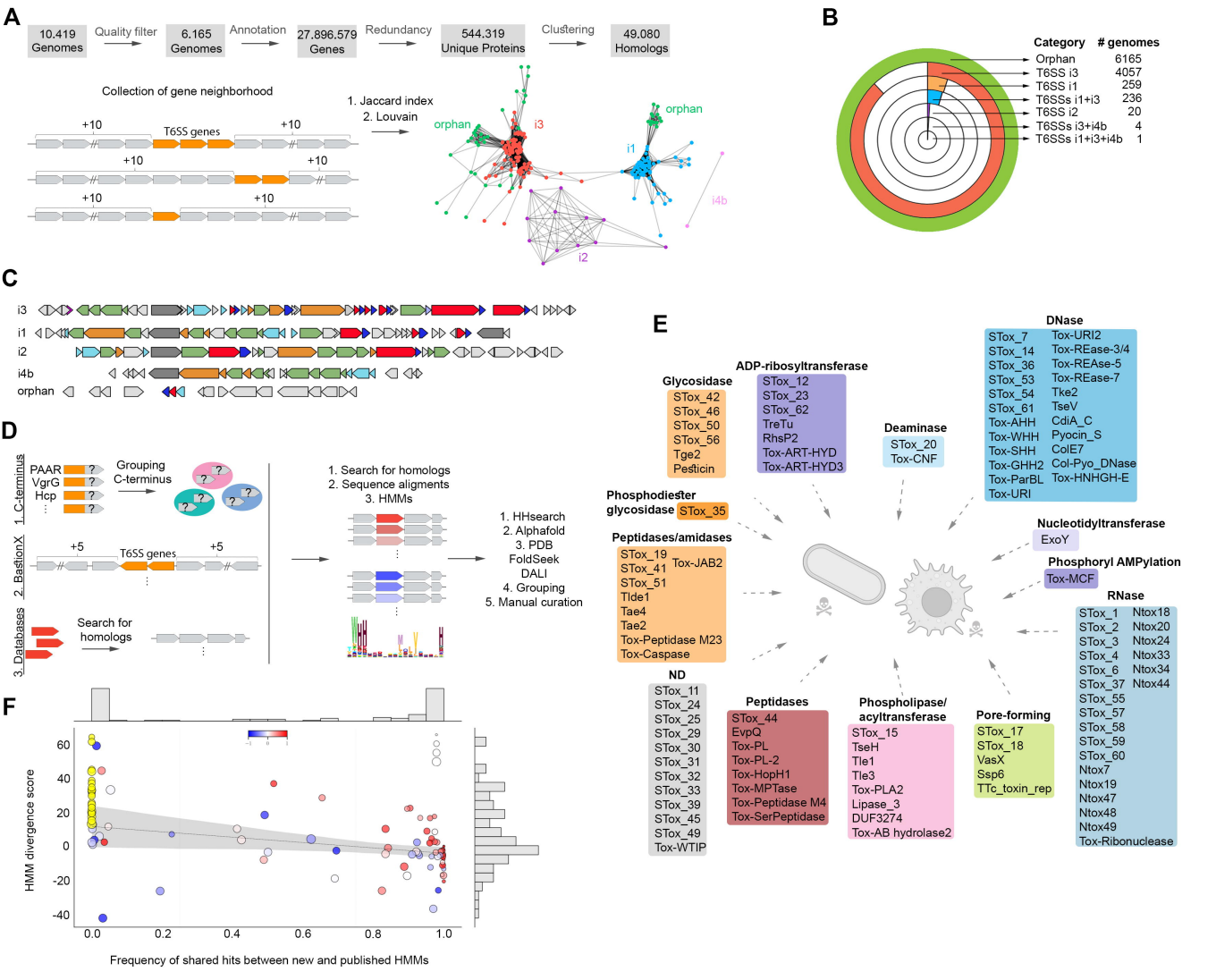
98 Belval, L. *et al.* A fast and simple method to eliminate Cpn60 from functional recombinant proteins produced by E. coli Arctic Express. *Protein Expr Purif* **109**, 29-34 (2015). https://doi.org/10.1016/j.pep.2015.01.009

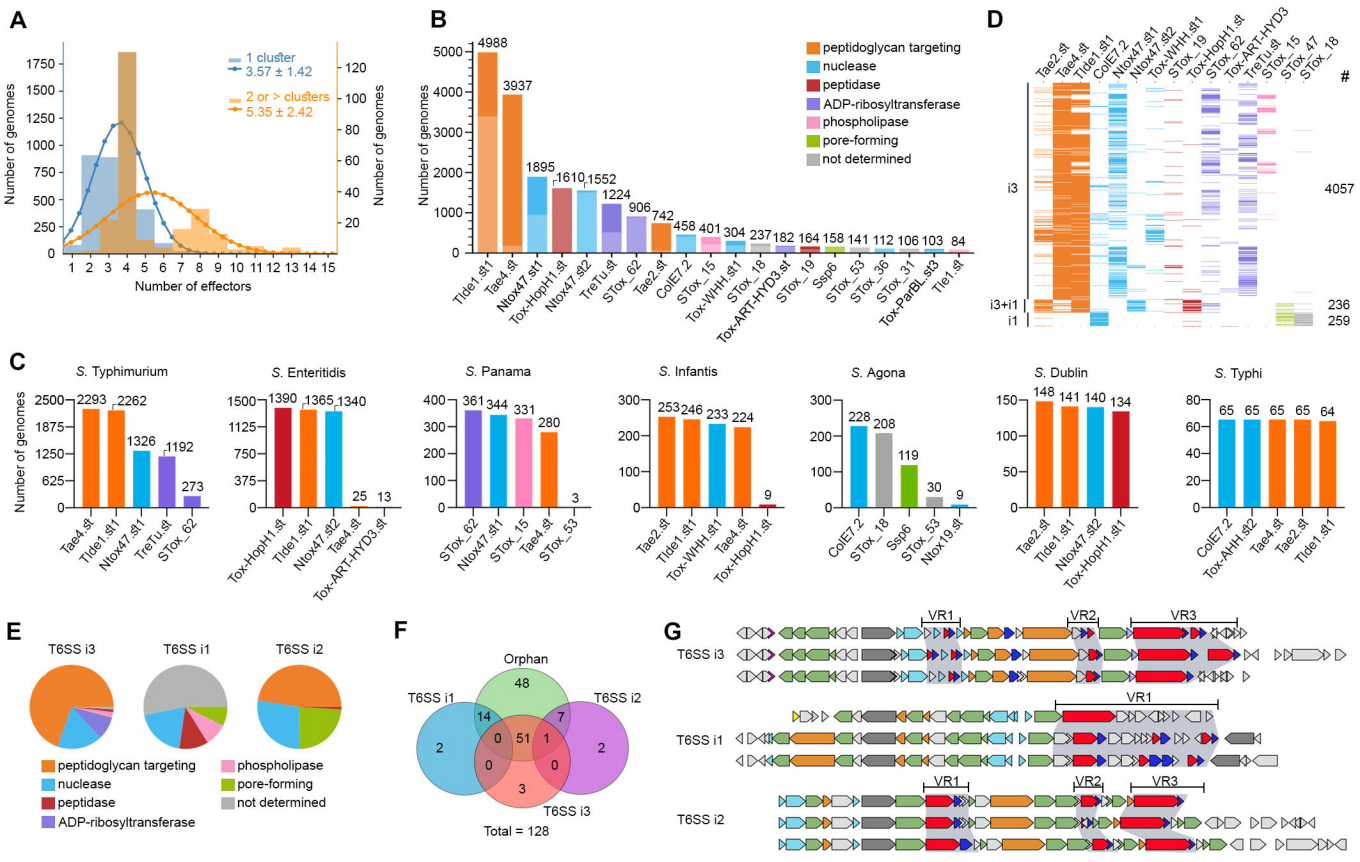
99 Yoshida, Y., Kodai, S., Takemura, S., Minamiyama, Y. & Niki, E. Simultaneous measurement of F2-isoprostane, hydroxyoctadecadienoic acid, hydroxyeicosatetraenoic acid, and hydroxycholesterols from physiological samples. *Anal Biochem* **379**, 105-115 (2008). https://doi.org/10.1016/j.ab.2008.04.028

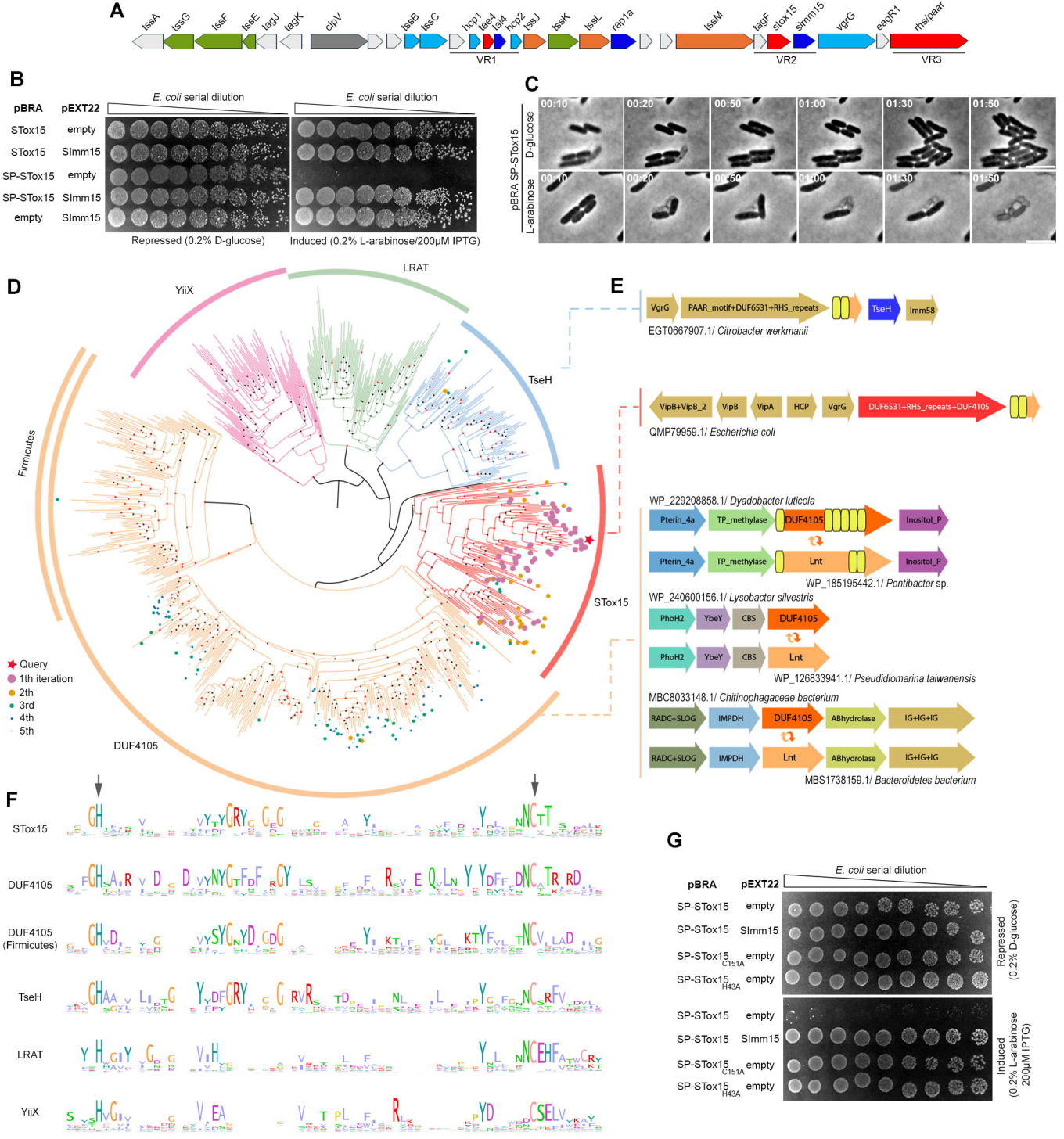
100 Chaves-Filho, A. B. *et al.* Alterations in lipid metabolism of spinal cord linked to amyotrophic lateral sclerosis. *Sci Rep* **9**, 11642 (2019). https://doi.org/10.1038/s41598-019-48059-7

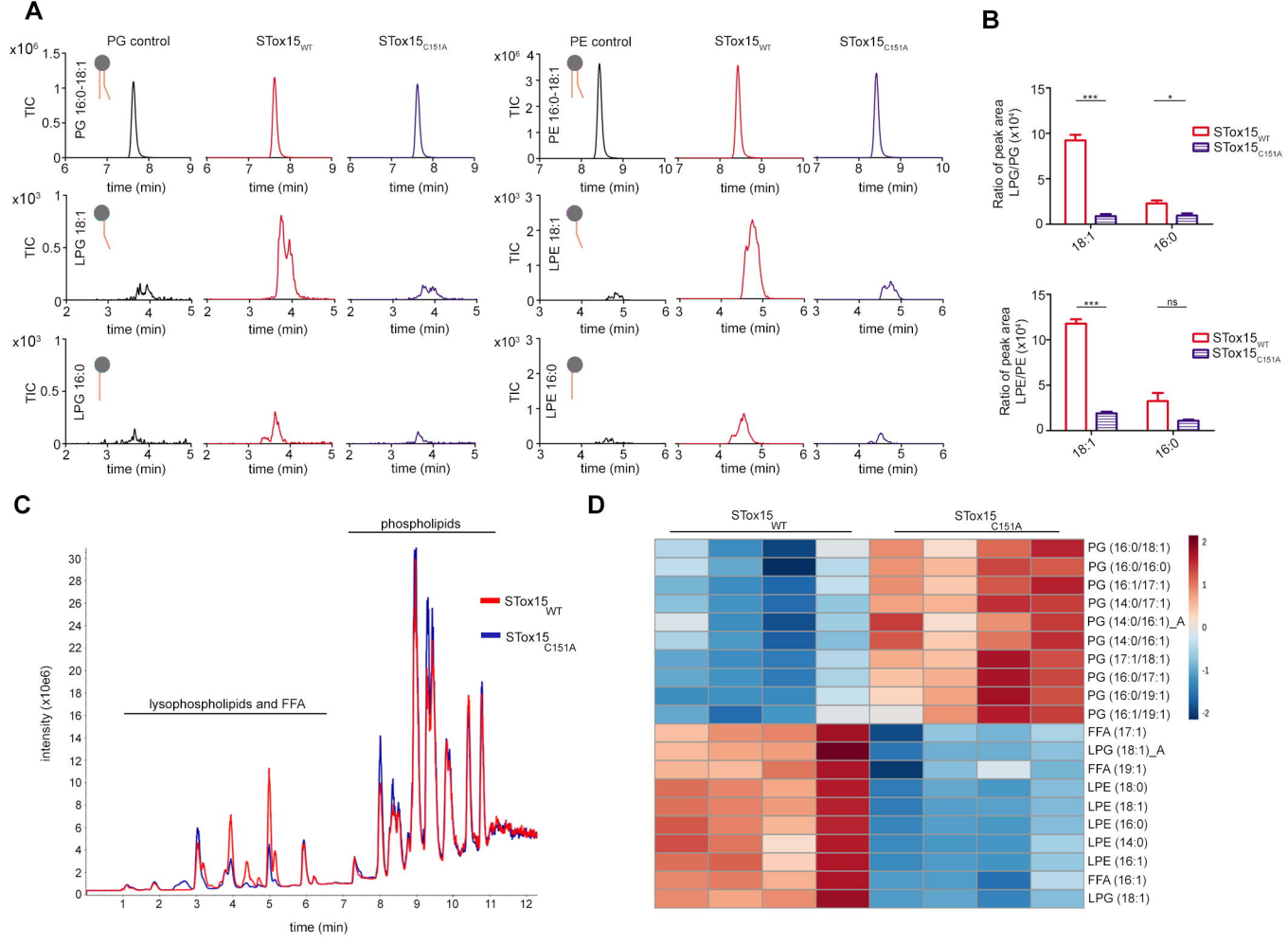
101 Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948 (2007). https://doi.org/10.1093/bioinformatics/btm404

102 Hallgren, J. *et al.* DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *bioRxiv*, 2022.2004.2008.487609 (2022). https://doi.org/10.1101/2022.04.08.487609


A
10.419 Genomes
Quality filter
6.165 Genomes
Annotation
27.896.579 Genes
Redundancy
544.319 Unique Proteins
Clustering
49.080 Homologs
Collection of gene neighborhood
+10
T6SS genes
1. Jaccard index
2. Louvain
orphan
i3
i1
i2
i4b
B
Category # genomes
Orphan 6165
T6SS i3 4057
T6SS i1 259
T6SSs i1+i3 236
T6SS i2 20
T6SSs i3+i4b 4
T6SSs i1+i3+i4b 1
C
i3
i1
i2
i4b
orphan
D
PAAR
VgrG
Hcp
1. C-terminus
Grouping C-terminus
2. BastionX
+5
T6SS genes
3. Databases
Search for homologs
1. Search for homologs
2. Sequence aligments
3. HMMs
1. HHsearch
2. Alphafold
3. PDB
FoldSeek
DALI
4. Grouping
5. Manual curation
E
Glycosidase
STox_42
STox_46
STox_50
STox_56
Tge2
Pesticin
ADP-ribosyltransferase
STox_12
STox_23
STox_62
TreTu
RhsP2
Tox-ART-HYD
Tox-ART-HYD3
Deaminase
STox_20
Tox-CNF
DNase
STox_7
STox_14
STox_36
STox_53
STox_54
STox_61
Tox-AHH
Tox-WHH
Tox-SHH
Tox-GHH2
Tox-ParBL
Tox-URI
Tox-URI2
Tox-REase-3/4
Tox-REAse-5
Tox-REase-7
Tke2
TseV
CdiA_C
Pyocin_S
ColE7
Col-Pyo_DNase
Tox-HNHGH-E
Phosphodiester glycosidase
STox_35
Peptidases/amidases
STox_19
STox_41
STox_51
Tlde1
Tae4
Tae2
Tox-Peptidase M23
Tox-Caspase
Tox-JAB2
Nucleotidyltransferase
ExoY
Phosphoryl AMPylation
Tox-MCF
RNase
STox_1
STox_2
STox_3
STox_4
STox_6
STox_37
STox_55
STox_57
STox_58
STox_59
STox_60
Ntox7
Ntox19
Ntox47
Ntox48
Ntox49
Tox-Ribonuclease
Ntox18
Ntox20
Ntox24
Ntox33
Ntox34
Ntox44
ND
STox_11
STox_24
STox_25
STox_29
STox_30
STox_31
STox_32
STox_33
STox_39
STox_45
STox_49
Tox-WTIP
Peptidases
STox_44
EvpQ
Tox-PL
Tox-PL-2
Tox-HopH1
Tox-MPTase
Tox-Peptidase M4
Tox-SerPeptidase
Phospholipase/ acyltransferase
STox_15
TseH
Tle1
Tle3
Tox-PLA2
Lipase_3
DUF3274
Tox-AB hydrolase2
Pore-forming
STox_17
STox_18
VasX
Ssp6
TTc_toxin_rep
F
HMM divergence score
Frequency of shared hits between new and published HMMs


A
1 cluster
3.57 ± 1.42
2 or > clusters
5.35 ± 2.42
Number of genomes
Number of effectors
B
peptidoglycan targeting
nuclease
peptidase
ADP-ribosyltransferase
phospholipase
pore-forming
not determined
Number of genomes
4988
3937
1895
1610
1552
1224
906
742
458
401
304
237
182
164
158
141
112
106
103
84
Tide1.st1
Tae4.st
Ntox47.st1
Tox-HopH1.st
Ntox47.st2
TreTu.st
STox_62
Tae2.st
ColE7.2
STox_15
Tox-WHH.st1
STox_18
Tox-ART-HYD3.st
STox_19
Ssp6
STox_53
STox_36
STox_31
Tox-ParBL.st3
Tie1.st
C
S. Typhimurium
S. Enteritidis
S. Panama
S. Infantis
S. Agona
S. Dublin
S. Typhi
Number of genomes
2293
2262
1326
1192
273
Tae4.st
Tide1.st1
Ntox47.st1
TreTu.st
STox_62
1390
1365
1340
25
13
Tox-HopH1.st
Tide1.st1
Ntox47.st2
Tae4.st
Tox-ART-HYD3.st
361
344
331
280
3
STox_62
Ntox47.st1
STox_15
Tae4.st
STox_53
253
246
233
224
9
Tae2.st
Tide1.st1
Tox-WHH.st1
Tae4.st
Tox-HopH1.st
228
208
119
30
9
ColE7.2
STox_18
Ssp6
STox_53
Ntox19.st
148
141
140
134
Tae2.st
Tide1.st1
Ntox47.st2
Tox-HopH1.st
65
65
65
65
64
ColE7.2
Tox-AHH.st2
Tae4.st
Tae2.st
Tide1.st1
D
Tae2.st
Tae4.st
Tide1.st1
ColE7.2
Ntox47.st1
Ntox47.st2
Tox-WHH.st1
STox_19
Tox-HopH1.st
STox_62
Tox-ART-HYD3
TreTu.st
STox_15
STox_47
STox_18
#
i3
4057
i3+i1
236
i1
259
E
T6SS i3
T6SS i1
T6SS i2
peptidoglycan targeting
nuclease
peptidase
ADP-ribosyltransferase
phospholipase
pore-forming
not determined
F
Orphan
T6SS i1
T6SS i2
T6SS i3
48
14
7
2
0
51
1
2
0
0
3
Total = 128
G
T6SS i3
T6SS i1
T6SS i2
VR1
VR2
VR3


A
tssA
tssG
tssF
tssE
tagJ
tagK
clpV
tssB
tssC
hcp1
tae4
tai4
hcp2
tssJ
tssK
tssL
rap1a
tssM
tagF
stox15
simm15
vgrG
eagR1
rhs/paar
VR1
VR2
VR3
B
E. coli serial dilution
E. coli serial dilution
pBRA
pEXT22
STox15
empty
STox15
SImm15
SP-STox15
empty
SP-STox15
SImm15
empty
SImm15
Repressed (0.2% D-glucose)
Induced (0.2% L-arabinose/200μM IPTG)
C
pBRA SP-STox15
D-glucose
L-arabinose
00:10
00:20
00:50
01:00
01:30
01:50
D
YiiX
LRAT
TseH
STox15
DUF4105
Firmicutes
Query
1th iteration
2th
3rd
4th
5th
E
VgrG
PAAR_motif+DUF6531+RHS_repeats
TseH
Imm58
EGT0667907.1/ Citrobacter werkmanii
VipB+VipB_2
VipB
VipA
HCP
VgrG
DUF6531+RHS_repeats+DUF4105
QMP79959.1/ Escherichia coli
WP_229208858.1/ Dyadobacter luticola
Pterin_4a
TP_methylase
DUF4105
Inositol_P
Pterin_4a
TP_methylase
Lnt
Inositol_P
WP_185195442.1/ Pontibacter sp.
WP_240600156.1/ Lysobacter silvestris
PhoH2
YbeY
CBS
DUF4105
PhoH2
YbeY
CBS
Lnt
WP_126833941.1/ Pseudidiomarina taiwanensis
MBC8033148.1/ Chitinophagaceae bacterium
RADC+SLOG
IMPDH
DUF4105
ABhydrolase
IG+IG+IG
RADC+SLOG
IMPDH
Lnt
ABhydrolase
IG+IG+IG
MBS1738159.1/ Bacteroidetes bacterium
F
STox15
DUF4105
DUF4105 (Firmicutes)
TseH
LRAT
YiiX
G
E. coli serial dilution
pBRA
pEXT22
SP-STox15
empty
SP-STox15
SImm15
SP-STox15 C151A
empty
SP-STox15 H43A
empty
Repressed (0.2% D-glucose)
SP-STox15
empty
SP-STox15
SImm15
SP-STox15 C151A
empty
SP-STox15 H43A
empty
Induced (0.2% L-arabinose 200μM IPTG)


A
PG control
STox15WT
STox15C151A
PE control
PG 16:0-18:1
PE 16:0-18:1
LPG 18:1
LPE 18:1
LPG 16:0
LPE 16:0
TIC
time (min)
B
Ratio of peak area LPG/PG (x10⁴)
Ratio of peak area LPE/PE (x10⁴)
18:1
16:0
***
*
ns
C
phospholipids
lysophospholipids and FFA
intensity (x10e6)
time (min)
D
PG (16:0/18:1)
PG (16:0/16:0)
PG (16:1/17:1)
PG (14:0/17:1)
PG (14:0/16:1)_A
PG (14:0/16:1)
PG (17:1/18:1)
PG (16:0/17:1)
PG (16:0/19:1)
PG (16:1/19:1)
FFA (17:1)
LPG (18:1)_A
FFA (19:1)
LPE (18:0)
LPE (18:1)
LPE (16:0)
LPE (14:0)
LPE (16:1)
FFA (16:1)
LPG (18:1)