



In silico Analyses of Core Proteins and Putative Effector and Immunity Proteins for T6SS in Enterohemorrhagic *E. coli*

Jaime Vazquez-Lopez and Fernando Navarro-Garcia*

Department of Cell Biology, Centro de Investigación y de Estudios Avanzados del IPN (CINVESTAV-IPN), Mexico City, Mexico

OPEN ACCESS

Edited by:

Roxane M. Piazza,
Butantan Institute, Brazil

Reviewed by:

Fernanda Fernandes Dos Santos,
Federal University of São Paulo, Brazil

Matthew L. Nilles,
University of North Dakota,
United States

Felipe Del Canto,
University of Chile, Chile

*Correspondence:

Fernando Navarro-Garcia
fnavarro@cell.cinvestav.mx

Specialty section:

This article was submitted to
Molecular Bacterial Pathogenesis,
a section of the journal
Frontiers in Cellular and Infection
Microbiology

Received: 28 January 2020

Accepted: 14 April 2020

Published: 05 May 2020

Citation:

Vazquez-Lopez J and
Navarro-Garcia F (2020) *In silico*
Analyses of Core Proteins and
Putative Effector and Immunity
Proteins for T6SS in
Enterohemorrhagic *E. coli*.
Front. Cell. Infect. Microbiol. 10:195.
doi: 10.3389/fcimb.2020.00195

Shiga-toxin-producing *Escherichia coli* (STEC) has become an important pathogen that can cause diarrhea, hemorrhagic colitis and hemolytic uremic syndrome (HUS) in humans. Recent reports show that the type VI secretion system (T6SS) from EHEC is required to produce infection in a murine model and its expression has been related to a higher prevalence of HUS. In this work, we use bioinformatics analyses to identify the core genes of the T6SS and compared the differences between these components among the two published genomes for EHEC O157:H7 strain EDL933. Prototype strain EDL933 was further compared with other O157:H7 genomes. Unlike other typical T6SS effectors found in *E. coli*, we identified that there are several *rhs* family genes in EHEC, which could serve as T6SS effectors. *In-silico* and PCR analyses of the differences between *rhs* genes in the two existing genomes, allowed us to determine that the most recently published genome is more reliable to study the *rhs* genes. Analyzing the putative tridimensional structure of Rhs proteins, as well as the motifs found in their C-terminal end, allowed us to predict their possible functions. A phylogenetic analysis showed that the orphan *rhs* genes are more closely related between them than the *rhs* genes belonging to *vgrG* islands and that they are divided into three clades. Analyses of the downstream region of the *rhs* genes for identifying hypothetical immunity proteins showed that every gene has an associated small ORF (129-609 nucleotides). These genes could serve as immunity proteins as they had several interaction motifs as well as structural homology with other known immunity proteins. Our findings highlight the relevance of the T6SS in EHEC as well as the possible function of the Rhs effectors of EHEC O157:H7 during pathogenesis and bacterial competition, and the identification of novel effectors for the T6SS using a structural approach.

Keywords: enterohemorrhagic *E. coli*, type 6 secretion system, effector proteins, Rhs, bacterial competition, pathogenesis

INTRODUCTION

Shiga-toxin-producing *Escherichia coli* (STEC) bacteria, including enterohemorrhagic *E. coli* (EHEC), are intestinal pathogens responsible for diseases such as enterohemorrhagic colitis and hemolytic uremic syndrome (HUS). HUS is a serious complication that may cause renal failure and heart damage, resulting in death in 10% of the affected (Croxen et al., 2013). Antibiotic treatment

against STEC infection increases HUS risk (Wong et al., 2000), and EHEC infectious dose is under 50 CFU (Tilden Jr et al., 1996), thus placing STEC as an important health risk for epidemic diseases. Cows are the principal STEC reservoir, in which the bacteria live as a commensal and do not cause infection, along with vegetables irrigated with contaminated waters, which have caused important outbreaks (Frank et al., 2011). Nowadays EHEC serotype O157:H7 and the hybrid strain between an enteroaggregative *E. coli* and STEC (EAEC/STEC O104:H4) are the most commonly associated strains to epidemic outbreaks (Yang et al., 2017). On one hand, EHEC O157:H7 has a locus of enterocyte effacement (LEE), which encodes a type 3 secretion system (T3SS) used for injecting effectors that subvert the cell physiology, causing the characteristic attaching and effacing (A/E) lesions (Schmidt, 2010). On the other hand, EAEC/STEC O104:H4 has the aggregative adherence fimbriae like AAFI as well as multiple antibiotic resistance genes (Rohde et al., 2011; Navarro-Garcia, 2014). Even if both serotypes have a different origin and have few virulence factors in common, both have the *stx2* gene, which encodes the Shiga toxin 2. The presence of this toxin in both genomes is probably due to horizontal gene transfer from a O157:H7 strain to EAEC strain 55989, the presumptive parental strain of EAEC/STEC O104:H4 (Rohde et al., 2011).

Both EHEC EDL933, an O157:H7 serotype used as a reference strain, and EAEC/STEC O104:H4, encode a type VI secretion system (T6SS) (Journet and Cascales, 2016), a nanomachine comprised from 13 core elements that make up for four defined structures: a membrane complex formed by TssJLM that attaches the T6SS to the inner membrane and makes the intermembrane pore; a baseplate comprised by TssEFGHK that connects the tail to the membrane complex and helps with the T6SS disassembly once it has contracted; a tail made up of stacked Hcp hexamers with a spike formed by a VgrG trimer; and a contractile sheath comprised by TssBC, which expel the Hcp tail to the exterior of the bacterial cell, while TssA regulates the T6SS assembly and contraction (Navarro-Garcia et al., 2019; Schneider et al., 2019). The core components are grouped in a genomic island and present homology with some of the components of the T4 bacteriophage machinery (Pukatzki et al., 2007; Leiman et al., 2009). The T6SS punctures other cell membranes, which can be from eukaryotic or prokaryotic organisms, and injects effectors directly into the cytoplasm that mediate bacterial competition and pathogenesis (Ma and Mekalanos, 2010; Basler et al., 2013). Effectors are usually classified into three families: (i) phospholipases that degrade the plasma membrane; (ii) murein-hydrolases and amidases that attack the cell wall; and (iii) nucleases which target DNA or RNA. The genes encoding those effectors are usually part of a bicistronic operon that also encodes an immunity protein, which binds to the effector protein and prevents auto-intoxication (Bingle et al., 2008). Finally, there are other effectors with non-canonical activities, including ADP ribosylation, ion chelation or even actin crosslinking (Lien and Lai, 2017).

Wan et al. (2017) have already shown that the T6SS of EHEC strain EDL933 is functional in an isogenic mutant for the master regulatory protein H-NS, and showed that EHEC can translocate KatN, a catalase effector, into the macrophage upon phagocytosis.

KatN neutralizes iROs present in the phagosome, thus avoiding bacterial killing. Wan et al. (2017) also showed that T6SS is essential to cause disease in a murine model, as isogenic mutants for *clpV* were unable to provoke death in the infected mice. ClpV is an ATPase that is essential for T6SS function by recycling the sheath components (Douzi et al., 2016). Nonetheless, mice death was independent of KatN, implying that other effectors acting as virulence factors in EHEC strain EDL933. Recently, Rhs family proteins have been associated with interbacterial competition in several STEC strains. Rhs family proteins are known for containing RHS repeats (recombination hotspot), often contain a PAAR domain that interacts with VgrG proteins (Bondage et al., 2016), and also have a C-terminal region with a nuclease, protease or deaminase activity (Ma et al., 2017b). In this work, we explore the differences between the T6SS genes and the putative effectors in two published genomes of EHEC O157:H7 strain EDL933 (Perna et al., 2001; Latif et al., 2014). Using bioinformatics analyses we could identify both the core genes of the T6SS as well as several putative effectors, and we provide evidence that the most-recently published genome (Latif et al., 2014) better represents the T6SS components and effectors. Additionally, analyzing sequence motifs and the modeled structure we identified several putative Rhs effectors, as well as the associated putative immunity proteins for each Rhs.

METHODS

Strains Used in This Study

EHEC strain EDL933 was kindly provided by Dr. Jose Luis Puente (Riley et al., 1983; Arenas-Hernández et al., 2014). The genomes bioinformatics comparison was made from EHEC strain EDL933 ATCC 700727 in the genome of Perna et al. (2001) (GenBank accession number: AE005174.2) and the genome of Latif et al. (2014) (GenBank accession number: CP008957.1). *E. coli* strain HB101 was obtained from ATCC 33694. The strains were grown overnight in LB plates with 1.5% agar at 37°C.

Bioinformatics Analyses

Motif searching for putative T6SS effectors was performed using the PROSITE, NCBI-CDD and Pfam databases (Marchler-Bauer et al., 2010; Sigrist et al., 2012; Finn et al., 2013) with the GenomeNet search engine. An *e-value* cutoff score of 0.01 was used. The tridimensional models were calculated using I-TASSER software (Yang and Zhang, 2015). For the analyses of the possible effector activity of the Rhs proteins, only the C-terminal end of each protein was modeled. The search for T6SS islands was performed using MacSyDB (Abby et al., 2016) and SecReT6 (Li et al., 2015). Relationship analyses between different Rhs proteins was conducted using STRING version 11 (Szklarczyk et al., 2019). Prediction of signal sequences was made using SignalIP (Juncker et al., 2003). Evolutionary analyses were conducted in MEGA X (Kumar et al., 2018) from the nucleotide sequences from the genome published by Latif et al. (2014), using the Maximum Likelihood method and Tamura-Nei model (Tamura and Nei, 1993). The tree presented is the consensus of 100 bootstrap repetitions (Felsenstein, 1985).

Primer Designing and Amplification for *rhsA* and *rhsC* Genes in EHEC Strain EDL933 DNA

Three primers were designed using Primer BLAST (Ye et al., 2012). The forward primer was designed to aligning with the *rhsA* gene in both genomes (*z5014* and *EDL933_4854*), while a genome-specific reverse primer was designed for each published genome (Perna et al., 2001; Latif et al., 2014). The sequence of the forward primer was 5'-CGCTATCTTTACGAC CCGCT-3'. The primer *rhsA1R* with the sequence 5'-GGC AAGGGGAATGGTCTAGG-3' was complementary with the *rhsA* gene (locus *z5014*) according to Perna et al. (2001), whereas the primer *rhsA2R* with the sequence 5'-GATGTG GGGGTACCATGCC-3' was complementary to the *rhsA* gene (locus *EDL933_4854*) according to Latif et al. (2014). For the amplification of the *rhsC* gene (locus *z0847*), the primers *rhsAF* and *rhsCR* (sequence 5'-TAGGCGTTTGTGGGTC TC-3') were used. The optimal aligning temperature for all the combinations were experimentally obtained by PCR using a temperature gradient and a final primer concentration of 0.4 μ M. The expected product was 800 bp for *rhsA1R* primer and 875 bp for *rhsA2R*, while a fragment of 832 bp fragment was expected for *rhsCR*. Chromosomal DNA from EHEC strain EDL933 was used as template. After PCR, the products were separated in a 1.5% agarose gel and then stained with ethidium bromide.

RESULTS AND DISCUSSION

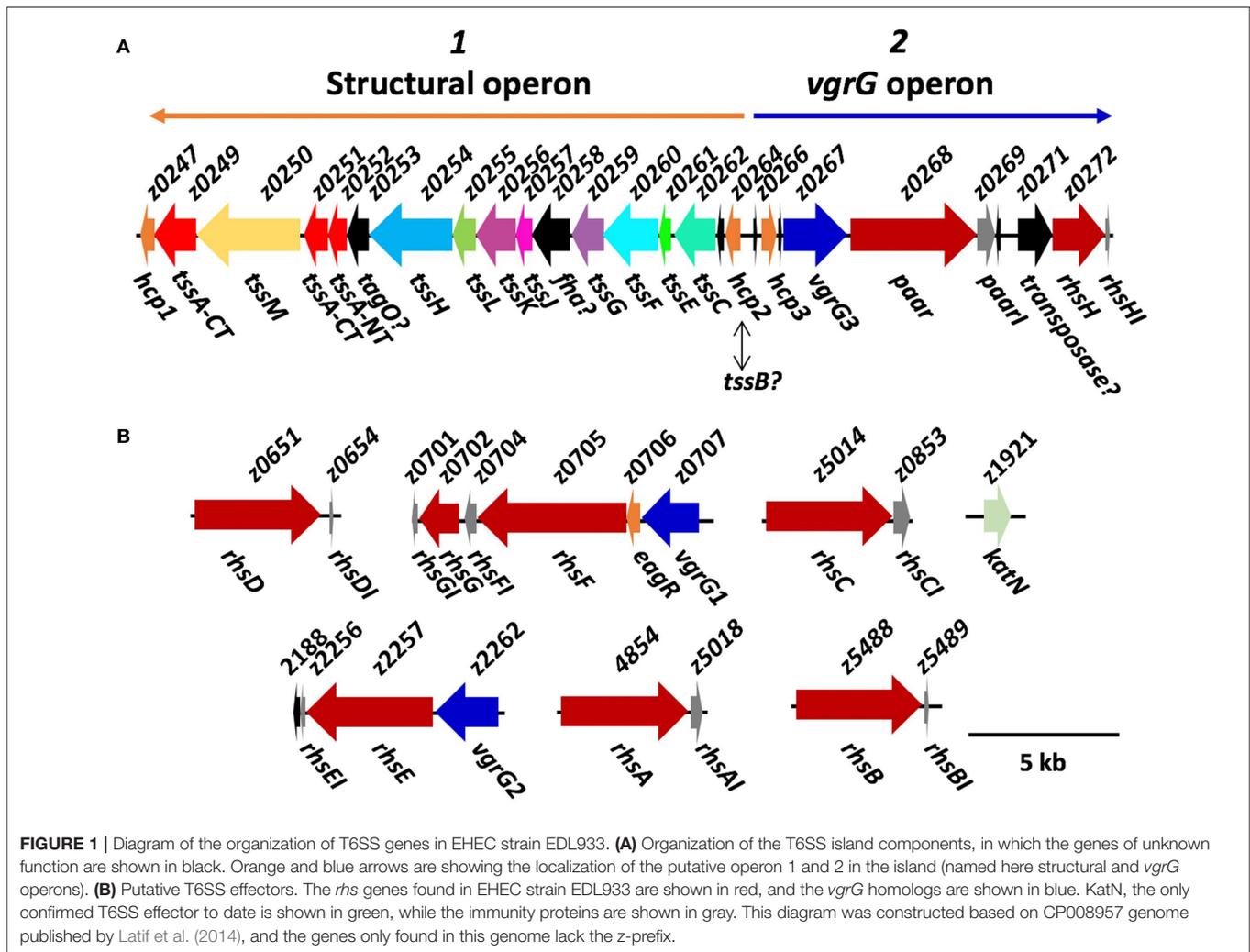
In-silico Analysis of the T6SS Island in EHEC Strain EDL933

Three distinct families of T6SS exist, which differ in their genetic organization as well as in the homology of their components (T6SS-1, T6SS-2 and T6SS-3) (Journet and Cascales, 2016). T6SS-1 and T6SS-3 have been associated with bacterial competition as well as with the invasive capacity of certain *E. coli* pathotypes, while T6SS-2 is commonly found in intestinal strains, including STECs like EHEC EDL933 and EAEC/STEC O104:H4 (Journet and Cascales, 2016). Wan et al. (2017) described the T6SS island in EHEC strain EDL933 using the genome published by Perna et al. (2001), which was obtained via shotgun sequencing as described by Mahillon et al. (1998). However, recent reports of structural proteins and the publication of a gapless genome without ambiguities for EHEC strain EDL933 using PacBio and Illumina sequencing (Latif et al., 2014) allowed us to perform a deeper analysis of the genetic island (Figure 1A). We found that the island encodes the 13 core components of the T6SS (TssA, TssBC, TssEFGHK, TssJLM, Hcp, and VgrG), which appears to be divided into two general operons: the first one (named here structural operon) would be codified on the complementary DNA strand and would contain both the baseplate components (TssEFGHK) and the membrane complex components (TssJLM). Besides, this operon includes two *hcp* genes (*tssD*, loci *z0247*, and *z0264*) and a *tssC* gene, an essential component of the T6SS sheath. Analyzing the genomes from O157:H7 strains in GenBank, we found that 137 strains out of 145 possessed the T6SS-2 structural genes and only 8 did

not (strains 09BKT002497, 121, 262, 611, EC10, EC4024, GZ-021210/cattle, and ZAP430). No significant similarity with T6SS-1 or T6SS-3 structural genes from EAEC 042, a reference strain for these T6SS families, was found in O157:H7 genomes. The second operon (named here *vgrG* operon) would be on the main DNA strand containing a group of genes that encode the tail components such as another *hcp* gene (*z0266*) and a *vgrG* gene (*tssI*, *z0267*). Additionally, in this operon, we found two genes downstream *vgrG*, and one of them encodes an Rhs family protein with a PAAR domain (*z0268*). Interestingly, both the *rhs* family gene and the gene directly downstream (*z0269*) were not found in every T6SS-2 islands, such as EAEC 042, EAEC 55989, or EAEC/STEC O104:H4, as previously reported (Journet and Cascales, 2016). This suggests that this pair of genes encode a T6SS effector as well as its associated immunity protein. Finally, *z0268* and *z0269* were found only in 66 EHEC O157:H7 strains from the lineage I and were absent in 79 strains from lineage II (Zhang et al., 2007).

One particularity of the T6SS island in EHEC strain EDL933 is that it contains three homologs of *hcp* genes (*z0247*, *z0264* and *z0266*) (Wan et al., 2017), so we decided to further analyze them. *hcp1* (*z0247*) is not recognized as an ORF in Latif et al. (2014), even though that region is 100% identical to Perna et al. (2001). This might be because there is a little overlapping between the 5' end of locus *z0247* and 3' end of locus *z0249*, so perhaps the region in this latter analysis was not properly recognized by the annotation software, RAST. For *hcp2* (*z0264*), it has already been shown that EHEC Δ *hns*, an isogenic mutant for the master regulatory protein H-NS repressing the T6SS expression in several bacteria, secretes the Hcp2 protein to the supernatant (Wan et al., 2017). However, upon inspecting the motifs found in the protein, we found a PF05591 motif, which may be found both in Hcp2 and TssB. Remarkably, the island reported by Wan et al. (2017) lacks the *tssB* gene, an essential component of the T6SS (Zhang et al., 2014). Our analyses using MasyDB and SecReT6 (Li et al., 2015; Abby et al., 2016) suggested that the gene is not *hcp* but *tssB*, but experimental data is required to completely clarify the identity of this gene. Finally, none of the three *hcp* genes belongs to the Hcp-ET family reported by Ma et al. (2017a), and no known effector domains fused with the Hcp core were found that might participate during pathogenesis or interbacterial competition in EHEC strain EDL933.

Furthermore, *tssA* gene (*z0251*) is half as long as the reported genes in other strains, and after inspecting the region between the loci *z0251* and *z0252* (Perna et al., 2001), we found that the gene seems to be interrupted, as a missing cytosine in the 249,998th nucleotide producing a stop codon. In fact, in the genome published by Latif et al. (2014), the segment containing *z0251* and *z0252* is recognized as a single pseudogene, even though this region is 100% identical as in the genome published by Perna et al. (2001). Being TssA an essential protein for the T6SS function, it would be odd that the gene was mutated, especially when the functionality of the T6SS has already been confirmed in this strain (Wan et al., 2017). Interestingly, Zoued et al. (2016) demonstrated that both N-terminal and C-terminal regions of TssA interact with different components of the T6SS, and it is possible that, even if the gene is divided in two,



both regions could properly oligomerize and perform TssA function during T6SS biogenesis. About a third of the O157:H7 strains we analyzed showed the same single-nucleotide deletion, while the rest encoded a full length TssA protein. Additionally, *z0249* also encodes a protein of the cluster of differentiation 3515 (COG3515) that would also correspond to the C-terminal fraction of TssA, although the identity between Z0251 and Z0249 is rather low (41.67% and 7% of query cover). The TssA C-terminal region forms dodecamers, suggesting that the gene could be duplicated on the genome, but the low homology could imply that both loci perform different activities during T6SS biogenesis and/or function.

Outside of the T6SS island, we found two islands that encode VgrG proteins (*vgrG* islands, **Figure 1B**), which Uchida et al. (2013) renamed *vgrG1* (*z0707*) and *vgrG2* (*z2262*), while the locus *z0267* inside the T6SS island was named *vgrG3*. Interestingly, downstream of *vgrG1*, we found an hypothetical protein with a domain of unknown function 1795 (DUF1795), which has been found to interact with Rhs and VgrG proteins in *Serratia marcescens* (Alcoforado Diniz et al., 2015), and is

essential for T6SS-mediated translocation of Rhs proteins in several STECs (Ma et al., 2017b). The three VgrG proteins have $\geq 90\%$ amino acid identity and resemble other VgrG proteins of the T6SS-2. Interestingly, the same 8 STEC O157:H7 strains that were negative for structural genes were also negative for *vgrG* genes, while the other possessed at least one *vgrG* homolog in their genome, having up to five copies of *vgrG*.

Several T6SS effectors are VgrG proteins that have a C-terminal extension with enzymatic activity (Ma and Mekalanos, 2010; Suarez et al., 2010; Brooks et al., 2013; Sana et al., 2015). However, none of the VgrG homologs in EHEC O157:H7 strains had a known effector domain in its C-terminal, though we found that the C-terminal region was highly divergent while the N-terminal region was conserved. T6SS-translocated proteins are carried by VgrG proteins, and those effectors usually interact with the C-terminal region of VgrG (Hachani et al., 2014; Whitney et al., 2014). Thereby, we postulate that the three different VgrG proteins present in EHEC O157:H7 strains carry different effectors to the prey cell.

Wan et al. (2017) already demonstrated that T6SS in EHEC EDL933 is functional in an *hns* isogenic mutant and that it can translocate the KatN catalase upon macrophage phagocytosis, thus neutralizing the iROs in the phagosome and avoiding bacterial death. They also showed that the T6SS is essential to cause disease in a murine model, as the isogenic *clpV* mutants were unable to provoke death in mice. Nonetheless, mice death was independent of *katN*, so there must be other unidentified effectors that could contribute to EHEC strain EDL933 pathogenesis.

Identification of Putative Effectors Using a Bioinformatics Approach

Each T6SS family has characteristic effectors according to its own functionality, and even if we could not find a single O157:H7 strain with a T6SS-1 or T6SS-3, we decided to look for effectors of any T6SS family in EHEC O157:H7 trying to determine the role of the secretion system. The commonly reported T6SS effectors in *E. coli* were not found in the genomes of EHEC O157:H7 available to date. EHEC O157:H7 lacks the effectors of the VT1-VT5 families described previously (Ma et al., 2018), as well as effectors with DUF2169, DUF4123 or DUF2345 domains. Proteins with the characteristic GX SXG or HXXXXXXD motifs were also absent (Lien and Lai, 2017). On the other hand, after identifying a PAAR domain-containing protein in the T6SS island of EHEC strain EDL933, we looked for other PAAR domain-containing proteins in EHEC strain EDL933 genome and found seven genes that encode these proteins. PAAR proteins have been found in T6SS-associated effectors in *Pseudomonas aeruginosa* and *Vibrio*

cholerae (Shneider et al., 2013), in EHEC strain EDL933, they also contained RHS (Recombination Hotspot) repeats, so they were classified as Rhs family proteins. In recent years, several T6SS-effectors related to Rhs proteins have been described in UPEC and STEC strains (Poole et al., 2011; Ma et al., 2017b), as well as in other species such as *P. aeruginosa* and *S. marcescens* (Kung et al., 2012; Diniz and Coulthurst, 2015). By using the I-TASSER software, we modeled the tridimensional structure of Z0268 (PAAR) and found that this protein resembled a shell-like structure built up by β -sheets that enclose a C-terminal region with enzymatic activity (Figure 2A). The N-terminal and central region of PAAR-Rhs proteins is present in both Gram-negative and Gram-positive bacteria, and there are even homologous proteins in eukaryotes, known as teneurins. The structure of the extracellular domain of teneurins is similar to Rhs proteins, and their presence in eukaryotes was probably due to horizontal gene transfer (Jackson et al., 2018). Teneurins bind to latrophilins and mediate cell-cell adhesion (Cruz-Ortega and Boucard, 2019), while the Rhs proteins in bacteria seem to mediate intercellular competition (Koskiniemi et al., 2013).

The C-terminal region of Rhs proteins frequently share homology with several bacterial toxins and the studied Rhs effectors usually fall into the canonical categories (nucleases, lipases, and murein-hydrolases). Ma et al. (2017b) showed that the PAAR domain was required for the translocation of Rhs proteins via T6SS. The PAAR domain interacts with the C-terminal end of VgrG for its transport to the cytoplasm of the prey cell, where the Rhs C-terminal exerts its functional role (Busby et al., 2013; Shneider et al., 2013).

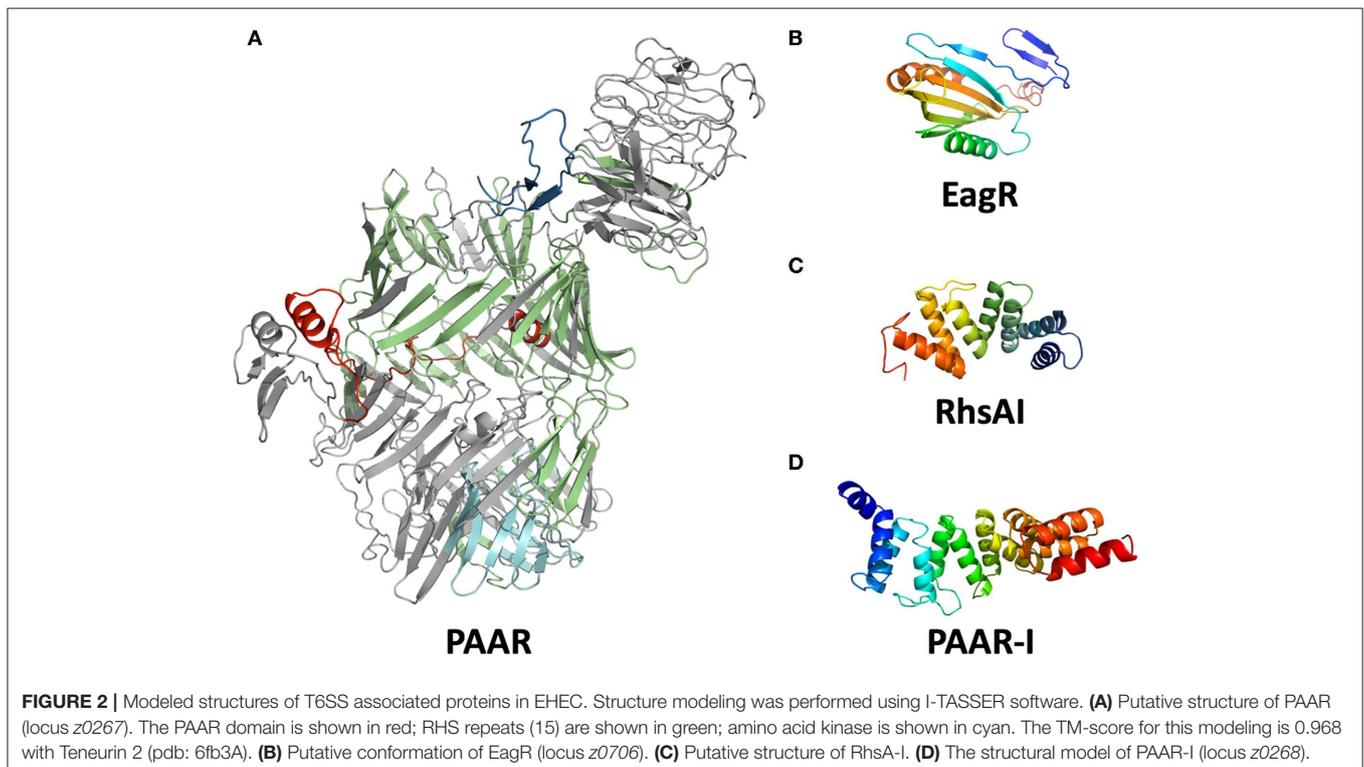


TABLE 1 | Comparison between Rhs proteins from two available genomes in EHEC EDL933.

Proposed name	Loci from Perna et al.	GenBank ID	Loci from Latif et al.	Genbank ID	Notes
RhsA	<i>z5014</i>	AAG58737	<i>EDL933_4854</i>	AIG70995	Fused with <i>z5017</i>
RhsB	<i>z5488</i>	AAG59133	<i>z5488</i>	AIG71408	99.7% Identical
RhsC	<i>z0847</i>	AAG55021	<i>z0847</i>	AIG66972	Fused with <i>z0851</i>
RhsD	<i>z0651</i>	AAG54854	<i>z0651</i>	AIG66789	100% Identical
RhsE	<i>z2257</i> <i>z2259</i> <i>z2261</i>	AAG56314 AAG56315 AAG56316	<i>z2257</i> <i>z2259</i> <i>z2261</i>	AIG68376	<i>z2257</i> - <i>z2261</i> are fused
RhsF	<i>z0705</i>	AAG54900	<i>z0705</i>	AIG66446	100% Identical
RhsG	<i>z0702</i>	AAG54898	<i>z0702</i>	AIG66832	100% Identical
RhsH	<i>z0272</i>	AAG54539	<i>z0272</i>	AIG66830	100% Identical
PAAR	<i>z0268</i>	AAG54536	<i>z0268</i>	AIG66450	100% Identical

In EHEC strain EDL933, we found several genes of the Rhs family, but these genes differ between both available genomes; 9 loci (Latif et al., 2014) vs. 11 loci (Perna et al., 2001) (Table 1). Although *z2488*, *z0651*, *z0268*, *z0272*, *z2257*, and *z0705* genes were identical in both genomes, we found several important differences. In the first genome published, for example, locus *z5014* encodes an Rhs protein with a PAAR motif that we identified as a putative DNase domain in its C-terminal end, while in the genome published by Latif et al. (2014), this C-terminal end was not a DNase and contained the sequence corresponding to locus *z5017* (Figures 3A,B). We also found that the protein encoded by locus *z5014* had different C-terminal regions, whereas locus *z2257* from Latif et al. (2014) genome contained regions from loci *z2256*, *z2257*, and *z2259* (Table 1), as there were 20 gaps and 42 mismatches between this region in both genomes. A possible reason for these discrepancies is that the genome published by Perna et al. (2001) might have assembly mistakes in the regions that encode Rhs proteins due to their size and the presence of multiple homologs in the genome. In fact, upon analysis of *rhs* genes in EHEC O157:H7 strain Sakai, we found that they were practically identical to those in EHEC strain EDL933 according to the genome published by Latif et al. (2014), both in the number of loci as in their sequences, including the C-terminal encoding region.

To test which of the two published genomes represented better the *rhs* genes in EHEC strain EDL933, we decided to amplify a region of an *rhs* gene present in EHEC EDL933 that should correspond to the same sequence in the different genomes reported; in Perna et al. (2001), it would be the locus *z5014* and in Latif et al. (2014) the locus *EDL933_4854*. Locus *z5014* from the genome published by Perna et al. (2001) contains a 3' end

that does not correspond to any other sequence from the genome published by Latif et al. (2014), while in this latter genome the locus *z5017* is fused to the 3' region of *z5014*, here named *EDL933_4854* (Figure 3A). Thus, we designed a forward primer (*rhsAF*) that could serve for both genomes, as well as a reverse primer (*rhsA1R*) specific for the 3' region of *z5014* in Perna et al. (2001) that would produce an 800 bp amplicon. We also designed another reverse primer (*rhsA2R*) that aligned with locus *z5017* from Perna et al. (2001) and locus *EDL933_4854* from Latif et al. (2014), which would produce a 2.6 kb long amplicon and an 875 bp long amplicon, respectively (Figure 3A). The reverse primer designed for the genome published by Perna et al. (2001) (*rhsA1R*) did not amplify the expected DNA fragment (800 bp) in EHEC EDL933 (Figure 3B), while the reverse primer for the genome published by Latif et al. (2014) (*rhsA2R*), amplified the 875 bp fragment from EHEC strain EDL933. To test if the reverse primer designed for the genome published by Perna et al. (2001) was reliable, we used *E. coli* strain HB101 as a positive control, since this strain has an *rhs* gene identical to the locus *z5014* reported by Perna et al. (2001). The amplification of an 800 bp fragment only in strain HB101 using *rhsA1R* primer (Figure 3B) suggested that EHEC EDL933 does not harbor the sequence reported by Perna et al. (2001). Using the same strategy (see Figure 3A), we analyzed the *rhsC* gene region, another chromosomal region with differences between both genomes. The region of the *rhsC* gene was amplified using the *rhsAF* primer and a specific primer (*rhsCR*) that would give a 2.7 kb fragment for the genome of Perna et al. (2001) and a 832 bp fragment for the genome of Latif et al. (2014). Similar to *rhsA*, a fragment of 832 bp was obtained both in EHEC EDL933 as in *E. coli* strain HB101, while the 2.6 kb fragment was not obtained in any strains. In conclusion, the genome published by Latif et al. (2014) better represents the *rhs* genes in EHEC strain EDL933, so we recommend using this genome in posterior analyses. In this genome, we found 9 genes of the *rhs* family, of which 7 are long (>4 kb) and have a PAAR domain, while the two remaining are relatively small (<1 kb) and do not have a PAAR domain. Because the locus *EDL933_4854* was similar to the *rhsA* gene from *E. coli* strain K-12, we named this gene as *rhsA*, named the other loci as *rhsB-H* according to their sequence similarity with genes in *E. coli* K-12, and named *pair* the *rhs* gene that is encoded inside the T6SS island to differentiate it from the others.

Once we determined which genome was better for the study of the Rhs family proteins, we wanted to elucidate the possible role of each protein, so we analyzed the C-terminal end of each Rhs looking for functional domains, and for effectors structural homology using I-TASSER software. We first analyzed the RhsA protein (locus *EDL933_4854*) and did not find any known motif, but after modeling its structure with I-TASSER we found that the C-terminal region had structural similarity with the toxin-A of *Clostridium difficile*, a toxin that inactivates GTP-binding proteins such as Rho, Rac, and Cdc42 (Voth et al., 2019) (pdb: 3HO6, TM-score: 0.73). The C-terminal end of RhsB (*z5488*) was structurally similar with the human proteasome 26S lid, which recognizes ubiquitinated proteins (Schweitzer et al., 2016) (pdb: 514kV, TM-score: 0.694), though we found

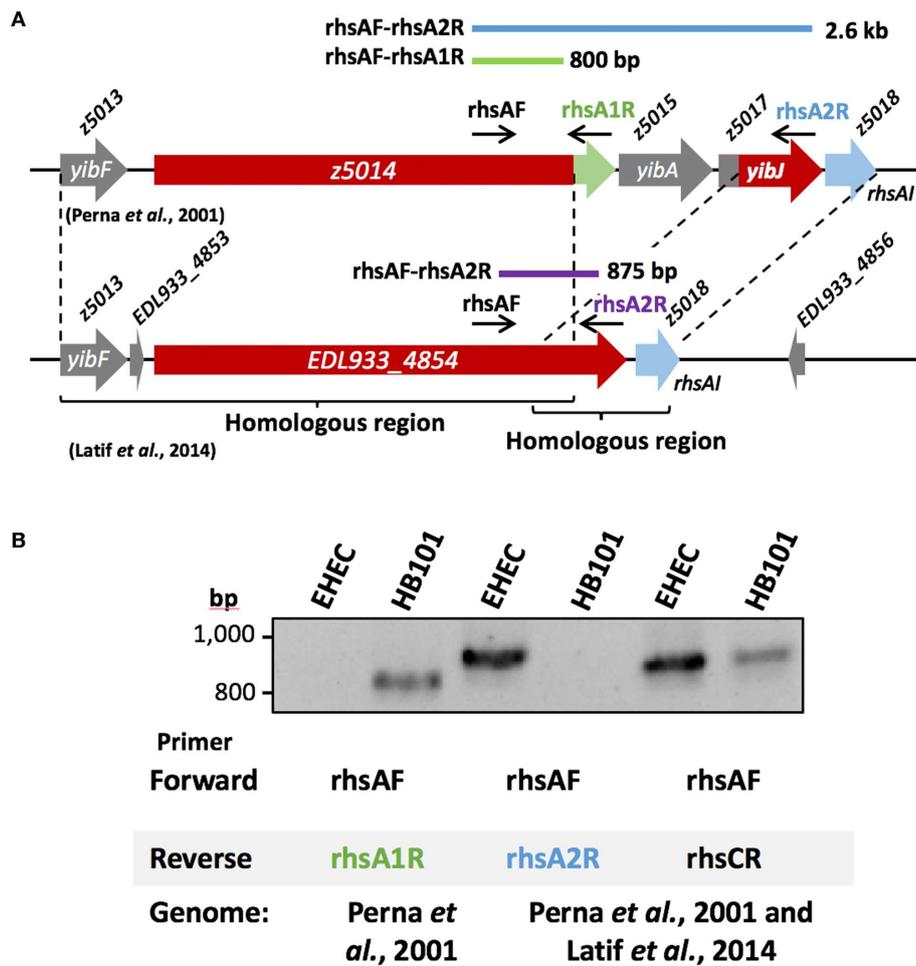


FIGURE 3 | Gene amplification of *rhsA* in EHEC strain EDL933. **(A)** Diagram of the region between loci *z5013* and *z5018* according to Perna et al. (2001) and according to Latif et al. (2014). The *rhs* gene (red) has a different C-terminal end in each genome, harboring a DNase domain (green) in and the sequence corresponding to locus *z5017* (red). **(B)** The 800 bp amplicon for *rhsA* gene (locus *z5014*) was not generated in EHEC strain EDL933 using the primer pair *rhsAF-rhsA1R* designed for the gene according to Perna et al. (2001). *E. coli* strain HB101 was used as a positive control for these primers (*rhsAF-rhsA1R*), which generated an 800 bp amplicon. Whereas, when the primer pair *rhsAF-rhsA2R* designed based on Perna et al. (2001) and Latif et al. (2014) was used, the expected 2.6 kb fragment was not obtained, but the 875 bp was, as expected from the genome from Latif et al. (2014). By using a specific primer (*rhsCR*) for *rhsC*, a 832 bp fragment was obtained, as expected from the genome of Latif et al. (2014), while the 2.7 kb fragment from the genome of Perna et al. (2001) was not obtained neither in EHEC EDL933 nor in *E. coli* strain HB101. A 63°C temperature was used for **(B)**. The PCR products were separated in an 1.5% agarose gel and stained using ethidium bromide.

a DUF4329 motif (PF05593, e -value: 5×10^{-13}) that according to Ma et al. (2017b) is an RNase domain. The C-terminal end of RhsC (*z5014*) had structural similarity with VipE protein of *Legionella pneumophila*, a protein that inhibits the vesicular traffic in yeast (Shohdy et al., 2005) (pdb: 4qn8A, TM-score: 0.574). The C-terminal region of RhsD (*z0651*) was structurally homologous with the human serum amyloid A1 protein (Lu et al., 2014) (pdb: 4ip8A, TM-score: 0.623), which participates in the inflammatory response, chemotaxis and opsonization upon binding with several integrins and G protein-coupled receptors (GPCRs) (Niemi et al., 2011). The C-terminal end of RhsE (*z2257*) had structural homology with the RsaA protein

of *Caulobacter crescentus* (Bharat et al., 2017) (pdb: 5n8pA, TM-score: 0.71), as well as the adhesin PfbA of *Streptococcus pneumoniae* (Suits and Boraston, 2013) (pdb: 3zpp, TM-score: 0.647). Both RsaA and PfbA are present in the S-layer of bacteria and could serve as adhesins. Finally, the C-terminal end of PAAR (*z0268*) presented homology with the cysteine-protease domain of the RtxA toxin of *V. cholerae* (Prochazkova et al., 2009) (pdb: 3fzyA, TM-score: 0.681), a toxin that crosslinks actin and actively participates during the pathogenesis of *V. cholerae*. Nonetheless, the C-terminal end did not have a RTX motif nor the actin cross-linking domain normally found in the RtxA toxin (Boardman et al., 2007) (Table 2).

TABLE 2 | Putative function of Rhs effectors in EHEC strain EDL933.

Locus	Domain homolog	TM-score	Homolog function
EDL933_4854 (rhsA)	Toxin A from <i>C. difficile</i>	0.73	Small-GTPase inhibition (like Rho and Rac)
z5488 (rhsB)	Proteasome lid 26S	0.694	Protein ubiquitination
z5014 (rhsC)	VipE	0.574	Vesicular-traffic inhibition
z0651 (rhsD)	human serum amyloid A1 protein	0.623	Chemotaxis, inflammatory response
z2257 (rhsE)	RsaA from <i>Caulobacter crescentus</i>	0.71	S-layer Adhesin
z0268 (paar)	Toxin RtxA from <i>V. cholerae</i>	0.681	Cysteine protease, actin-crosslinking
z0705 (rhsF)	Capsid protein from betanodavirus	0.863	Capsid protein

RhsF protein (z0705) did not harbor any apparent motif, but the structural modeling showed that the C-terminal end resembled the structure of capsid proteins present in viruses such as betanodavirus (pdb: 3jbmA, TM-score: 0.863) or Orsay-like virus (pdb: 4nwvA, TM-score: 0.825). Just downstream of the *rhsF* gene, we found another *rhs* gene (*rhsG*, locus z0702). RhsG protein had a metallopeptidase-4 domain (PF15640, *e-value*: 1×10^{-32}), which has already been described in a similar STEC strain (Ma et al., 2017b), but lacked the PAAR domain in EHEC strain EDL933 and it seems to be interrupted (this region is identical in both published genomes). The presence of a C-terminal region similar to capsid proteins has not been reported previously, and upon inspecting the 3' region of the *rhsF* gene and the linker sequence between *rhsF* and *rhsG*, we found that both sequences shared similarity with Pithovirus LCDPAC02. This suggests that the *rhsG* gene was originally fused with *rhsF*, but a viral sequence was inserted in this region and separated the gene. It is unclear if the C-terminal of the RhsF protein has a function related to interbacterial competition or pathogenesis. To corroborate the predicted functions of the Rhs proteins as T6SS effectors, a translocation assay is needed, and this becomes complicated due to the current difficulty to activate the T6SS *in vitro*. As there are several methodologies to achieve expression of the T6SS *in vitro* (Gueguen and Cascales, 2013; Miyata et al., 2013; Wan et al., 2017), co-expression of the Rhs proteins and the T6SS components needs to be achieved to properly investigate their function. Heterologous expression can help to study the effector function, and murine infection models with isogenic mutants for these genes could help to establish the importance of the Rhs proteins during intestine colonization and pathogenesis.

The genes *rhsE*, *rhsF*, and *paar* were located downstream of a distinct *vgrG* gene (Figure 1), suggesting that these Rhs proteins are translocated by a its specific VgrG protein. On the other hand, *rhsA*, *rhsB*, *rhsC*, and *rhsD* were orphan genes that did not belong to a *vgrG* island, so presumably, they can be translocated by any VgrG protein. If each VgrG protein translocates specific Rhs proteins, then it would be logical to think that the *rhs* genes are divided in three categories according to the number of *vgrG*

genes in EHEC EDL933. Using phylogenetic analysis, we found that *rhs* genes were indeed divided into three clades (Figure 4). The first clade grouped *rhsA*, *rhsB*, and *rhsC* genes being the last two more closely related. The second clade contained *rhsD*, *rhsE*, and *paar* genes, being the latter one more dissimilar of them. Finally, *rhsF* gene was clearly different from the others, being the only member of its clade. While *rhsF* is adjacent to *vgrG1*, it would seem that VgrG1 translocates RhsF, however, both *rhsE* and *paar* are in the same clade and are adjacent to *vgrG2* and *vgrG3*. Besides, the first clade containing *rhsA*, *rhsB*, and *rhsC* did not include any *rhs* gene encoded around a *vgrG* island, so it is unclear which VgrG protein, if any, translocates these proteins. Ma et al. (2017b) claimed that, in STEC, the *rhs* genes encoded in the vicinity of the EagR chaperone are phylogenetically distinct from the orphan genes (*rhs* genes not encoded in an *eagR* island), which is consistent in EHEC EDL933, as *eagR* was directly upstream of the *rhsF* gene, the less conserved gene in EHEC EDL933. When we modeled the structure of EagR (locus z0706) (Figure 2B), we found that the protein resembles PA0094 (TM-score: 0.888) from *P. aeruginosa*, whose crystal structure was obtained by Osipiuk et al. (unpublished, pdb: 1TU1). PA0094 acts as a chaperone for the Tse6 (PA0093), a PAAR-domain containing protein that also has a Ntox46 domain with nuclease activity. The high structural homology between PA0094 and EagR suggests that the EagR protein also functions as a chaperone for RhsF in EHEC EDL933. Nevertheless, because of the absence of a known motif in RhsF C-terminus, more studies are needed to determine if this island and its encoded proteins are indeed functional.

We have found that *rhs* genes are widely distributed among bacteria through data base analyses. By searching all EHEC O157:H7 genomes in GenBank, we found that there are some strains having up to 11 of these genes. Other studies have already classified PAAR-Rhs proteins as T6SS effectors, being mostly nucleases or peptidases (Koskiniemi et al., 2013; Ma et al., 2017b). The RNase motif present in Z5014 from the genome published by Perna et al. (2001) (PF15606) is not present in other Rhs proteins from O157:H7 strains, suggesting that this motif is not present in EHEC O157:H7. However, in EHEC O157:H7 strains there are 9 PAAR-Rhs proteins with a DUF4329 that has a RNase activity, including RhsB from EDL933 as showed above ($e \leq 0.01$). Interestingly, 7 PAAR-Rhs proteins in EHEC O157:H7 strains contain a serum amyloid A protein motif in their C-terminal region (PF00277; $e \leq 0.01$). As we discussed earlier, this motif could play a role in the immunological response but experimental evidences are required. Though we found 26 Rhs proteins with metallopeptidase-4 domain in EHEC O157:H7 strains, none had a PAAR domain that would identify them as T6SS effectors. Finally, 11 PAAR-Rhs proteins contained an amino acid kinase motif (AAK, cd04259) in their C-terminal, which is commonly found in proteins related to lysine and aspartate metabolism. Proteins with this activity could phosphorylate prey proteins and change their behavior, thus contributing to the bacterium survival and/or pathogenesis. However, EHEC O157:H7 strain EDL933 does not possess this interesting motif.

scrambling, ion transport and regulation of other membrane proteins (Pedemonte and Galletta, 2014). On the other hand, RhsBI also contained a dsRNA-binding domain found in the Dead-end protein homolog-1 (DND1) protein (*e-value*: 0.087). Both motifs have a high *e-value*, meaning that RhsB shares poor homology with those proteins, so the function of the RhsBI protein remains unclear until experimental evidence. However, it is worth to mention that DND1 is a RNA-binding protein, and it is known to protect eukaryotic mRNA from miRNAs, so the RhsBI protein could have a protective effect from mRNA degradation from the RNase motif DUF4329 present in RhsB, as we discussed earlier.

For RhsCI (locus *z0853*) and RhsDI (locus *z0654*) there was a TPR motif and a TIGR03373 motif, respectively, but both had high *e-values* (0.21 and 0.35), suggesting that there is no sequence similarity between these proteins and the consensus sequence. STRING analyses suggest that the *rhcC-D* genes are often in the same neighborhood as their cognate immunity genes, but there is no evidence of co-expression or interaction for any Rhs protein with their associated immunity protein.

Locus *EDL933_2188* would encode the immunity protein of RhsE, and the protein encoded was of 61 amino acids that bear a lipoprotein signal peptide according to the software SignalP (Juncker et al., 2003), indicating that the immunity protein is probably anchored to the bacterial internal membrane. As we mentioned earlier, RhsE protein possessed structural homology with RsaA protein, present in the S-layer. S-layer located proteins (SLPs) are involved in several activities, such as cell-wall biogenesis, cell division, and swimming. Besides, SLPs have been involved in pathogenesis, altering the immune response of the host and promoting bacterial adhesion to host cells (Fagan and Fairweather, 2014). This signal peptide might lead to the translocation of RhsEI to the periplasm, suggesting that RhsEI binds to RhsE in the outer membrane, where RhsE would normally act. Dong et al. (2013) have shown that Tai3, the immunity protein for the amidase effector Tae3 in *Ralstonia picketti*, also harbors a signal sequence. In the S-layer, the C-terminal end of RhsE could damage the plasma membrane or the cell-wall of other bacteria, and the immunity protein would be on the membrane to avoid cell death provoked by RhsE in sister cells. Additionally, RhsEI possessed another motif (PF16855) present in small proteins of the viral external capsid (*e-value*: 0.072), though the high *e-value* indicates poor homology with the consensus sequence of PF16855.

Finally, the candidate for the PAAR-I protein (locus *z0269*) was 203 amino acid long, larger than other putative immunity proteins. We identified 7 SLR/TPR motifs, as previously mentioned, are motifs found in immunity proteins. Upon analysis of the putative tridimensional structure of PAARI by I-TASSER, we found that the immunity protein resembled the *Helicobacter* cysteine-rich protein C (TM-score: 0.928, pdb: 1OUV) (Figure 2D). Although the exact role of this protein is still unknown, it seems to participate during cell-wall biogenesis. Besides, patients infected with *H. pylori* usually bear high antibody titers against this protein. Lüthy et al. (2004) demonstrated that the structural conformation of the *Helicobacter* cysteine-rich protein C resembles the Hsp60/Hsp70,

which provides more evidence of the possible chaperone activity of PAARI. Finally, the presence of loci *z0268* and *z0267* is specific to a specific lineage of O157:H7 that includes EHEC EDL933 and EHEC Sakai, as previously reported (Zhang et al., 2007).

Small Rhs proteins were also part of bicistronic operons, and we named them RhsH (*z0272*) and RhsG (*z0702*). Both effectors were part of an island that also contained another *rhc* gene. RhsG seems to be the former C-terminal region of RhsF that was probably divided when a viral sequence was inserted, as we mentioned above. Since these proteins were 444 and 586 amino acids in length, compared to the >1,400 amino acids of other Rhs proteins, their genes were probably part of a larger former gene. The putative immunity proteins for RhsG and RhsH were also small (69 and 56 amino acids respectively) and did not have a clear structural homolog. The absence of the PAAR motif in both proteins suggests that these proteins are not secreted via T6SS, and the small size and poor structural homology of the putative immunity proteins suggest that their genes are actually pseudogenes.

The immunity proteins protect the bacteria from sister cells attacks when the effectors modulate antibacterial competition. Interestingly, Wan et al. (2017) have shown that EHEC Δ *hms* does not have a bactericidal activity, suggesting that the Rhs proteins do not have a toxic effect on other bacteria, although *rhc* gene expression was not measured. As we show here, several effector-immunity proteins seem to be related to the cell membrane and cell-wall biogenesis, so more studies are necessary to elucidate the function of the T6SS both in bacterial competition and during pathogenesis. It is also possible that the putative immunity proteins described have a different function than predicted, maybe as chaperones like EagR.

CONCLUSIONS

The two existing genome sequences exhibit critical differences among the Rhs proteins, and as demonstrated experimentally and bioinformatically we suggest using the genome published by Latif et al. (2014) to study the Rhs proteins in EHEC strain EDL933. Upon analysis of the core components of the T6SS island, we found that the TssA gene seems to be split in two, but these two proteins should still be able to perform their function, as the functionality of the T6SS has been demonstrated by Wan et al. (2017). Although EHEC strain EDL933 lacks the commonly found T6SS effectors in *E. coli*, many *rhc* genes appear to serve as T6SS effectors. Using a structural homology approach, we postulate that RhsA may interact with small GTPases such as Rho and Rac; RhsB could interfere in the protein ubiquitination; RhsC might participate in the vesicular traffic; RhsD could auto-aggregate; RhsE might act as an S-layer adhesin; and PAAR could act as a cysteine protease. Finally, RhsF would not have any function due to a viral DNA sequence inserted on the 3' end of the gene, which separated the former metalloproteinase into RhsF and RhsG. Additionally, we did not find direct *in silico* information supporting EHEC strain EDL933 is involved in bacterial competition, as suggested by Wan et al. (2017). The role of Rhs proteins during pathogenesis

or bacterial competition requires experimental support. The exact role of T6SS and its effectors in EHEC and other STECs might provide new strategies to fight diseases caused by those pathogens.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: AE005174.2, CP008957.1.

REFERENCES

- Abby, S. S., Cury, J., Guglielmini, J., Néron, B., Touchon, M., and Rocha, E. P. (2016). Identification of protein secretion systems in bacterial genomes. *Sci. Rep.* 6:23080. doi: 10.1038/srep23080
- Alcoforado Diniz, J., Liu, Y. C., and Coulthurst, S. J. (2015). Molecular weaponry: diverse effectors delivered by the type VI secretion system. *Cell. Microbiol.* 17, 1742–1751. doi: 10.1111/cmi.12532
- Andrews, B., Carroll, J., Ding, S., Fearnley, I. M., and Walker, J. E. (2013). Assembly factors for the membrane arm of human complex I. *Proc. Natl. Acad. Sci. U.S.A.* 110, 18934–18939. doi: 10.1073/pnas.1319247110
- Arenas-Hernández, M. M., Rojas-López, M., Medrano-López, A., Nuñez-Reza, K. J., Puente, J. L., Martínez-Laguna, Y., et al. (2014). Environmental regulation of the long polar fimbriae 2 of enterohemorrhagic *Escherichia coli* O157: H7. *FEMS Microbiol. Lett.* 357, 105–114. doi: 10.1111/1574-6968.12513
- Basler, M., Ho, B., and Mekalanos, J. (2013). Tit-for-tat: type VI secretion system counterattack during bacterial cell-cell interactions. *Cell* 152, 884–894. doi: 10.1016/j.cell.2013.01.042
- Bharat, T. A., Kureisaite-Ciziene, D., Hardy, G. G., Ellen, W. Y., Devant, J. M., Hagen, W. J., et al. (2017). Structure of the hexagonal surface layer on *Caulobacter crescentus* cells. *Nature Microbiol.* 2:17059. doi: 10.1038/nmicrobiol.2017.59
- Bingle, L. E., Bailey, C. M., and Pallen, M. J. (2008). Type VI secretion: a beginner's guide. *Curr. Opin. Microbiol.* 11, 3–8. doi: 10.1016/j.mib.2008.01.006
- Boardman, B. K., Meehan, B. M., and Satchell, K. J. F. (2007). Growth phase regulation of *Vibrio cholerae* RTX toxin export. *J. Bacteriol.* 189, 1827–1835. doi: 10.1128/JB.01766-06
- Bondage, D. D., Lin, J.-S., Ma, L.-S., Kuo, C.-H., and Lai, E.-M. (2016). VgrG C-terminus confers the type VI effector transport specificity and is required for binding with PAAR and adaptor-effector complex. *Proc. Natl. Acad. Sci. U.S.A.* 113, E3931–E3940. doi: 10.1073/pnas.1600428113
- Brooks, T. M., Unterweger, D., Bachmann, V., Kostiuk, B., and Pukatzki, S. (2013). Lytic activity of the *vibrio cholerae* type VI secretion toxin VgrG-3 is inhibited by the antitoxin TsaB. *J. Biol. Chem.* 288, 7618–7625. doi: 10.1074/jbc.M112.436725
- Busby, J. N., Panjikar, S., Landsberg, M. J., Hurst, M. R., and Lott, J. S. (2013). The BC component of ABC toxins is an RHS-repeat-containing protein encapsulation device. *Nature* 501, 547. doi: 10.1038/nature12465
- Croxen, M. A., Law, R. J., Scholz, R., Keeney, K. M., Wlodarska, M., and Finlay, B. B. (2013). Recent advances in understanding enteric pathogenic *Escherichia coli*. *Clin. Microbiol. Rev.* 26, 822–880. doi: 10.1128/CMR.00022-13
- Cruz-Ortega, J. S., and Boucard, A. A. (2019). Actin cytoskeleton remodeling defines a distinct cellular function for adhesion G protein-coupled receptors ADGRL/latrophilins 1, 2 and 3. *Biol. Open* 8:bio039826. doi: 10.1242/bio.039826
- Diniz, J. A., and Coulthurst, S. J. (2015). Intraspecies competition in *Serratia marcescens* is mediated by type VI-secreted Rhs effectors and a conserved effector-associated accessory protein. *J. Bacteriol.* 197, 2350–2360. doi: 10.1128/JB.00199-15
- Dong, C., Zhang, H., Gao, Z.-Q., Wang, W.-J., She, Z., Liu, G.-F., et al. (2013). Structural insights into the inhibition of type VI effector Tae3 by its immunity protein Tai3. *Biochem. J.* 454, 59–68. doi: 10.1042/BJ20130193

AUTHOR CONTRIBUTIONS

JV-L and FN-G participated in the design of the study, data analysis and writing of the manuscript. JV-L carried out the PCR experiments. Both authors read and approved the final manuscript.

FUNDING

FN-G was supported by a Conacyt grant (501212).

- Douzi, B., Brunet, Y. R., Spinelli, S., Lensi, V., Legrand, P., Blangy, S., et al. (2016). Structure and specificity of the Type VI secretion system ClpV-TssC interaction in enteroaggregative *Escherichia coli*. *Sci. Rep.* 6, 1–13. doi: 10.1038/srep34405
- Fagan, R. P., and Fairweather, N. F. (2014). Biogenesis and functions of bacterial S-layers. *Nat. Rev. Microbiol.* 12, 211–222. doi: 10.1038/nrmicro3213
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.1111/j.1558-5646.1985.tb00420.x
- Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2013). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223
- Frank, C., Werber, D., Cramer, J. P., Askar, M., Faber, M., and Heiden, M., et al. (2011). Epidemic profile of shiga-toxin-producing *Escherichia coli* O104: H4 outbreak in Germany. *New Engl. J. Med.* 365, 1771–1780. doi: 10.1056/NEJMoa1106483
- Gueguen, E., and Cascales, E. (2013). Promoter swapping unveils the role of the *Citrobacter rodentium* CTS1 type VI secretion system in interbacterial competition. *Appl. Environ. Microbiol.* 79, 32–38. doi: 10.1128/AEM.02504-12
- Hachani, A., Allsopp, L. P., Oduko, Y., and Filloux, A. (2014). The VgrG proteins are “a la carte” delivery systems for bacterial type VI effectors. *J. Biol. Chem.* 289, 17872–17884. doi: 10.1074/jbc.M114.563429
- Jackson, V. A., Meijer, D. H., Carrasquero, M., van Bezouwen, L. S., Lowe, E. D., Kleanthous, C., et al. (2018). Structures of teneurin adhesion receptors reveal an ancient fold for cell-cell interaction. *Nat. Commun.* 9:1079. doi: 10.1038/s41467-018-03460-0
- Jiang, F., Waterfield, N. R., Yang, J., Yang, G., and Jin, Q. (2014). A *Pseudomonas aeruginosa* type VI secretion phospholipase D effector targets both prokaryotic and eukaryotic cells. *Cell Host Microbe* 15, 600–610. doi: 10.1016/j.chom.2014.04.010
- Journet, L., and Cascales, E. (2016). The type VI secretion system in *Escherichia coli* and related species. *EcoSal. Plus* 7:1–20. doi: 10.1128/ecosalplus.ESP-0009-2015
- Juncker, A. S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H., and Krogh, A. (2003). Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* 12, 1652–1662. doi: 10.1110/ps.0303703
- Koskiniemi, S., Lamoureux, J. G., Nikolakakis, K. C., t'Kint de Roodenbeke, C., Kaplan, M. D., Low, D. A., et al. (2013). Rhs proteins from diverse bacteria mediate intercellular competition. *Proc. Natl. Acad. Sci. U.S.A.* 110, 7032–7037. doi: 10.1073/pnas.1300627110
- Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evolut.* 35, 1547–1549. doi: 10.1093/molbev/msy096
- Kung, V. L., Khare, S., Stehlik, C., Bacon, E. M., Hughes, A. J., and Hauser, A. R. (2012). An rhs gene of *Pseudomonas aeruginosa* encodes a virulence protein that activates the inflammasome. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1275–1280. doi: 10.1073/pnas.1109285109
- Latif, H., Li, H. J., Charusanti, P., Palsson, B. O., and Aziz, R. K. (2014). A Gapless, unambiguous genome sequence of the enterohemorrhagic *Escherichia coli* O157:H7 strain EDL933. *Genome Announc.* 2:e00821-14. doi: 10.1128/genomeA.00821-14
- Leiman, P. G., Basler, M., Ramagopal, U. A., Bonanno, J. B., Sauder, J. M., Pukatzki, S., et al. (2009). Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin.

- Proc. Natl. Acad. Sci. U.S.A.* 106, 4154–4159. doi: 10.1073/pnas.0813360106
- Li, J., Yao, Y., Xu, H. H., Hao, L., Deng, Z., Rajakumar, K., et al. (2015). SecReT6: a web-based resource for type VI secretion systems found in bacteria. *Environ. Microbiol.* 17, 2196–2202. doi: 10.1111/1462-2920.12794
- Lien, Y.-W., and Lai, E.-M. (2017). Type VI secretion effectors: methodologies and biology. *Front. Cell. Infect. Microbiol.* 7:254. doi: 10.3389/fcimb.2017.00254
- Lu, J., Yu, Y., Zhu, L., Cheng, Y., and Sun, P. D. (2014). Structural mechanism of serum amyloid A-mediated inflammatory amyloidosis. *Proc. Natl. Acad. Sci. U.S.A.* 111, 5189–5194. doi: 10.1073/pnas.1322357111
- Lüthy, L., Grütter, M. G., and Mittl, P. R. (2004). The crystal structure of helicobacter cysteine-rich protein C at 2.0 Å resolution: similar peptide-binding sites in TPR and SEL1-like repeat proteins. *J. Mol. Biol.* 340, 829–841. doi: 10.1016/j.jmb.2004.04.055
- Ma, A. T., and Mekalanos, J. J. (2010). *In vivo* actin cross-linking induced by *Vibrio cholerae* type VI secretion system is associated with intestinal inflammation. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4365–4370. doi: 10.1073/pnas.0915156107
- Ma, J., Pan, Z., Huang, J., Sun, M., Lu, C., and Yao, H. (2017a). The hcp proteins fused with diverse extended-toxin domains represent a novel pattern of antibacterial effectors in type VI secretion systems. *Virulence* 8, 1189–1202. doi: 10.1080/21505594.2017.1279374
- Ma, J., Sun, M., Dong, W., Pan, Z., Lu, C., and Yao, H. (2017b). PAAR-Rhs proteins harbor various C-terminal toxins to diversify the antibacterial pathways of type VI secretion systems. *Environ. Microbiol.* 19, 345–360. doi: 10.1111/1462-2920.13621
- Ma, J., Sun, M., Pan, Z., Song, W., Lu, C., and Yao, H. (2018). Three hcp homologs with divergent extended loop regions exhibit different functions in avian pathogenic *Escherichia coli*. *Emerg. Microbes Infect.* 7, 1–13. doi: 10.1038/s41426-018-0042-0
- Mahillon, J., Kirkpatrick, H. A., Kijenski, H. L., Bloch, C. A., Rode, C. K., Mayhew, G. F., et al. (1998). Subdivision of the *Escherichia coli* K-12 genome for sequencing: manipulation and DNA sequence of transposable elements introducing unique restriction sites. *Gene* 223, 47–54. doi: 10.1016/S0378-1119(98)00365-5
- Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., et al. (2010). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* 39(Suppl. 1), D225–D229. doi: 10.1093/nar/gkq1189
- Mittl, P. R., and Schneider-Brachert, W. (2007). Sel1-like repeat proteins in signal transduction. *Cell. Signal.* 19, 20–31. doi: 10.1016/j.cellsig.2006.05.034
- Miyata, S. T., Bachmann, V., and Pukatzki, S. (2013). Type VI secretion system regulation as a consequence of evolutionary pressure. *J. Med. Microbiol.* 62, 663–676. doi: 10.1099/jmm.0.053983-0
- Navarro-Garcia, F. (2014). *Escherichia coli* O104:H4 pathogenesis: an enteroaggregative *E. coli/shiga toxin-producing E. coli explosive cocktail of high virulence*. *Microbiol. Spectr.* 2:1–19. doi: 10.1128/microbiolspec.EHEC-0008-2013
- Navarro-Garcia, F., Ruiz-Perez, F., Cataldi, A. A., and Larzabal, M. (2019). Type VI secretion system in pathogenic *Escherichia coli*: structure, role in virulence and acquisition. *Front. Microbiol.* 10:1965. doi: 10.3389/fmicb.2019.01965
- Niemi, K., Teirilä, L., Lappalainen, J., Rajamäki, K., Baumann, M. H., Öörni, K., et al. (2011). Serum amyloid A activates the NLRP3 inflammasome via P2X7 receptor and a cathepsin B-sensitive pathway. *J. Immunol.* 186, 6119–6128. doi: 10.4049/jimmunol.1002843
- Pedemonte, N., and Galletta, L. J. (2014). Structure and function of TMEM16 proteins (anocytins). *Physiol. Rev.* 94, 419–459. doi: 10.1152/physrev.00039.2011
- Perna, N. T., Plunkett, G. 3rd, Burland, V., Mau, B., Glasner, J. D., Rose, D. J., et al. (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409, 529–533. doi: 10.1038/35054089
- Poole, S. J., Diner, E. J., Aoki, S. K., Braaten, B. A., de Roodenbeke, C., t'K., Low, D. A., et al. (2011). Identification of functional toxin/immunity genes linked to contact-dependent growth inhibition (CDI) and rearrangement hotspot (Rhs) systems. *PLoS Genet.* 7:e1002217. doi: 10.1371/journal.pgen.1002217
- Prochazkova, K., Shuvalova, L. A., Minasov, G., Voburka, Z., Anderson, W. F., and Satchell, K. J. (2009). Structural and molecular mechanism for autoprocessing of MARTX toxin of *Vibrio cholerae* at multiple sites. *J. Biol. Chem.* 284, 26557–26568. doi: 10.1074/jbc.M109.025510
- Pukatzki, S., Ma, A. T., Revel, A. T., Sturtevant, D., and Mekalanos, J. J. (2007). Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15508–15513. doi: 10.1073/pnas.0706532104
- Riley, L. W., Remis, R. S., Helgerson, S. D., McGee, H. B., Wells, J. G., Davis, B. R., et al. (1983). Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *New Engl. J. Med.* 308, 681–685. doi: 10.1056/NEJM198303243081203
- Rohde, H., Qin, J., Cui, Y., Li, D., Loman, N. J., Hentschke, M., et al. (2011). Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104: H4. *New Engl. J. Med.* 365, 718–724. doi: 10.1056/NEJMoa1107643
- Sana, T. G., Baumann, C., Merdes, A., Soscia, C., Rattei, T., Hachani, A., et al. (2015). Internalization of *Pseudomonas aeruginosa* strain PAO1 into epithelial cells is promoted by interaction of a T6SS effector with the microtubule network. *MBio* 6, e00712–00715. doi: 10.1128/mBio.00712-15
- Schmidt, M. A. (2010). LEEways: tales of EPEC, ATEC and EHEC. *Cell. Microbiol.* 12, 1544–1552. doi: 10.1111/j.1462-5822.2010.01518.x
- Schneider, J. P., Nazarov, S., Adaixo, R., Liuzzo, M., Ringel, P. D., Stahlberg, H., et al. (2019). Diverse roles of TssA-like proteins in the assembly of bacterial type VI secretion systems. *EMBO J.* 38:e100825. doi: 10.15252/embj.2018100825
- Schweitzer, A., Auferheide, A., Rudack, T., Beck, F., Pfeifer, G., Plitzko, J. M., et al. (2016). Structure of the human 26S proteasome at a resolution of 3.9 Å. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7816–7821. doi: 10.1073/pnas.1608050113
- Shneider, M. M., Buth, S. A., Ho, B. T., Basler, M., Mekalanos, J. J., and Leiman, P. G. (2013). PAAR-repeat proteins sharpen and diversify the type VI secretion system spike. *Nature* 500, 350–353. doi: 10.1038/nature12453
- Shohdy, N., Efe, J. A., Emr, S. D., and Shuman, H. A. (2005). Pathogen effector protein screening in yeast identifies *Legionella* factors that interfere with membrane trafficking. *Proc. Natl. Acad. Sci. U.S.A.* 102, 4866–4871. doi: 10.1073/pnas.0501315102
- Sigrist, C. J., De Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., et al. (2012). New and continuing developments at PROSITE. *Nucleic Acids Res.* 41, D344–D347. doi: 10.1093/nar/gks1067
- Suarez, G., Sierra, J., Erova, T., Sha, J., Horneman, A., and Chopra, A. (2010). A type VI secretion system effector protein, VgrG1, from *Aeromonas hydrophila* that induces host cell toxicity by ADP ribosylation of actin. *J. Bacteriol.* 192, 155–168. doi: 10.1128/JB.01260-09
- Suits, M. D., and Boraston, A. B. (2013). Structure of the streptococcus pneumoniae surface protein and adhesin PfbA. *PLoS ONE* 8:e67190. doi: 10.1371/journal.pone.0067190
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. doi: 10.1093/nar/gky1131
- Tamura, K., and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10, 512–526.
- Tilden Jr, J., Young, W., McNamara, A.-M., Custer, C., Boesel, B., Lambert-Fair, M. A., et al. (1996). A new route of transmission for *Escherichia coli*: infection from dry fermented salami. *Am. J. Public Health* 86, 1142–1145. doi: 10.2105/AJPH.86.8_Pt_1.1142
- Uchida, K., Leiman, P. G., Arisaka, F., and Kanamaru, S. (2013). Structure and properties of the C-terminal β -helical domain of VgrG protein from *Escherichia coli* O157. *J. Biochem.* 155, 173–182. doi: 10.1093/jb/mvt109
- Urosev, D., Ferrer-Navarro, M., Pastorello, I., Cartocci, E., Costenaro, L., Zhulenkova, D., et al. (2013). Crystal structure of c5321: a protective antigen present in uropathogenic *Escherichia coli* strains displaying an SLR fold. *BMC Struct. Biol.* 13:19. doi: 10.1186/1472-6807-13-19
- Voth, K. A., Chung, I. Y. W., van Straaten, K., Li, L., Boniecki, M. T., and Cygler, M. (2019). The structure of *Legionella* effector protein LpnE provides insights into its interaction with Oculocerebrorenal syndrome of Lowe (OCRL) protein. *FEBS J.* 286, 710–725. doi: 10.1111/febs.14710
- Wan, B., Zhang, Q., Ni, J., Li, S., Wen, D., Li, J., et al. (2017). Type VI secretion system contributes to Enterohemorrhagic *Escherichia coli* virulence by secreting catalase against host reactive oxygen species (ROS). *PLoS Pathogens* 13:e1006246. doi: 10.1371/journal.ppat.1006246

- Whitney, J. C., Beck, C. M., Goo, Y. A., Russell, A. B., Harding, B. N., De Leon, J. A., et al. (2014). Genetically distinct pathways guide effector export through the type VI secretion system. *Mol. Microbiol.* 92, 529–542. doi: 10.1111/mmi.12571
- Wong, C. S., Jelacic, S., Habeeb, R. L., Watkins, S. L., and Tarr, P. I. (2000). The risk of the hemolytic-uremic syndrome after antibiotic treatment of *Escherichia coli* O157: H7 infections. *New Engl. J. Med.* 342, 1930–1936. doi: 10.1056/NEJM200006293422601
- Yang, J., and Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* 43, W174–W181. doi: 10.1093/nar/gkv342
- Yang, S.-C., Lin, C.-H., Aljuffali, I. A., and Fang, J.-Y. (2017). Current pathogenic *Escherichia coli* foodborne outbreak cases and therapy development. *Arch. Microbiol.* 199, 811–825. doi: 10.1007/s00203-017-1393-y
- Yang, X., Long, M., and Shen, X. (2018). Effector-immunity pairs provide the T6SS nanomachine its offensive and defensive capabilities. *Molecules* 23:1009. doi: 10.3390/molecules23051009
- Yang, X. Y., Li, Z. Q., She, Z., Geng, Z., Xu, J. H., Gao, Z. Q., et al. (2016). Structural analysis of *Pseudomonas aeruginosa* H3-T6SS immunity proteins. *FEBS Lett.* 590, 2787–2796. doi: 10.1002/1873-3468.12291
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinform.* 13:134. doi: 10.1186/1471-2105-13-134
- Zhang, L., Xu, J., Xu, J., Zhang, H., He, L., and Feng, J. (2014). TssB is essential for virulence and required for type VI secretion system in *Ralstonia solanacearum*. *Microb. Pathogen.* 74, 1–7. doi: 10.1016/j.micpath.2014.06.006
- Zhang, Y., Laing, C., Steele, M., Ziebell, K., Johnson, R., Benson, A. K., et al. (2007). Genome evolution in major *Escherichia coli* O157: H7 lineages. *BMC genomics* 8:121. doi: 10.1186/1471-2164-8-121
- Zoued, A., Durand, E., Brunet, Y. R., Spinelli, S., Douzi, B., Guzzo, M., et al. (2016). Priming and polymerization of a bacterial contractile tail structure. *Nature* 531, 59. doi: 10.1038/nature17182

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Vazquez-Lopez and Navarro-Garcia. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.