

# Functional and Computational Analysis of Amino Acid Patterns Predictive of Type III Secretion System Substrates in *Pseudomonas syringae*

Lisa M. Schechter<sup>1\*</sup>, Joy C. Valenta<sup>1</sup>, David J. Schneider<sup>2,3</sup>, Alan Collmer<sup>3</sup>, Eric Sakk<sup>4</sup>

**1** Department of Biology, University of Missouri-St. Louis, St. Louis, Missouri, United States of America, **2** Plant-Microbe Interactions Research Unit, Robert W. Holley Center for Agriculture and Health, Agricultural Research Service, United States Department of Agriculture, Ithaca, New York, United States of America, **3** Department of Plant Pathology and Plant-Microbe Biology, Cornell University, Ithaca, New York, United States of America, **4** Department of Computer Science, Morgan State University, Baltimore, Maryland, United States of America

## Abstract

Bacterial type III secretion systems (T3SSs) deliver proteins called effectors into eukaryotic cells. Although N-terminal amino acid sequences are required for translocation, the mechanism of substrate recognition by the T3SS is unknown. Almost all actively deployed T3SS substrates in the plant pathogen *Pseudomonas syringae* pathovar tomato strain DC3000 possess characteristic patterns, including (i) greater than 10% serine within the first 50 amino acids, (ii) an aliphatic residue or proline at position 3 or 4, and (iii) a lack of acidic amino acids within the first 12 residues. Here, the functional significance of the *P. syringae* T3SS substrate compositional patterns was tested. A mutant AvrPto effector protein lacking all three patterns was secreted into culture and translocated into plant cells, suggesting that the compositional characteristics are not absolutely required for T3SS targeting and that other recognition mechanisms exist. To further analyze the unique properties of T3SS targeting signals, we developed a computational algorithm called TERE (Type III Effector Relative Entropy Evaluation) that distinguishes DC3000 T3SS substrates from other proteins with a high sensitivity and specificity. Although TERE did not efficiently identify T3SS substrates in *Salmonella enterica*, it was effective in another *P. syringae* strain and *Ralstonia solanacearum*. Thus, the TERE algorithm may be a useful tool for identifying new effector genes in plant pathogens. The nature of T3SS targeting signals was additionally investigated by analyzing the N-terminus of FtsX, a putative membrane protein that was classified as a T3SS substrate by TERE. Although the first 50 amino acids of FtsX were unable to target a reporter protein to the T3SS, an AvrPto protein substituted with the first 12 amino acids of FtsX was translocated into plant cells. These results show that the T3SS targeting signals are highly mutable and that secretion may be directed by multiple features of substrates.

**Citation:** Schechter LM, Valenta JC, Schneider DJ, Collmer A, Sakk E (2012) Functional and Computational Analysis of Amino Acid Patterns Predictive of Type III Secretion System Substrates in *Pseudomonas syringae*. PLoS ONE 7(4): e36038. doi:10.1371/journal.pone.0036038

**Editor:** Ching-Hong Yang, University of Wisconsin-Milwaukee, United States of America

**Received:** February 9, 2012; **Accepted:** March 29, 2012; **Published:** April 27, 2012

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

**Funding:** This work was supported by start-up research funds from the University of Missouri-St. Louis, National Science Foundation award DBI-0077622, and United States Department of Agriculture-Agriculture Research Service project 1907-21000-017-00. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: schechterl@umsl.edu

## Introduction

Gram-negative bacteria have developed a wide variety of mechanisms to export proteins. One of the best studied protein secretion devices is the type III secretion system (T3SS), which transports extracellular components of the flagellum [1]. Some Gram-negative pathogens and symbionts also contain T3SSs that deliver proteins called effectors directly from the bacterial cytoplasm into host cells during infection [2]. Recent findings suggest that T3SSs may additionally translocate extracellular bacterial proteins into host cells [3,4]. Once inside the host cell cytoplasm, effectors mimic host proteins and manipulate signaling pathways to promote bacterial survival and growth during infection [5,6].

Identifying the complete collection of T3SS effectors produced by a particular bacterium has proven difficult for several reasons. First, many effectors have similar or redundant functions inside host cells, which may mask phenotypes in screens for less virulent

mutants. Studies in *Salmonella enterica* and *Pseudomonas syringae* have shown that deletion of multiple effector genes is often required to observe attenuation in virulence assays [7–9]. Second, genetic screens to identify new effectors are often labor intensive [10–13]. Proteomic analysis of culture supernatants may be a more efficient way to identify T3SS substrates [14,15]. However, this method may fail to discover effectors that are secreted in small amounts or are only deployed upon host cell contact. Finally, many effectors appear to be unique to certain species or even strains of bacteria. Thus, homology searches have only been successful at identifying a subset of the effectors present in any one bacterium.

Understanding how effector proteins are targeted for secretion is crucial for discovering all of the effector genes in bacteria, as well as for developing new methods to inhibit T3SS function. Although the mechanism of substrate recognition by the T3SS is unclear, two models have been proposed to explain how effectors are distinguished from other bacterial proteins. In the first model, effectors are targeted to the T3SS by N-terminal amino acid

sequences. This model is based on studies showing that the first ~15 amino acids of the *Yersinia* effector YopE are essential for secretion into the extracellular milieu [16,17]. A larger region (~50 N-terminal amino acids) is required for effector translocation into host cells [16,17]. The additional sequences required for efficient translocation may be involved in mediating the delivery of effectors from an extracellular location into host cells [3].

In the second model of T3SS substrate recognition, sequences within the first 15 codons of mRNAs form secondary structures that target effector proteins for cotranslational export through the T3SS [18]. In support of this hypothesis, frameshift mutations that drastically change the N-terminal amino acid sequences of effector proteins but minimally alter the mRNA sequence do not abrogate effector secretion or translocation by the T3SS [18–21]. However, effector secretion is also unaffected by synonymous changes within the first 15 codons that considerably affect the mRNA secondary structure without altering the protein sequence [21,22]. The observation that effectors are deployed in the presence of translation inhibitors additionally casts doubt on the cotranslational secretion theory [22]. Altogether, these findings indicate that the T3SS targeting signal within the N-terminal 15 amino acids of effectors is highly degenerate and tolerant of mutations. Thus, it may be impossible to identify a consensus T3SS recognition sequence within effector proteins.

In addition to endogenous targeting signals, effectors may be guided to the T3SS by accessory factors called chaperones. T3SS chaperones are small, usually acidic proteins that have similar structures, even though their amino acid sequences are not significantly similar. Chaperone genes are generally encoded adjacent to effector genes, or within T3SS gene clusters. They bind to the effector chaperone-binding domain (CBD), a ~50–100 amino acid region that is directly downstream from the N-terminal secretion targeting signal [2]. Although many chaperones are dedicated to binding only one effector, some chaperones are promiscuous and bind to several different effectors [23]. Two lines of evidence support a role for chaperones in effector targeting. First, deletion or mutation of the CBDs in the *Salmonella* effectors SopA, SopE, SptP, and SipA causes these proteins to be secreted into culture via the flagellar export pathway, rather than the *Salmonella* pathogenicity island 1 (SPI-1)-encoded T3SS [24–26]. This finding indicates that at least some effectors require chaperones for targeting to the proper T3SS. Second, chaperones can interact with proteins at the base of the T3SS in *Salmonella*, enteropathogenic *E. coli*, and *Chlamydia* [27–29]. However, in certain situations YopE from *Yersinia* does not require its dedicated chaperone, SycE (YerA), for T3SS-mediated secretion or translocation [17,30,31]. Thus, the N-terminal 15 amino acids of effectors are sufficient for T3SS targeting, and chaperones may serve to enhance the process.

Although the molecular mechanisms that underlie effector targeting to the T3SS remain obscure, several structural, bioinformatic, and computational analyses indicate that the N-termini of effector proteins possess common features, including: (i) flexibility and disorder in solution [32,33], (ii) amphipathicity [22,34,35], and (iii) bias for particular amino acids [35–38]. In fact, the N-terminal amino acid sequences of actively deployed effectors in *P. syringae* pathovar tomato (*P. s. tomato*) strain DC3000 have been examined extensively and generally contain three patterns. First, DC3000 effector N-termini are enriched in polar amino acids, especially serine [11,37,39]. Second, DC3000 effectors usually contain an aliphatic amino acid or proline at the third or fourth position [39–41]. Finally, DC3000 T3SS substrates also generally lack negatively charged amino acids within the first 12 residues [39–41]. These characteristics have been successfully

used for their predictive value as part of a bioinformatic workflow for identifying candidate effectors in *P. syringae* genomes [41,42].

The targeting patterns in *P. syringae* effectors are also found in flagellar secretion substrates and in a subset of T3SS effectors in other plant and animal pathogens. For example, the *Yersinia* effector YopE possesses the three major patterns, including an unusually high serine content of 28% in the first 50 residues. However, many T3SS substrates from animal pathogens lack one or more of these characteristic patterns [43]. This observation suggests that the characteristic targeting patterns of *P. syringae* effectors may not mediate secretion, or that two or more classes of effectors exist in bacteria with quite different N-terminal amino acid patterns.

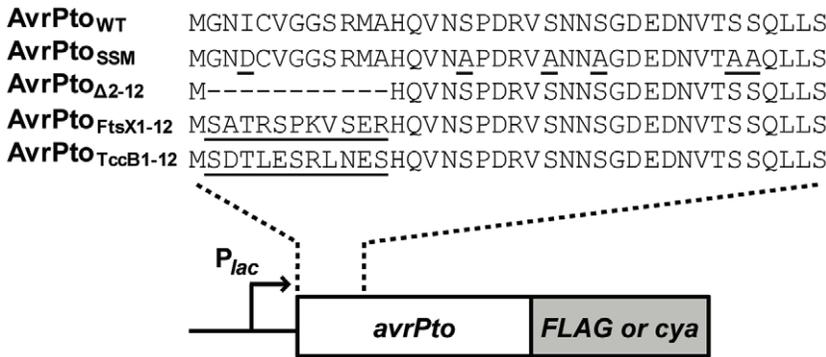
In this study, we sought to better understand how T3SS substrates are distinguished from other proteins in *P. s. tomato* DC3000, a model pathogen of the important crop tomato and the model plant *Arabidopsis*. This organism is an ideal subject for bioinformatic and computational studies on T3SS targeting signals because its genome sequence has been determined and it encodes over 50 experimentally validated T3SS substrates [41,44–46]. We first analyzed whether the characteristic targeting patterns found in most DC3000 effectors are required for the T3SS-dependent secretion of AvrPto. An altered AvrPto protein lacking all of the patterns was targeted for secretion as well as wild-type AvrPto. To determine whether DC3000 effectors have other distinctive properties, we developed a computational algorithm that measures differences between the amino acid sequences of T3SS substrates and nonsecreted proteins. In contrast to other computational T3SS substrate prediction models that utilize Naïve Bayesian, artificial neural network (ANN), or support vector machine (SVM) classification algorithms, our method is based on an information theory approach. The performance of our algorithm was analyzed in *P. syringae* and other bacteria with T3SSs, and in comparison to other T3SS prediction models. We show that our computational algorithm is a useful tool for recognizing T3SS substrates in three plant pathogens.

## Results

### Examination of T3SS targeting patterns in the *P. syringae* AvrPto effector protein

Despite the value of the characteristic T3SS targeting patterns in predicting high-probability *P. syringae* effector candidates, the importance of these sequences in mediating secretion has not been examined. We therefore analyzed the significance of the targeting patterns in AvrPto, a well-studied *P. syringae* effector that suppresses plant immune responses triggered by pathogen-associated molecular patterns (PAMPs) [47]. A previous study showed that the first ~50 amino acids of AvrPto are required for efficient secretion into culture and translocation into plant cells by the *P. s. tomato* DC3000 T3SS [40]. Plasmids were constructed that express C-terminally FLAG epitope-tagged wild-type or secretion signal mutant versions of AvrPto (AvrPto<sub>WT</sub> and AvrPto<sub>SSM</sub>, respectively). AvrPto<sub>SSM</sub> contains several mutations. The fourth residue (isoleucine) is substituted with aspartate, and most of the serines within the first 50 amino acids are changed to alanine (Figure 1). The mutant thus lacks all three of the *P. syringae* characteristic T3SS targeting patterns.

The two plasmids expressing AvrPto<sub>WT</sub> or AvrPto<sub>SSM</sub> were transferred into wild-type DC3000 and a  $\Delta hrp$  mutant derivative, which lacks the entire T3SS coding region [48]. These strains were grown in *hrp*-derepressing minimal medium (HDM) to induce T3SS gene expression, and cellular and supernatant protein samples were collected. Approximately equal levels of AvrPto<sub>SSM</sub>



**Figure 1. Schematic diagram of AvrPto mutants examined in this study.** Plasmids were constructed that express wild-type or mutant versions of the *avrPto* gene fused in frame to either *FLAG* epitope tag sequences or *cya* (calmodulin-dependent adenylate cyclase). Each gene was expressed from an upstream *lac* promoter (*P<sub>lac</sub>*). The sequences of the first 50 amino acids of each protein are shown above the *avrPto* gene. Amino acids in the mutant proteins that differ from the wild-type AvrPto sequence are underlined. Dashes within the AvrPto<sub>Δ2-12</sub> sequence indicate deleted residues.

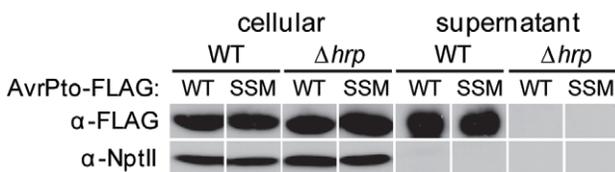
doi:10.1371/journal.pone.0036038.g001

and AvrPto<sub>WT</sub> were isolated from the culture supernatants of wild-type DC3000 (Figure 2). In addition, secretion of both AvrPto<sub>SSM</sub> and AvrPto<sub>WT</sub> was dependent on an intact T3SS. As a control, we examined the location of neomycin phosphotransferase II (NptII), a cytoplasmic protein. NptII was detected in bacterial cells but not culture supernatants, showing that cytoplasmic proteins did not leak into the growth medium during the experiment (Figure 2). Overall, these results show that the characteristic targeting patterns of *P. syringae* T3SS substrates are not required for the secretion of AvrPto.

Although the secretion signal mutations did not affect AvrPto export into the extracellular milieu, we suspected that they might reduce AvrPto translocation into plant cells. In a previous study, we showed that *P. s. tomato* DC3000 efficiently translocates an AvrPto-Cya hybrid protein into the leaves of tomato or *Nicotiana benthamiana* plants in a T3SS-dependent manner [40]. Cya is a bacterial adenylate cyclase that produces cAMP only when it is delivered into the cytoplasm of eukaryotic cells, where it can bind to its cofactor calmodulin [16]. To test whether the characteristic effector targeting patterns are required for translocation of AvrPto, we constructed four plasmids that express different versions of *avrPto-cya* (Figure 1). Two of these plasmids express Cya hybrid

proteins that include the entire 164 amino acids of AvrPto<sub>WT</sub> or AvrPto<sub>SSM</sub>. The other two plasmids express the first 50 amino acids of AvrPto<sub>WT</sub> or AvrPto<sub>SSM</sub> fused to Cya. These smaller hybrid proteins were constructed because we hypothesized that the SSM mutations might have a stronger effect in the context of the minimal AvrPto translocation signal. Expression of the appropriate sized proteins in DC3000 was confirmed by immunoblot analysis (Figure 3). Smaller protein bands were detected by the anti-Cya antibodies in some lanes of the immunoblot, as has been observed in previous studies [40,41]. These species may result from processing of the Cya hybrid protein.

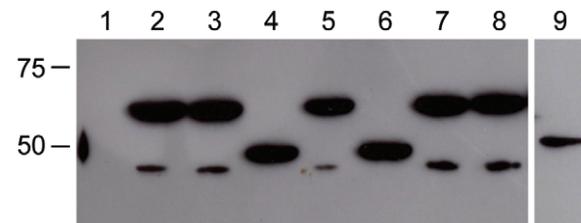
To analyze AvrPto translocation, accumulation of cAMP was measured in *N. benthamiana* leaves after inoculation with wild-type or *Δhrp* DC3000 strains expressing the various AvrPto-Cya hybrid proteins. Similar levels of cAMP were detected in *N. benthamiana* leaves inoculated with wild-type DC3000 expressing AvrPto<sub>WT(1-164)</sub>-Cya or AvrPto<sub>SSM(1-164)</sub>-Cya (Table 1). *N. benthamiana* leaves inoculated with DC3000 strains expressing AvrPto<sub>WT(1-50)</sub>-Cya or AvrPto<sub>SSM(1-50)</sub>-Cya also produced nearly the same levels of cAMP. Translocation was dependent on the T3SS, since little cAMP accumulation occurred when plant leaves were inoculated with DC3000 *Δhrp* mutants expressing the hybrid proteins. Thus, despite



**Figure 2. Secretion of AvrPto<sub>WT</sub> and AvrPto<sub>SSM</sub> by DC3000.**

Wild-type and T3SS mutant (*Δhrp*) DC3000 strains containing plasmids that express AvrPto<sub>WT</sub> or AvrPto<sub>SSM</sub> were grown in *hrp*-derepressing fructose minimal medium (HDM). Cultures were separated into cellular and supernatant fractions by centrifugation and filtration, and an immunoblot analysis was performed after electrophoresis of protein samples through a 12.5% SDS-PAGE gel. The supernatant samples are 15-fold more concentrated than the cellular samples. The 21 kDa AvrPto<sub>WT</sub> and AvrPto<sub>SSM</sub> proteins were detected using primary antibodies against the FLAG epitope. The NptII protein (29.1 kDa) expressed from pUFR034 was also detected as a cytoplasmic control using primary antibodies against NptII. The results shown were taken from samples collected during a single experiment. Similar results were observed in an independently conducted experiment.

doi:10.1371/journal.pone.0036038.g002



**Figure 3. Expression of AvrPto-Cya hybrid proteins in *P. s. tomato* DC3000.**

DC3000 strains containing plasmids that express Cya fusion proteins were grown in culture and protein samples were separated in a 12.5% SDS-PAGE gel. An immunoblot analysis was performed using primary antibodies against Cya. The protein in each lane and its estimated molecular weight is: Lane 1, empty vector; lane 2, AvrPto<sub>Δ2-12</sub>-Cya (60.9 kDa); lane 3, AvrPto<sub>WT(1-164)</sub>-Cya (62.0 kDa); lane 4, AvrPto<sub>WT(1-50)</sub>-Cya (48.9 kDa); lane 5, AvrPto<sub>SSM(1-164)</sub>-Cya (61.9 kDa); lane 6, AvrPto<sub>SSM(1-50)</sub>-Cya (48.9 kDa); lane 7, AvrPto<sub>FtsX(1-12)</sub>-Cya (62.1 kDa); lane 8, AvrPto<sub>TccB(1-12)</sub>-Cya (62.2 kDa); lane 9, FtsX<sub>1-50</sub>-Cya (50.8 kDa). The positions of protein standards on the gel are indicated to the left of the blot.

doi:10.1371/journal.pone.0036038.g003

**Table 1.** Translocation of AvrPto-Cya hybrid proteins into *N. benthamiana* by *P. s. tomato* DC3000.

Cya fusion protein	Translocation by DC3000 (pmol cAMP/μg protein)	
	Wild-type	Δhrp
AvrPto <sub>WT(1-164)</sub>	139.3±5.5	0.2±0.2
AvrPto <sub>WT(1-50)</sub>	112.6±14.2	0.6±0.2
AvrPto <sub>SSM(1-164)</sub>	152.3±15.1	0.8±0.6
AvrPto <sub>SSM(1-50)</sub>	124.4±19.3	0.8±0.2

doi:10.1371/journal.pone.0036038.t001

lacking the common T3SS secretion signal targeting patterns, the AvrPto<sub>SSM</sub> mutant was translocated into cells as well as AvrPto<sub>WT</sub>.

### Amino acid composition comparisons between T3SS substrates and other DC3000 proteins

The characteristic targeting patterns in *P. syringae* effectors were initially identified by manual examination of amino acid sequences. We reasoned that a computational approach would more comprehensively determine properties that are unique to T3SS substrates. To begin our analysis, a substrate training set was constructed that contained most of the experimentally confirmed DC3000 T3SS substrates, which came to 38 proteins in total (Table 2). HopP1, HopAO1, HopT1-2, HopAA1-2, and HopAM1-2 were excluded from the substrate training set because they are highly homologous to other DC3000 effector proteins and thus might bias results. Several other validated effectors were also omitted because the genes that encode them in DC3000 are not expressed or are interrupted by transposons [41]. The rest of the proteins encoded in the DC3000 genome (~5600) were used as a background data set for comparison. It is important to note that the background data set could contain T3SS substrates that have not yet been identified.

To compare the composition of T3SS substrates to other nonsecreted proteins, we used an information theory-based classifier that involves computations of relative entropy [49,50]. This classifier analyzed the T3SS substrate and background training sets using a sliding window size of 3. For each window block, a probability score (or entropy estimate) was calculated as described in the Materials and Methods. For the T3SS substrate sequences, the entropy estimate for each sliding window was fairly constant at about 4.1 bits (Figure 4). In contrast, the background data set differed in information content from the T3SS substrate set by between 0.1 to 0.5 bits. This result, along with findings from others, confirms that differences in amino acid composition can be exploited to develop computational models that recognize T3SS substrates [38,51].

### Classification of DC3000 T3SS substrates based on relative entropy measurements

To distinguish DC3000 T3SS substrates from other proteins, we developed an algorithm incorporating a symmetric version of the Kullback-Liebler distance. The classifier, which we named the TERE (Type III Effector Relative Entropy Evaluation) algorithm, was trained on the DC3000 T3SS substrate and background data sets. The algorithm was then used to evaluate all annotated protein coding sequences in the *P. s. tomato* DC3000 genome. Each protein received a relative entropy score between

−47 and +34. All T3SS substrates that were used to construct the T3SS substrate training set scored between −47 and −11 (Table 2 and Table S1). Classifier performance was tested by constructing a negative training set of proteins known not to be secreted by the T3SS. Table S2 shows the score distribution for the supervised performance test. Based upon this table, we chose −13 as the cutoff score for predicting T3SS substrates. For blind classification tests involving the complete genome, Table 2 and Table S1 indicate that all but one protein in the substrate training set (HopA11) had a score below (more negative than) −13.

In addition to the proteins in the substrate training set, the TERE algorithm classified several other DC3000 proteins as potential T3SS substrates. These proteins, which scored between −47 and −13, fell into three classes: (i) experimentally validated T3SS substrates that were omitted from the substrate training set, (ii) predicted substrates of the flagellar T3SS, or (iii) unlikely T3SS substrates. Proteins in the first class included HopD, HopO1-3, HopP1, HopS1, HopT1-2, HopAA1-2, HopAG1, HopAH2-1, HopAM1-2, HopAO1, HopAQ1, HopAS1, and PSPTO\_0907 (Table 2). The fact that these omitted effectors earned scores similar to proteins in the T3SS substrate training set showed that the TERE algorithm effectively identified DC3000 effector proteins. In fact, only one known DC3000 T3SS substrate omitted from the substrate training set, HopAH2-2, did not score within the −47 to −13 range. Proteins in the second class included FlgC (flagellin), FlgM, FlgK, FlgE (two homologs), and FlgK (Table S1). These results were not surprising, because flagellin can be secreted by nonflagellar T3SSs in other bacteria, and effectors can also be secreted through the flagellum [24,52–54]. Finally, TERE identified 63 proteins in the third class. We classified these proteins as unlikely T3SS substrates because they have predicted functions in bacterial cell physiology, metabolism, or transcription regulation. Furthermore, none of the genes encoding these proteins are regulated by HrpL, an extracytoplasmic function (ECF) family sigma factor that induces expression of almost all T3SS substrates in DC3000 [13,41,55,56].

To further evaluate the effectiveness of TERE, we performed several statistical tests on the results. First, we measured the sensitivity, which determines how accurately the algorithm identifies known T3SS substrates. At the cutoff score of −13, the sensitivity of the TERE algorithm was 96.2%. This value is comparable to or better than the sensitivities achieved by other T3SS substrate predictive models [38,43,51,57,58]. Second, we determined the specificity, which assesses the proportion of proteins that are correctly identified as non-substrates of the T3SS. The specificity was 98.9%, which indicates that only about 1% of the proteins encoded by the DC3000 genome were incorrectly identified by TERE as T3SS substrates. This value is significantly higher than the specificity values reported by most other computational models [38,43,51,57,58]. We also constructed a receiver operator characteristic (ROC) curve by plotting the sensitivity versus the specificity at each score output of the TERE algorithm and calculated the area under the curve (AUC). The AUC measures the overall effectiveness of the algorithm at predicting T3SS substrates; a value of 1.0 indicates that all proteins were categorized correctly, whereas a value of 0.5 indicates that all proteins were randomly classified. The AUC for the TERE algorithm was .992, indicating that it is highly accurate. Finally, TERE performance was evaluated by a 5-fold cross validation test, in which 7–8 different effectors were randomly omitted from the positive training set in 5 distinct repetitions. The average sensitivity for the 5-fold cross validation was 90%, whereas the average specificity was 99.1% (data not shown). Therefore, the TERE algorithm retained its predictive

**Table 2.** DC3000 T3SS substrates and their scores after analysis by the TERE algorithm.

Substrate <sup>a</sup>	Score	Amino acid sequence used to construct the T3SS substrate training set
HrpA1	-19	MVAFAGLTSKLTNLGNSAVGGVGGALQGVNTVASNATLQKNI LLGTGDSL
HrpH1	-14	MFAVAFVSSPRLRLARAVQ IAVLAMGALCVGCQSVDYSPPRQDRPPRLVS
HrpJ1	-27	MKIVAPPIMRILPVAPTRVVT PAAQPLPNADLHNSGTSPPQVSRFAALI
HrpK1	-15	MRISSSPFVI VNPQTPGELALAVESPLAKALPTPVGGGQAGVQFGQPAG
HrpW1	-29	MSIGITPRPQQTTT PLDFSALS GKSPQNTFGEQNTQQAI DPSSALLFGSD
HrpZ1	-27	MQALNSISSLQTSASLFPVSLNSDVANTSTSSKELKAVIDQLVQALTQS
HopAK1	-31	MNTINRNIYPVSGISAQDAPVQTDQLQPQGGIRPGHNSNLI DFGLIQQA
AvrE1	-25	MQSPSIHRNTGSI IQFTVTPDARAATDLQERAEQPRQRSSHSLSSVGKRA
AvrPto1	-35	MGNICVGGSRMAHQVNSPDRVSNNSGDEDNVTSSQLLSVRHQLAESAGLP
HopA1	-15	MNPIQSRFSSVQELRRSNVDI PALKANGQLEVDGKRYEIRAADDGTISVL
HopB1	-33	MRPVGGPAPGYPPTYEAERPTAQ AAGNDRARSQASSSPAASVAPETPM
HopC1	-40	MTIVSGHIGKHPSLTTVQAGSSASVENQMPDPAQFSDGRWKKLPTQLSSI
HopD*	-31	
HopD1	-35	MNPLRSIQHNIATPPI SGGQPLDAVGPQAQQSHPKRISPSQLSQAHQAL
HopE1	-27	MNRVSGSSSATWQAVNDLVEQV SERTTLSTTG YQTAMGRLNKPEKSDADA
HopF2	-39	MGNICGTSGRHVYSPSHTQRITS APSTSTHVGGDTLTSIHQLSHSQREQ
HopG1	-21	MQIKNSHLYSASRMVQNTFNASPKMEV TNAIAKNNEPAALSATQTAKTHE
HopH1	-19	MITPSRYPGIYIAPLSNEPTAAHTFKEQAE EALDHISAAPSGDKLLRKIS
HopI1	-24	MINLTHIASSLARAALSDSTKPKMERA INVASHIAGKVALQVTSLLLEQK
HopK1	-35	MNRISTSSVNSSFNYTAPTEEAQNRFASAPDNS PLVTTTTSIAQASEGLQ
HopM1	-37	MISSRIGGAGGVKLSRVNQHD TVPAQTAHPNAV TAGMNPPLTPDQSGSH
HopN1	-23	MYIQQSGAQSGVAAKTQHDKPS SLSGLAPGSSDAFARFHPKAGAFVPLE
HopO1-1	-47	MGNICGTSGSNHVYSPPI SPQHASGSSTPVPSASG TMLSLSHEQILSQNY
HopO1-2	-36	MNISPVSGAHGSSYPSAQSTASTASKG PSGSFLKQLGGCFSPCLGSSSTG
HopO1-3*	-29	
HopP1*	-22	
HopQ1-1	-34	MHRPITAGHTT SRLILDQSKQISRTPSESSAQ SALSQQASMSSPVLERSK
HopR1	-23	MVKVTSSGFTANPLSHHADSVPANSP PQLPEPVHLVDLSESSRKGGMRN
HopS1*	-19	
HopS2	-17	MKKSAGTQAYALFASATGSSSKV LSTIARHLTGCFAPNK TALHSATAV
HopT1-1	-31	MKTVSNHSIPSTNLVVDAGTETS AQKSQ PVCSEIQRNSKIEKAVIEHIAD
HopT1-2*	-26	
HopU1	-23	MNINRQLPVSGSERLLTPDVGVS RQACSERHYSTGQDRHDFYRFAARLHV
HopV1	-22	MRFDAARGQKPKAPMDAPSSRLRLRAI AGMPSEEAGTTAPADVNQPPPAD
HopX1	-34	MKIHNAGLTPPLPGISNGNVGKAAQSSITQP QSQQGSYGLPPESETRPD
HopY1	-25	MNITPLTSAAGKSSAQGTDKISIPNSTRM INAASIKWLNKVRSAISDHI
HopAA1-1	-33	MHINRRVQQPPVTATDSFR TASDASLASSSVRSVSSDQOREINAIADYLT
HopAA1-2	-29	
HopAB2	-43	MAGINRAGPSGAYFVGHTDPEPVSGQAHGSGSGASSNSPQVQPRPSNTP
HopAD1	-26	MLIGHSLHHRPTAVDSSLPTSATSQ TISNTKSRLDPHRVRELTFIGVGS
HopAF1	-43	MGLCISKHSGSSYSYSDSDRWQVPACPPNARSVSSHQTASASDIASGDVD
HopAG*	-15	
HopAH1	-39	MSMNTSVSNNGPVWSPVSSGNHAPSDFSGKSSSNAVHFLSPESAHRSPS
HopAH2-1*	-19	
HopAH2-2*	+2	
HopAI1	-11	MLALKLNTSIAQAPLKKNAEAE LRHMNHAEVRAHTPTRFTLNHRAPTIEV
HopAM1-1	-38	MHANPLSSFNRAQHGNLTNVEASQVKSAGT SSTTNIDSKNIEEHVADRLS

**Table 2. Cont.**

Substrate <sup>a</sup>	Score	Amino acid sequence used to construct the T3SS substrate training set
HopAM1-2*	-38	
HopAO1*	-29	
HopAQ1*	-33	
HopAS1*	-25	
PSPTO_0907*	-21	

<sup>a</sup>Experimentally validated T3SS substrates that were not included in the positive training set are denoted by asterisks.  
doi:10.1371/journal.pone.0036038.t002

value in identifying DC3000 T3SS substrates even when the positive training set was varied.

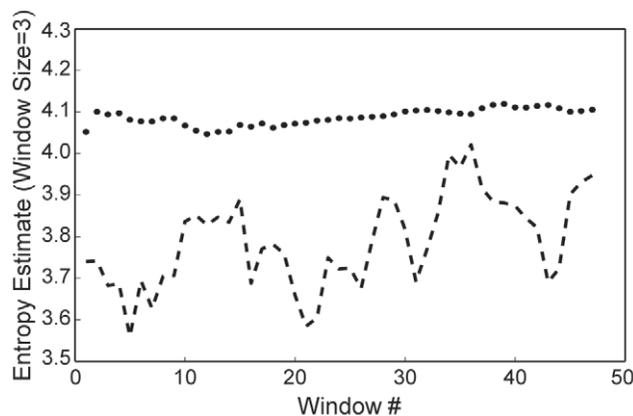
**TEREE algorithm performance on other bacterial genomes**

The universality of TERE was evaluated by conducting analyses on other bacterial genomes that encode T3SSs. In each case, the algorithm was trained on the T3SS substrate and background data sets from DC3000. First, we examined *P. syringae* pathovar phaseolicola (*P. s. phaseolicola*) strain 1448a, which is closely related to *P. s. tomato* DC3000, but has a different host range. Although *P. s. tomato* DC3000 and *P. s. phaseolicola* 1448a encode many homologous effectors, they also each express several distinct effectors [6,13,42,45]. TERE identified 78.1% of the known T3SS substrates in 1448a and had a specificity of 98.7% (Table 3 and Table S3). In addition, the PSPPH\_1525 and PSPPH\_A0133 proteins were classified by TERE as potential T3SS substrates. These proteins are likely to be effectors because they are both: i) encoded by genes that are regulated by HrpL [42], and ii) homologous to SKWP2, a verified effector protein in *Ralstonia solanacearum* [59,60]. When another T3SS computational SVM-based model called SIEVE (SVM-based Identification and Evaluation of Virulence Effectors) analyzed the 1448a genome, the results were more sensitive but less specific than those of the TERE algorithm (Table 3) [43]. We also compared TERE and SIEVE by determining the number of validated T3SS substrates within the top 50 scoring proteins. TERE recognized 20 1448a

T3SS substrates within the top 50 hits, whereas SIEVE identified only 9. Thus, TERE is more accurate than SIEVE at recognizing effectors in a bacterium that is closely related to DC3000.

TERE was also used to identify effectors in a more distantly related bacterium, *Salmonella enterica* serovar Typhimurium (*S. e. Typhimurium*) strain LT2. Although *P. syringae* and *S. enterica* are both in the  $\gamma$ -Proteobacteria, *S. e. Typhimurium* is an animal pathogen that causes a typhoid-like disease in mice and gastroenteritis in humans. In addition, *P. s. tomato* DC3000 and *S. e. Typhimurium* LT2 do not appear to have any effector genes in common. When TERE was used to identify T3SS substrates encoded by the LT2 genome, the sensitivity was 20.5% and the specificity was 98.8% (Table 3 and Table S4). In comparison, SIEVE recognized 86.4% of the LT2 T3SS substrates at a specificity of 91.9% [43]. When we lowered the specificity of TERE to 90.9%, the sensitivity rose to only 47.7%. Thus, SIEVE outperforms TERE on the *S. enterica* Typhimurium LT2 genome. However, both computational models identified a similar number of T3SS substrates within the top 50 highest scoring proteins (Table 3) [43].

TERE performance was also assessed on the *Ralstonia solanacearum* GMI1000 genome. This bacterium is a plant pathogen in the  $\beta$ -Proteobacteria and a more distant phylogenetic relative of *P. syringae* than *S. enterica*. Although the *R. solanacearum* GMI1000 and *P. s. tomato* DC3000 genomes encode several homologous effectors, these plant pathogens also secrete many distinct effectors [6,59]. Interestingly, the TERE algorithm was more effective at recognizing T3SS substrates in *Ralstonia* than in *Salmonella*, generating a sensitivity of 50.0% and specificity of 98.2% (Table 3). Although the sensitivity may seem low, it is important to note that within the top 50 hits, TERE identified 28 validated T3SS substrates, 2 putative T3SS substrates, and 2 secreted flagellar proteins (Table 3 and Table S5) [59–61]. In addition, TERE identified more than 25 *R. solanacearum* GMI1000 effectors that do not have homologs in DC3000 (Table S5). Another SVM-based computational T3SS substrate prediction model called BPBAac performed somewhat better than TERE, with a sensitivity of 63.8%, and a specificity of 99.0% (Table 3) [38]. BPBAac also identified 42 bona fide effectors within the top 50 hits of the algorithm [38]. Overall, these results indicate that TERE performance is in many respects comparable to other computational T3SS substrate prediction methods.



**Figure 4. Entropy estimates for the N-terminal regions of DC3000 T3SS substrates and nonsecreted proteins.** The dashed line represents the negative (background) training set, whereas the dotted line represents the T3SS substrate set. The estimates were calculated for residues 2–47 using a sliding window size of 3.  
doi:10.1371/journal.pone.0036038.g004

**Analysis of a potential T3SS targeting signal in FtsX**

Several of the proteins classified as T3SS substrates by TERE are not likely effector proteins because they have known or predicted intracellular functions. An example of such a protein is FtsX, the transmembrane component of an ABC transporter involved in cell division [62]. This protein was classified as a T3SS substrate by the TERE algorithm in *P. s. tomato* DC3000, *P. s.*

**Table 3.** Comparison of TERE to other computational T3SS substrate prediction models.

Genome	Method	Sensitivity <sup>a</sup>	Specificity <sup>b</sup>	# of known T3SS substrates in top 50 hits
<i>P. s. phaseolicola</i> 1448a	TEREE	78.1%	98.7%	20
	SIEVE <sup>c</sup>	87.5%	90.1%	9
<i>S. e. Typhimurium</i> LT2	TEREE	20.5%	98.8%	8
	SIEVE <sup>c</sup>	86.4%	91.9%	9
<i>R. solanacearum</i> GMI1000	TEREE	50.0%	98.2%	28
	BPBAc <sup>c</sup>	63.9%	99.0%	42

<sup>a</sup>Values were calculated by dividing the number of validated effectors, or true positives, by the sum of the true positives and false negatives. The second columns of Tables S3, S4, and S5 list the validated effectors for *P. s. phaseolicola* 1448a, *S. e. Typhimurium* LT2, and *R. solanacearum* GMI1000, respectively.

<sup>b</sup>Values were calculated by dividing the number of true negatives (non-substrates of the T3SS) by the sum of the false positives and true negatives.

<sup>c</sup>The sensitivity and specificity values for SIEVE and BPBAc were calculated based on published data sets [38,43].

doi:10.1371/journal.pone.0036038.t003

phaseolicola 1448a, and *S. e. Typhimurium* (Tables S1, S3, and S4). We reasoned that the N-terminal region of FtsX might contain functional T3SS targeting signals, while other features of the protein might prevent secretion. For example, the TMPred program ([http://www.ch.embnet.org/software/TMPRED\\_form.html](http://www.ch.embnet.org/software/TMPRED_form.html)) estimates that FtsX contains four hydrophobic segments that span the cytoplasmic membrane. These membrane spanning regions might prevent FtsX secretion despite the presence of N-terminal T3SS targeting signals.

To determine whether FtsX contains T3SS targeting signals, we created an FtsX-Cya hybrid protein. According to TMPred, the N-terminus of the DC3000 FtsX protein is located in the bacterial cytoplasm and the first membrane spanning segment begins at amino acid 69. We thus removed all of the membrane spanning domains by fusing the first 50 amino acids of FtsX to Cya. This protein was expressed in wild-type or  $\Delta hrp$  DC3000 strains, which were then inoculated into *N. benthamiana* leaves. As controls, we simultaneously measured the translocation of AvrPto<sub>(1-164)</sub>-Cya and AvrPto $\Delta_{2-12}$ -Cya by DC3000. The AvrPto $\Delta_{2-12}$ -Cya mutant lacks amino acids 2 to 12 of AvrPto, which removes most of the core signal required for targeting to the T3SS (Figure 1) [19]. Similar levels of cAMP were quantitated in *N. benthamiana* leaves inoculated with DC3000 strains expressing FtsX<sub>(1-50)</sub>-Cya or AvrPto $\Delta_{2-12}$ -Cya, indicating that the N-terminal region of FtsX does not contain a functional T3SS targeting signal (Table 4). The lack of AvrPto $\Delta_{2-12}$ -Cya and FtsX-Cya translocation was not due to poor protein expression or protein degradation, as both hybrids were detected in DC3000 (Figure 3). This finding highlights the

importance of subjecting the results of computational prediction programs to experimental testing.

### The extreme N-termini of unlikely T3SS substrates do not prevent secretion of AvrPto

A number of studies on different T3SS substrates have shown that the minimal signal for targeting to the T3SS is located within the first 15 amino acids (or codons) of substrates [16–19,21,63–67]. We thus hypothesized that the extreme N-termini of nonsecreted proteins might prevent secretion of AvrPto. To test this idea, the first 12 amino acids of AvrPto<sub>WT(1-164)</sub>-Cya were replaced with the first 12 amino acids of FtsX to yield AvrPto<sub>1-12FtsX</sub>-Cya (Figure 1). Another similar fusion was constructed in which the first 12 amino acids of AvrPto were replaced with the same region of PSPTO\_4342 (Figure 1). Because it is homologous to the TccB insecticidal toxin of *Photorhabdus luminescens*, we will refer to PSPTO\_4342 as TccB. We predicted that the AvrPto<sub>1-12TccB</sub>-Cya fusion would not be translocated into plant cells for two reasons: i) TccB had a score of +18 in our computational model (Table S1), considerably outside of the range for T3SS substrates, and ii) a TccB-Cya fusion was not translocated into *N. benthamiana* by DC3000 in a previous study [41]. Both the AvrPto<sub>1-12FtsX</sub>-Cya and AvrPto<sub>1-12TccB</sub>-Cya hybrid proteins were efficiently expressed in DC3000 (Figure 3).

When the AvrPto-Cya hybrids with mutant N-termini were tested for translocation by the DC3000 T3SS into *N. benthamiana*, we unexpectedly observed that both the AvrPto<sub>1-12FtsX</sub>-Cya and AvrPto<sub>1-12TccB</sub>-Cya mutants were effectively delivered into plant cells in a T3SS-dependent manner (Table 4). The levels of cAMP that accumulated for each mutant were not much lower than that of the positive control, AvrPto<sub>WT(1-164)</sub>-Cya. Therefore, the minimal secretion signal of AvrPto appears to tolerate a number of substitutions. AvrPto is still a T3SS substrate even when its core secretion signal is replaced with sequences from proteins that are not translocated by the T3SS into host cells.

**Table 4.** Translocation of unlikely T3SS substrates into *N. benthamiana* by *P. s. tomato* DC3000.

Cya fusion protein	Translocation by DC3000 (pmol cAMP/ $\mu$ g protein)	
	Wild-type	$\Delta hrp$
FtsX <sub>(1-50)</sub>	4.0 $\pm$ 0.5	0.3 $\pm$ 0.1
AvrPto <sub>WT(1-164)</sub>	155.4 $\pm$ 23.5	0.5 $\pm$ 0.1
AvrPto $\Delta_{2-12}$	7.5 $\pm$ 1.0	0.1 $\pm$ 0.0
AvrPto <sub>FtsX(1-12)</sub>	107.0 $\pm$ 8.2	0.0 $\pm$ 0.0
AvrPto <sub>TccB(1-12)</sub>	118.5 $\pm$ 8.5	0.0 $\pm$ 0.0

doi:10.1371/journal.pone.0036038.t004

### Comparison of the abilities of computational models to accurately predict T3SS substrates

In addition to TERE, SIEVE, and BPBAc, other computational models that predict T3SS substrates have been described [38,43,51,57,58,68]. Because most of these programs are accessible as web-based prediction tools, we determined whether they could accurately classify the Cya hybrid proteins examined in this study as T3SS substrates. All of the computational models correctly predicted that wild-type AvrPto is a T3SS substrate,

**Table 5.** Computational T3SS substrate predictions for proteins experimentally tested in this study.

Protein	Computational T3SS substrate predictions					Experimental results
	TEREE	SIEVE [36]	Effective T3 <sup>a</sup> [51]	T3SS prediction <sup>b</sup> [57]	T3SEdb [68]	
AvrPto <sub>WT</sub>	+	+	+	+/-	+	+
AvrPto <sub>SSM</sub>	+	+	+/-	+/-	+	+
FtsX (PSPTO_0429)	+	-	+/-	+/-	+	-
TccB (PSPTO_4342)	-	-	-	-/-	-	-
AvrPto <sub>Δ2-12</sub>	+	-	+	+/-	+	-
AvrPto <sub>FtsX(1-12)</sub>	+	-	+	+/-	+	+
AvrPto <sub>TccB(1-12)</sub>	+	-	+	+/-	+	+

<sup>a</sup>This model could be run at more stringent (selective) or less stringent (sensitive) settings. Symbols in this column indicate that the protein was classified as a T3SS substrate at: (+) the selective level, (+/-) the sensitive level, or (-) neither level.

<sup>b</sup>This model could be run using either an ANN or SVM classifier. Symbols in this column indicate that the protein was classified as a T3SS substrate using: (+/-) ANN, (-) SVM, (+/+) both, or (-/-) neither classifiers.

doi:10.1371/journal.pone.0036038.t005

and that TccB is not secreted by the T3SS (Table 5). However, none of the models were able to successfully classify all of the other mutant AvrPto proteins. Interestingly, FtsX was predicted to be a T3SS substrate by three computational models other than ours, despite the fact that DC3000 was not able to translocate FtsX<sub>(1-50)</sub>-Cya into plant cells. Thus, computational tools may be helpful in identifying potential T3SS substrates, but the results can be misleading.

## Discussion

Previous studies on AvrPto have shown that N-terminal amino acids are important for targeting to the T3SS. The first 15 amino acids of AvrPto are sufficient to target the Npt protein to the *Yersinia enterocolitica* T3SS for secretion into the extracellular milieu [19]. In DC3000, the first ~50 amino acids of AvrPto are required for efficient secretion and translocation of an AvrPto-Cya hybrid protein [40]. AvrPto also possesses the characteristic N-terminal amino acid patterns associated with proteins traveling the T3SS pathway. The vast majority of actively deployed *P. s. tomato* T3SS substrates contain (i) greater than 10% serine, (ii) an aliphatic amino acid or proline at position 3 or 4, and (iii) no negatively charged residues within the first 12 amino acids [41]. However, some *P. syringae* effectors and many of the T3SS substrates from animal pathogens lack one or more of these characteristic patterns.

In this study, we tested the functional significance of the *P. syringae* T3SS targeting patterns in AvrPto. We found that AvrPto secretion into the extracellular milieu and translocation into plants was unaffected by multiple mutations that removed the three major patterns (Figure 2, Table 1). In fact, even though the first 15 amino acids of AvrPto are sufficient to target the Npt protein to the *Yersinia enterocolitica* T3SS for secretion into the culture medium, we found that replacing the first 12 amino acids of AvrPto-Cya with the same regions of the nonsecreted FtsX or TccB proteins did not appreciably reduce translocation into plant cells (Table 4). Therefore, instead of relying on a single targeting signal, AvrPto may have several characteristics that additively or redundantly contribute to its recognition by the T3SS. This model is consistent with our previous findings that secretion and translocation efficiency increases for AvrPto-Cya hybrids that contain progressively larger portions of AvrPto [40]. One feature of AvrPto that may play a role in recognition by the T3SS is a pH-folding switch controlled by histidine 87. This switch allows AvrPto to maintain an unfolded conformation in the bacterial

cytoplasm [69]. Alternatively, AvrPto may interact with a chaperone that contributes to T3SS targeting. Another DC3000 effector, HopV, naturally lacks all three T3SS targeting patterns and interacts with the chaperone ShcV [70]. Thus, ShcV may compensate for a poor secretion signal by guiding HopV to the T3SS. However, there is currently no experimental evidence that chaperones mediate AvrPto secretion. Genes in the vicinity of *avrPto* do not possess features of T3SS chaperones, and promiscuous chaperones that interact with AvrPto have not been identified. In addition, AvrPto is secreted by *E. coli* containing a plasmid expressing the *hrp/hrc* T3SS gene cluster from *Dickeya dadantii* [71]. Thus, if AvrPto binds a chaperone, it is most likely encoded within the *hrp/hrc* gene cluster and conserved between *P. syringae* and *D. dadantii*.

Although our experimental analysis of T3SS secretion signals was limited to AvrPto, substantial changes have been made to the N-termini of several other effectors without radically reducing secretion. For example, AvrBs2 is delivered into plant cells by *Xanthomonas campestris* even when it contains frameshift mutations that alter the sequence of its first 18 amino acids [20]. Furthermore, YopE and YopD mutants that contain synthetic amphipathic amino acid sequences in their extreme N-termini are still secreted by the *Y. pseudotuberculosis* T3SS [22,34,67]. It has been proposed that substrate recognition by the T3SS is influenced by accessory proteins as well as the overall physical properties of substrates, rather than specific amino acid sequences [2]. Thus, it is possible that the AvrPto<sub>1-12</sub>FtsX-Cya and AvrPto<sub>1-12</sub>TccB-Cya hybrid proteins are translocated into plants because the FtsX or TccB amino acid sequences do not appreciably affect the structure of the AvrPto N-terminus.

To further examine compositional differences between DC3000 T3SS substrates and nonsecreted proteins, we employed a computational approach. According to our analysis, the amino acid sequences of T3SS targeting signals are substantially different than nonsecreted proteins (Figure 4). Other computational analyses have also recognized differences between the compositions of T3SS substrates and nonsecreted proteins [38,51]. These differences were exploited to develop a computational algorithm based on a symmetric version of the Kullback-Liebler distance [50]. Unlike other computational T3SS substrate prediction algorithms that are based on SVM, ANN, or Naïve Bayesian classifiers, our method is based on information theory [38,43,51,57,58,68]. The algorithm, called TERE, distinguishes between T3SS substrates and other DC3000 proteins by

calculating differences in relative entropy. The TERE algorithm differentiated T3SS substrates in DC3000 with a high sensitivity; only two known effector proteins were not scored as positives (Table 2, Table S1). Another remarkable feature of TERE is its high specificity. In other words, the majority of the top hits of the algorithm were known effectors, and only about 1% of the proteins in the DC3000 genome were scored as false positives.

Although TERE performed extremely well in DC3000, its effectiveness in other bacteria varied (Table 3). The algorithm was efficient at recognizing effectors in *P. s. phaseolicola* 1448a and *R. solanacearum* GMI1000, but not in *S. e. Typhimurium* LT2. These results might be explained by the fact that *P. syringae* and *R. solanacearum* have several homologous effector genes [6,59]. However, TERE identified more than 25 *R. solanacearum* T3SS substrates that are not found in DC3000. Thus, the success of TERE in *R. solanacearum* is not simply due to common effector genes. In contrast, *P. s. tomato* DC3000 and *S. e. Typhimurium* have different pathogenic lifestyles and completely distinct sets of effectors. Many *S. enterica* T3SS effectors function to promote bacterial entry into intestinal epithelial cells or survival within macrophages, while *P. syringae* effectors primarily suppress plant defense responses [72,73]. TERE performance on the *S. e. Typhimurium* genome thus might be improved by including *Salmonella* or other animal pathogen effectors in the T3SS substrate training set. Another reason that TERE may not be as effective in *S. e. Typhimurium* is that *P. syringae* and *S. enterica* effectors have different amino acid biases. A recent analysis reported that plant pathogens contain more alanine, proline, and arginine in their effector targeting signals than animal pathogens [38]. In addition, animal pathogen effectors are more enriched in isoleucine, asparagine, and threonine than plant pathogen effectors [38]. Including animal pathogen effectors in the T3SS substrate training set for the TERE algorithm might also compensate for this problem.

One false positive that was recognized as a T3SS substrate in several iterations of the TERE algorithm was FtsX, a transmembrane protein that functions in cell division [62]. To explain these results, we reasoned that the N-terminal region of FtsX may possess a T3SS targeting signal that is obstructed by other features of the protein. In fact, when YopE is fused to a tightly folded protein such as dihydrofolate reductase (DHFR), it is rejected as a T3SS substrate [74,75]. However, the first 50 amino acids of FtsX did not target the Cya reporter protein to the T3SS for translocation into plant cells (Table 4). Thus, even though the TERE algorithm is quite sensitive, it does not rule out all nonsecreted proteins as T3SS substrates. TERE is not alone in this regard. Several other computational T3SS substrate prediction programs were unable to precisely predict the secretion status of all the mutant AvrPto-Cya proteins examined in this study (Table 5).

In conclusion, advances in genome sequencing technologies have led to the availability of many new bacterial genome sequences. Computational T3SS substrate prediction models will be useful tools for identifying new effector genes within the genomes of bacteria that contain T3SSs. Our results show that the TERE algorithm performed well on the genomes of three plant pathogens. No computational T3SS substrate prediction model is 100% accurate at identifying effector genes [38,43,51,57,58,68]. Thus, comparing the results of a few different computational models and constructing a short-list of common hits may be the most effective way to identify potential T3SS effector candidates within bacterial genome sequences.

## Materials and Methods

### Bacterial strains and growth conditions

The *P. syringae* strains used in this study are listed in Table 6 and were grown in King's B medium (KB) at 29°C [76] or *hup*-derepressing minimal medium supplemented with fructose (HDM) at 22°C [77]. *Escherichia coli* DH5 $\alpha$  or TOP10 strains were used for cloning and propagating plasmids. They were grown in Luria-Bertani or Terrific Broth at 37°C [78]. Antibiotics were used at the following concentrations: ampicillin, 100  $\mu$ g/ml; chloramphenicol, 20  $\mu$ g/ml; gentamicin, 10  $\mu$ g/ml; kanamycin, 50  $\mu$ g/ml; rifampin, 50  $\mu$ g/ml; spectinomycin, 50  $\mu$ g/ml.

### Construction of plasmids

pBBR1-based plasmids that express FLAG-tagged versions of wild-type and mutant AvrPto proteins were constructed in several steps. First, *avrPto* from *P. syringae* pv. tomato JL1065 was amplified by PCR using the primers P830C and P403C (Table 7). The product was digested with *Nde*I and *Sal*I and cloned into pFLAG-CTC. The resulting plasmid, pCPP3156, encodes an AvrPto protein that lacks amino acids 2–12 and contains a FLAG epitope (DYKDDDDK) at its C-terminus. The plasmid also contains a single point mutation that introduces an *Hpa*I cleavage site between codons 15 and 16 of *avrPto*, but does not change the amino acid sequence of AvrPto. The *avrPto*<sub>A2-12-FLAG</sub> sequence from pCPP3156 was then subcloned into pBBR1-MCS5 to create pCPP3178 (Table 6). To construct pCPP3384, which encodes AvrPto<sub>WT</sub>, pCPP3178 was digested with *Nde*I and *Hpa*I and ligated to a double-stranded DNA fragment formed by the hybridization of P831C and P832C (Table 7). Plasmids pLMS153 and pLMS154 were constructed in a similar manner, except that the double-stranded DNA fragments were formed by the hybridization P154 and P155, and P156 and P157, respectively (Table 7). To create pCPP3407, which expresses AvrPto<sub>SSM</sub>, pCPP3384 was digested with *Hpa*I and *Bsp*I and ligated to a double-stranded DNA fragment that was formed by hybridizing four overlapping oligonucleotides designated APS1, APS2, APS3, and APS4 (Table 7). All oligonucleotides were phosphorylated by T4 polynucleotide kinase prior to hybridization.

The plasmids that express full length AvrPto<sub>WT</sub>, AvrPto<sub>SSM</sub>, AvrPto<sub>A2-12</sub>, AvrPto<sub>2-12FtsX</sub>, or AvrPto<sub>2-12TccB</sub> fused to Cya (pND4, pND2, pLMS155, pLMS157, and pLMS158, respectively) were constructed in two steps. First, *avrPto* sequences were amplified from pCPP3178, pCPP3384, pCPP3407, pLMS153, or pLMS154 using the primers P1 and P3 (Table 7). Next, the PCR products were digested with *Xba*I and *Xma*I, and ligated to pCPP3214 digested with the same enzymes. The plasmids that express the first 50 amino acids of AvrPto<sub>WT</sub> or AvrPto<sub>SSM</sub> fused to Cya (pND3 and pND1, respectively) were constructed in a similar manner, except that P1 and P2 were used to amplify *avrPto* sequences from pCPP3384 or pCPP3407.

The plasmid that encodes the FtsX-Cya fusion protein (pCPP5170) was constructed using Gateway cloning technology (Invitrogen). *PSPTO\_0429* sequences were amplified from DC3000 chromosomal DNA by PCR using P1211C and P1256C (Table 7). The PCR product was then cloned into pENTR/SD/D-TOPO to create the entry vector pCPP5168. A recombination (or LR) reaction between the entry vector and the destination vector pCPP3234 was then performed to create pCPP5170 (Table 6).

### DNA manipulations and sequencing

Plasmid DNA was isolated and manipulated according to standard protocols [79]. T4 polynucleotide kinase (New England

**Table 6.** Bacterial strains and plasmids used in this study.

Strain or Plasmid	Genotype or relevant characteristics <sup>a</sup>	Source
<i>E. coli</i>		
DH5 $\alpha$	F <sup>-</sup> $\Phi$ 80 <i>lacZ</i> $\Delta$ M15 $\Delta$ ( <i>lacZYA-argF</i> )U169 <i>deoR recA1endA1 hsdR17</i> ( <i>r<sub>K</sub><sup>-</sup> m<sub>K</sub><sup>+</sup></i> ) <i>phoA supE44 thi-1 gyrA96 relA1 <math>\lambda</math><sup>-</sup></i>	Invitrogen
TOP10	F <sup>-</sup> <i>mcrA</i> $\Delta$ ( <i>mrr-hsdRMS-mcrBC</i> ) $\Phi$ 80 <i>lacZ</i> $\Delta$ M15 $\Delta$ <i>lacX74 recA1 araD139</i> $\Delta$ ( <i>ara-leu</i> )7697 <i>galE15 galK16 rpsL endA1 nupG</i>	Invitrogen
<i>P. syringae</i> pv. tomato		
DC3000	Wild type; Rf <sup>r</sup>	[83]
CUCPB5114	DC3000 $\Delta$ <i>hrpK-hrpR::<math>\Omega</math>Cm</i> ; Rf <sup>r</sup> , Cm <sup>r</sup>	[48]
Plasmids		
pUFR034	Broad-host-range vector; Km <sup>r</sup>	[84]
pFLAG-CTC	Vector for expression of C-terminal FLAG fusion proteins; Ap <sup>r</sup>	Sigma-Aldrich
pCPP3156	pFLAG-CTC:: <i>avrPto</i> <sub><math>\Delta</math>12-12</sub>	This work
pBBR1MCS-5	Broad-host-range expression vector containing P <sub><i>lacI</i></sub> ; Gm <sup>r</sup>	This work
pCPP3178	pBBR1MCS-5:: <i>avrPto</i> <sub><math>\Delta</math>12-12</sub> -FLAG	This work
pCPP3384	pBBR1MCS-5:: <i>avrPto</i> <sub>WT</sub> -FLAG	This work
pCPP3407	pBBR1MCS-5:: <i>avrPto</i> <sub>SSM</sub> -FLAG	This work
pLMS153	pBBR1MCS-5:: <i>avrPto</i> <sub>2-12ftsX</sub> -FLAG	This work
pLMS154	pBBR1MCS-5:: <i>avrPto</i> <sub>2-12tccB</sub> -FLAG	This work
pENTR/SD/D-TOPO	Gateway entry vector; Km <sup>r</sup>	Invitrogen
pCPP5168	pENTR/SD/D-TOPO:: <i>ftsX</i> <sub>(1-50)</sub>	This work
pCPP3214	Vector for expression of C-terminal Cya fusion proteins; Sp <sup>r</sup>	[40]
pCPP3234	Gateway destination vector version of pCPP3214; Sp <sup>r</sup> , Cm <sup>r</sup>	[40]
pCPP5170	pCPP3234:: <i>ftsX</i> <sub>(1-50)</sub>	This work
pND1	pCPP3214:: <i>avrPto</i> <sub>SSM</sub> (1-50)	This work
pND2	pCPP3214:: <i>avrPto</i> <sub>SSM</sub> (1-164)	This work
pND3	pCPP3214:: <i>avrPto</i> <sub>WT</sub> (1-50)	This work
pND4	pCPP3214:: <i>avrPto</i> <sub>WT</sub> (1-164)	This work
pLMS155	pCPP3214:: <i>avrPto</i> <sub><math>\Delta</math>2-12</sub>	This work
pLMS157	pCPP3214:: <i>avrPto</i> <sub>2-12ftsX</sub>	This work
pLMS158	pCPP3214:: <i>avrPto</i> <sub>2-12tccB</sub>	This work

<sup>a</sup>Rf<sup>r</sup>, Cm<sup>r</sup>, Ap<sup>r</sup>, Gm<sup>r</sup>, Sp<sup>r</sup>, and Km<sup>r</sup> indicate resistance to rifampicin, chloramphenicol, ampicillin, gentamicin, spectinomycin, and kanamycin, respectively.  
doi:10.1371/journal.pone.0036038.t006

Biolabs), restriction enzymes (New England Biolabs), and DNA ligase (Takara) were used according to the manufacturer's protocols. PCR was performed with either ExTaq (Takara) or Vent (New England Biolabs), and oligonucleotide primers were obtained from Integrated DNA Technologies (IDT). All cloned PCR products were sequenced to ensure that no mutations were introduced. DNA sequencing was performed at either the Cornell University Life Sciences Core Laboratories Center or the University of Missouri DNA Core Facility using an Applied Biosystems 3730 DNA analyzer (Applied Biosystems).

**Secretion assays, protein sample preparation, and immunoblot analysis**

Secretion assays were carried out using a previously described procedure [40]. Cya hybrid protein expression from plasmids was monitored by inoculating *P. s.* tomato DC3000 strains into KB containing spectinomycin and 200  $\mu$ M isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG). Cultures were grown at 28°C for 4 h, and bacteria were pelleted and suspended in protein sample buffer. Equal amounts of cells, based on OD<sub>600</sub>, were loaded onto an SDS-PAGE gel. Following separation by electrophoresis and

transfer onto a nitrocellulose membrane, proteins were detected using a standard Western analysis procedure [79]. Primary antibodies, either anti-FLAG M2 mouse monoclonal immunoglobulin G (IgG) (Sigma-Aldrich), anti-Cya (3D1) mouse monoclonal IgG (Santa Cruz Biotechnology), or anti-NptII rabbit polyclonal IgG (United States Biological, Swampscott, MA), were used at 1:5000. Secondary anti-mouse or anti-rabbit IgG-horseradish peroxidase conjugate antibodies (Sigma-Aldrich) were used at 1:30,000. Blots were developed using the Pierce Super-Signal West Pico chemiluminescent substrate (Thermo Fisher Scientific).

**Adenylate cyclase assays**

Cyclic AMP levels in infected *N. benthamiana* leaf tissue were determined as previously described [40,80]. Briefly, *P. syringae* strains were grown as lawns on KB plates and then suspended to an OD<sub>600</sub> of 0.3 (~1 $\times$ 10<sup>8</sup> cfu/ml) in 10 mM MgCl<sub>2</sub>-100 mM sucrose solution supplemented with 100  $\mu$ M isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG). Bacteria were infiltrated into the third or fourth oldest leaves of *N. benthamiana* with a blunt syringe, and plants were incubated in a growth chamber set at 23°C and

**Table 7.** Oligonucleotides used in this study.

Name	Sequence <sup>a</sup>
P403C	5'-ATTGTA <u>GTCGACTT</u> GCCAGTTACGGTACGGG-3'
P830C	5'-GCGATACATATGCATCAGTTAACTCCCCAGACCGAGT-3'
P831C	5'-TATGGGAAATATATGTGTCGGCGGATCCAGGATGGCCCATCAGGTT-3'
P832C	5'-AACCTGATGGGCCATCTGGATCCGCCGACACATATTTCCCA-3'
P154	5'-TATGAGTGCACACGCAGCCCAAGGTTTCAGAGCGCCATCAGGTT-3'
P155	5'-AACCTGATGGCGCTCTGAAACCTTGGGGCTGCGTGTGGCACTCA-3'
P156	5'-TATGTCGGATACCCCTTGAAGCCGGCTCAACGAATCTCATCAGGTT-3'
P157	5'-AACCTGATGAGATTCGTTGAGCCGCTTTCAAGGTATCGGACA-3'
APS1	5'-AACGCCCCAGACCGAGTTGCCAACAAAC-3'
APS2	5'-GCGGGTACGAAGATAACGTAACGGCCGCCAACTGC-3'
APS3	5'-CTTCGTACCCGCGTTGTTGGCAACTCGGTCTGGGGCGTT-3'
APS4	5'-TCAGCAGTTGGGGCGCCGTTACGTTAT-3'
P1	5'-CGGTTCTAGAACAAATTCACACAGGAG-3'
P2	5'-TAATATCCCGGGTGGTAGACCAGCAGACTC-3'
P3	5'-ATTTAAACCCGGTTGCCAGTTACGGTACG-3'
P1211C	5'-CACCATGAGTGCACACGCA-3'
P1256C	5'-ATGACTCTCCAGCCAGGCAG-3'

<sup>a</sup>Important restriction enzyme sites are underlined.  
doi:10.1371/journal.pone.0036038.t007

80% humidity, with a 16 h/8 h light/dark cycle. Two leaf disks were collected from each infiltrated area 6 h post-inoculation with a 0.8-cm-diameter cork borer. Leaf disks were then frozen in liquid nitrogen, ground to a powder, and suspended in 300 µl of 0.1 M HCl. cAMP was quantitated using a cAMP ELISA assay kit (Enzo Life Sciences) and protein levels were determined by Bradford assay (Bio-Rad) according to the manufacturer’s directions.

### Computational analysis of T3SS substrates

To characterize the composition of T3SS substrates, we divided the protein coding sequences of *P. s. tomato* DC3000 into two groups: i) a positive training set consisting of the amino acid sequences of 38 experimentally tested T3SS substrates (Table 2), and ii) the remaining ~5600 protein sequences, which were used for background statistics. For the TERE analysis, we extracted the first 50 amino acids of each sequence.

The block entropy calculation referred to in Figure 4 was accomplished by applying a sliding window to the positive training set and the background set. Let *W* represent the window size, *N* represent the number of amino acids (i.e. *N* = 20) and *M* represent the number of sequences in a given set. Under these circumstances, an *M* × *W* block of symbols was examined starting at sequence position *m* and ending at sequence position *m* + *W* – 1. The block symbol probability at the *m*<sup>th</sup> position, *p<sub>i</sub>*, was then estimated as *p<sub>i</sub>* = *n<sub>i</sub>* / (*MW*), where *n<sub>i</sub>* is the number of times the *i*<sup>th</sup> amino acid appears in the block for *i* = 1, ..., *N*. The following equation was then used to determine entropy (*H<sub>e</sub>*) estimates for each window:

$$H_e = - \sum_{i=1}^N p_i \log_2 p_i \quad (1)$$

To identify T3SS substrates within bacterial genomes, the TERE algorithm applies a symmetric version of the Kullback-

Liebler distance [50]. Given two discrete probability mass functions *P* and *Q* each containing *N* elements, the symmetric Kullback-Liebler distance is defined as [49]:

$$D_s(P||Q) \equiv D(P||Q) + D(Q||P) \quad (2)$$

where

$$D(P||Q) = - \sum_{i=1}^N p_i \log_2 \frac{p_i}{q_i} \quad (3)$$

is generally referred to as the relative entropy.

To characterize a protein sequence of unknown classification as being close or far from the substrate distribution, *D<sub>s</sub>* was evaluated over a series of sliding windows of length *W*. Given the window size, there were *K* = *L* - *W* + 1 positions to consider where *L* = 50. At each window position, three discrete probability mass functions (*Q*, *P*<sup>1</sup>, *P*<sup>2</sup>) were computed: i) *Q* was constructed by computing *q<sub>i</sub>* = *n<sub>i</sub>* / *W* (*i* = 1, ..., 20) for the sequence of unknown classification, ii) *P*<sup>1</sup> represents the background probability mass function, and iii) *P*<sup>2</sup> represents the T3SS substrate probability mass function derived from *P. s. tomato* DC3000 sequences (Table 2). For the background and substrate distributions, similar to the block entropy calculation, we estimated the symbol probability as *p<sub>i</sub>* = *n<sub>i</sub>* / (*MW*), where *n<sub>i</sub>* was the number of times amino acid *i* appeared in the window and *M* represents the number of sequences in a given set.

Given these distributions, the TERE algorithm then calculated *D<sub>s</sub>*(*P<sup>k</sup>* | *Q*) for *k* = 1, 2 where *P*<sup>1</sup> was the background distribution and *P*<sup>2</sup> was the T3SS substrate distribution derived from *P. s. tomato* DC3000 sequences (Table 2). Finally, category 2 was chosen if *D<sub>s</sub>*(*P*<sup>1</sup> | *Q*) > *D<sub>s</sub>*(*P*<sup>2</sup> | *Q*); otherwise, category 1 was chosen. To decide upon the class membership of a given sequence, the choice for each of the *K* windows was examined and the majority was chosen. In other words, over *K* instances there were *k<sub>1</sub>* instances in favor of the background and *k<sub>2</sub>* instances in favor of the substrate distribution. A score *S* was created by taking the difference *S* = *k<sub>1</sub>* - *k<sub>2</sub>*. For the purposes of robustness, we ran our algorithm three times with window sizes *W* = 1, 2, 3. For each sequence tested, we took the minimum score from each of the three tests. All computations for this work were performed using MATLAB.

### Performance Evaluation

The performance of TERE was evaluated by calculating three measures: i) sensitivity, or the number of true positives divided by the sum of the true positives and false negatives (TP / (TP + FN)), ii) specificity, or the number of true negatives divided by the sum of the false positives and true negatives (TN / (TN + FP)), and iii) the area under the ROC curve (AUC) that is generated when the sensitivity and specificity are plotted against each other for each output score. The AUC represents the probability that TERE algorithm will rank a randomly chosen positive sequence at a score less than a randomly chosen negative sequence (Table S2). Specifically, the Wilcoxon Rank Sum Test was applied in order to compute the AUC [81,82]. The 5-fold cross validation test was performed by creating 5 different T3SS substrate training sets that each lacked 7–8 different effector sequences. Each different training set was then utilized by TERE to analyze DC3000 coding regions, and the sensitivity and specificity were calculated for each run.

**Supporting Information**

**Table S1** TERE algorithm scores for annotated coding regions in *P. s. tomato* DC3000. (XLSX)

**Table S2** TERE algorithm scores for proteins in T3SS substrate and negative training sets. (XLSX)

**Table S3** TERE algorithm scores for annotated coding regions in *P. s. phaseolicola* 1448a. (XLSX)

**Table S4** TERE algorithm scores for annotated coding regions in *S. e. Typhimurium* LT2. (XLSX)

**References**

1. Macnab RM (2003) How bacteria assemble flagella. *Annu Rev Microbiol* 57: 77–100.
2. Galán JE, Wolf-Watz H (2006) Protein delivery into eukaryotic cells by type III secretion machines. *Nature* 444: 567–573.
3. Akopyan K, Edgren T, Wang-Edgren H, Rosqvist R, Fahlgren A, et al. (2011) Translocation of surface-localized effectors in type III secretion. *Proc Natl Acad Sci USA* 108: 1639–1644.
4. Vidal JE, Navarro-García F (2008) EspC translocation into epithelial cells by enteropathogenic *Escherichia coli* requires a concerted participation of type V and III secretion systems. *Cell Microbiol* 10: 1975–1986.
5. Galán JE (2009) Common themes in the design and function of bacterial effectors. *Cell Host Microbe* 5: 571–579.
6. Grant SR, Fisher EJ, Chang JH, Mole BM, Dangl JL (2006) Subterfuge and manipulation: type III effector proteins of phytopathogenic bacteria. *Annu Rev Microbiol* 60: 425–449.
7. Zhou D, Chen L-M, Hernandez L, Shears SB, Galán JE (2001) A *Salmonella* inositol polyphosphatase acts in conjunction with other bacterial effectors to promote host cell actin cytoskeleton rearrangements and bacterial internalization. *Mol Microbiol* 39: 248–260.
8. Zhang S, Santos RL, Tsois RM, Stender S, Hardt W-D, et al. (2002) The *Salmonella enterica* serotype Typhimurium effector proteins SipA, SopA, SopB, SopD, and SopE2 act in concert to induce diarrhea in calves. *Infect Immun* 70: 3843–3855.
9. Kvitko BH, Park DH, Velázquez AC, Wei C-F, Russell AB, et al. (2009) Deletions in the repertoire of *Pseudomonas syringae* pv. tomato DC3000 type III secretion effector genes reveal functional overlap among effectors. *PLoS Pathog* 5: e1000388.
10. Roden JA, Belt B, Ross JB, Tachibana T, Vargas J, et al. (2004) A genetic screen to isolate type III effectors translocated into pepper cells during *Xanthomonas* infection. *Proc Natl Acad Sci U S A* 101: 16624–16629.
11. Guttman DS, Vinatzer BA, Sarkar SF, Ranall MV, Kettler G, et al. (2002) A functional screen for the Type III (Hrp) secretome of the plant pathogen *Pseudomonas syringae*. *Science* 295: 1722–1726.
12. Subtil A, Delevoeye C, Balana ME, Tastevin L, Perrinet S, et al. (2005) A directed screen for chlamydial proteins secreted by a type III mechanism identifies a translocated protein and numerous other new candidates. *Mol Microbiol* 56: 1636–1647.
13. Chang JH, Urbach JM, Law TF, Arnold LW, Hu A, et al. (2005) A high-throughput, near-saturating screen for type III effector genes from *Pseudomonas syringae*. *Proc Natl Acad Sci USA* 102: 2549–2554.
14. Niemann GS, Brown RN, Gustin JK, Stufkens A, Shaikh-Kidwai AS, et al. (2011) Discovery of novel secreted virulence factors from *Salmonella enterica* serovar Typhimurium by proteomic analysis of culture supernatants. *Infect Immun* 79: 33–43.
15. Deng W, de Hoog CL, Yu HB, Li Y, Croxen MA, et al. (2010) A comprehensive proteomic analysis of the type III secretome of *Citrobacter rodentium*. *J Biol Chem* 285: 6790–6800.
16. Sory M-P, Boland A, Lambermont I, Cornelis GR (1995) Identification of the YopE and YopH domains required for secretion and internalization into the cytosol of macrophages, using the *cyoA* gene fusion approach. *Proc Natl Acad Sci USA* 92: 11998–12002.
17. Schesser K, Frithz-Lindsten E, Wolf-Watz H (1996) Delineation and mutational analysis of the *Yersinia pseudotuberculosis* YopE domains which mediate translocation across bacterial and eukaryotic cellular membranes. *J Bacteriol* 178: 7227–7233.
18. Anderson DM, Schneewind O (1997) A mRNA signal for the type III secretion of Yop proteins by *Yersinia enterocolitica*. *Science* 278: 1140–1143.
19. Anderson DM, Fouts DE, Collmer A, Schneewind O (1999) Reciprocal secretion of proteins by the bacterial type III machines of plant and animal pathogens suggests universal recognition of mRNA targeting signals. *Proc Natl Acad Sci USA* 96: 12839–12843.

**Table S5** TERE algorithm scores for annotated coding regions in *R. solanacearum* GMI1000. (XLSX)

**Acknowledgments**

We thank Monica Moll, Nichole Deluca, and Matthew Sidwell for assistance with plasmid construction.

**Author Contributions**

Conceived and designed the experiments: LMS DJS AC ES. Performed the experiments: LMS JCV. Analyzed the data: LMS ES. Contributed reagents/materials/analysis tools: LMS JCV ES. Wrote the paper: LMS ES. Designed the computational algorithm used in analysis: DJS ES.

20. Mudgett MB, Chesnokova O, Dahlbeck D, Clark ET, Rossier O, et al. (2000) Molecular signals required for type III secretion and translocation of the *Xanthomonas campestris* AvrBs2 protein to pepper plants. *Proc Natl Acad Sci USA* 97: 13324–13329.
21. Rüssmann H, Kubori T, Sauer J, Galán JE (2002) Molecular and functional analysis of the type III secretion signal of the *Salmonella enterica* InvJ protein. *Mol Microbiol* 46: 769–779.
22. Lloyd SA, Norman M, Rosqvist R, Wolf-Watz H (2001) *Yersinia* YopE is targeted for type III secretion by N-terminal, not mRNA, signals. *Mol Microbiol* 39: 520–532.
23. Ghosh P (2004) Process of protein transport by the type III secretion system. *Microbiol Mol Biol Rev* 68: 771–795.
24. Lee SH, Galán JE (2004) *Salmonella* type III secretion-associated chaperones confer secretion-pathway specificity. *Mol Microbiol* 51: 483–495.
25. Lilić M, Vujanac M, Stebbins CE (2006) A common structural motif in the binding of virulence factors to bacterial secretion chaperones. *Mol Cell* 21: 653–664.
26. Higashide W, Zhou D (2006) The first 45 amino acids of SopA are necessary for InvB binding and SPI-1 secretion. *J Bacteriol* 188: 2411–2420.
27. Spaeth KE, Chen Y-S, Valdivia RH (2009) The *Chlamydia* type III secretion system C-ring engages a chaperone-effector protein complex. *PLoS Pathog* 5: e1000579.
28. Gauthier A, Finlay BB (2003) Translocated intimin receptor and its chaperone interact with ATPase of the type III secretion apparatus of enteropathogenic *Escherichia coli*. *J Bacteriol* 185: 6747–6755.
29. Akeda Y, Galán JE (2005) Chaperone release and unfolding of substrates in type III secretion. *Nature* 437: 911–915.
30. Woestyn S, Sory M-P, Boland A, Lequenne O, Cornelis GR (1996) The cytosolic SycE and SycH chaperones of *Yersinia* protect the region of YopE and YopH involved in translocation across the eukaryotic cell membranes. *Mol Microbiol* 20: 1261–1271.
31. Boyd AP, Lambermont I, Cornelis GR (2000) Competition between the Yops of *Yersinia enterocolitica* for delivery into eukaryotic cells: role of the SycE chaperone binding domain of YopE. *J Bacteriol* 182: 4811–4821.
32. Buchko GW, Niemann G, Baker ES, Belov ME, Smith RD, et al. (2010) A multi-pronged search for a common structural motif in the secretion signal of *Salmonella enterica* serovar Typhimurium type III effector proteins. *Mol BioSyst* 6: 2448–2458.
33. Gazi AD, Charova SN, Panopoulos NJ, Kokkinidis M (2009) Coiled-coils in type III secretion systems: structural flexibility, disorder and biological implications. *Cell Microbiol* 11: 719–729.
34. Lloyd S, Sjöström M, Andersson S, Wolf-Watz H (2002) Molecular characterization of type III secretion signals via analysis of synthetic N-terminal amino acid sequences. *Mol Microbiol* 43: 51–59.
35. Arnold R, Jehl A, Rattei T (2010) Targeting effectors: the molecular recognition of type III secreted proteins. *Microbes Infect* 12: 346–358.
36. McDermott JE, Corrigan A, Peterson E, Oehmen C, Niemann G, et al. (2011) Computational prediction of type III and IV secreted effectors in Gram-negative bacteria. *Infect Immun*. IAL00537-00510 p.
37. Greenberg JT, Vinatzer BA (2003) Identifying type III effectors of plant pathogens and analyzing their interaction with plant cells. *Curr Opin Microbiol* 6: 20–28.
38. Wang Y, Zhang Q, Sun M, Guo D (2011) High-accuracy prediction of bacterial type III secreted effectors based on position-specific amino acid composition profiles. *Bioinformatics* 27: 777–784.
39. Petnicki-Owiczka T, Schneider DJ, Tam VC, Chanccy ST, Shan L, et al. (2002) Genomewide identification of proteins secreted by the Hrp type III protein secretion system of *Pseudomonas syringae* pv. tomato DC3000. *Proc Natl Acad Sci USA* 99: 7652–7657.

40. Schechter LM, Roberts KA, Jamir Y, Alfano JR, Collmer A (2004) *Pseudomonas syringae* type III secretion system targeting signals and novel effectors studied with a Cya translocation reporter. *J Bacteriol* 186: 543–555.
41. Schechter LM, Vencato M, Jordan KL, Schneider SE, Schneider DJ, et al. (2006) Multiple approaches to a complete inventory of *Pseudomonas syringae* pv. tomato DC3000 type III secretion system effector proteins. *Mol Plant-Microbe Interact* 19: 1180–1192.
42. Vencato M, Tian F, Alfano J, Buell C, Cartinhour S, et al. (2006) Bioinformatics-enabled identification of the HrpL regulon and type III secretion system effector proteins of *Pseudomonas syringae* pv. phaseolicola 1448A. *Mol Plant-Microbe Interact* 19: 1193–1206.
43. Samudrala R, Heffron F, McDermott JE (2009) Accurate prediction of secreted substrates and identification of a conserved putative secretion signal for type III secretion systems. *PLoS Pathog* 5: e1000375.
44. Buell C, Joardar V, Lindeberg M, Selengut J, Paulsen I, et al. (2003) The complete genome sequence of the Arabidopsis and tomato pathogen *Pseudomonas syringae* pv. tomato DC3000. *Proc Natl Acad Sci USA* 100: 10181–10186.
45. Lindeberg M, Cartinhour S, Myers CR, Schechter LM, Schneider DJ, et al. (2006) Closing the circle on the discovery of genes encoding Hrp regulon members and type III secretion system effectors in the genomes of three model *Pseudomonas syringae* strains. *Mol Plant-Microbe Interact* 19: 1151–1158.
46. Cumnac S, Lindeberg M, Collmer A (2009) *Pseudomonas syringae* type III secretion system effectors: repertoires in search of functions. *Curr Opin Microbiol* 12: 53–60.
47. Zong N, Xiang T, Zou Y, Chai J, Zhou J-M (2008) Blocking and triggering of plant immunity by *Pseudomonas syringae* effector AvrPto. *Plant Signal Behav* 3: 583–585.
48. Fouts DE, Badel JL, Ramos AR, Rapp RA, Collmer A (2003) A *Pseudomonas syringae* pv. tomato DC3000 Hrp (type III secretion) deletion mutant expressing the Hrp system of bean pathogen *P. syringae* pv. *syringae* 61 retains normal host specificity for tomato. *Mol Plant-Microbe Interact* 16: 43–52.
49. Cover TM, Thomas JA (1991) Elements of Information Theory. New York: J. Wiley & Sons. xxii, 542 p.
50. Sakk E, Schneider D, Vencato M, Collmer A, Cartinhour S (2005) Computational identification and characterization of type III secretion substrates. Stanford, CA. pp 191–192.
51. Arnold R, Brandmaier S, Kleine F, Tischler P, Heinz E, et al. (2009) Sequence-based prediction of type III secreted proteins. *PLoS Pathog* 5: e1000376.
52. Young BM, Young GM (2002) YplA is exported by the Ysc, Ysa, and flagellar type III secretion systems of *Yersinia enterocolitica*. *J Bacteriol* 184: 1324–1334.
53. Sun Y-H, Rolán HG, Tsalis RM (2007) Injection of flagellin into the host cell cytosol by *Salmonella enterica* serotype Typhimurium. *J Biol Chem* 282: 33897–33901.
54. Badaea L, Beatson S, Kaparakis M, Ferrero R, Hardland E (2009) Secretion of flagellin by the LEE-encoded type III secretion system of enteropathogenic *Escherichia coli*. *BMC Microbiol* 9: 30.
55. Ferreira AO, Myers CR, Gordon JS, Martin GB, Vencato M, et al. (2006) Whole-genome expression profiling defines the HrpL regulon of *Pseudomonas syringae* pv. tomato DC3000, allows *de novo* reconstruction of the Hrp *cis* element, and identifies novel coregulated genes. *Mol Plant-Microbe Interact* 19: 1167–1179.
56. Lan L, Deng X, Zhou J, Tang X (2006) Genome-wide gene expression analysis of *Pseudomonas syringae* pv. tomato DC3000 reveals overlapping and distinct pathways regulated by *hrpL* and *hrpRS*. *Mol Plant-Microbe Interact* 19: 976–987.
57. Löwer M, Schneider G (2009) Prediction of type III secretion signals in genomes of gram-negative bacteria. *PLoS ONE* 4: e5917.
58. Yang Y, Zhao J, Morgan R, Ma W, Jiang T (2010) Computational prediction of type III secreted proteins from gram-negative bacteria. *BMC Bioinformatics* 11: S47.
59. Poueymiro M, Genin S (2009) Secreted proteins from *Ralstonia solanacearum*: a hundred tricks to kill a plant. *Curr Opin Microbiol* 12: 44–52.
60. Mukaihara T, Tamura N (2009) Identification of novel *Ralstonia solanacearum* type III effector proteins through translocation analysis of *hrpB*-regulated gene products. *Microbiology* 155: 2235–2244.
61. Mukaihara T, Tamura N, Iwabuchi M (2010) Genome-wide identification of a large repertoire of *Ralstonia solanacearum* type III effector proteins by a new functional screen. *Mol Plant-Microbe Interact* 23: 251–262.
62. Schmidt KL, Peterson ND, Kustusch RJ, Wissel MC, Graham B, et al. (2004) A predicted ABC transporter, FisEX, is needed for cell division in *Escherichia coli*. *J Bacteriol* 186: 785–793.
63. Goss JW, Sorg JA, Ramamurthi KS, Ton-That H, Schneewind O (2004) The secretion signal of YopN, a regulatory protein of the *Yersinia enterocolitica* type III secretion pathway. *J Bacteriol* 186: 6320–6324.
64. Blaylock B, Sorg JA, Schneewind O (2008) *Yersinia enterocolitica* type III secretion of YopR requires a structure in its mRNA. *Mol Microbiol* 70: 1210–1222.
65. Ramamurthi KS, Schneewind O (2003) *Yersinia yopQ* mRNA encodes a bipartite type III secretion signal in the first 15 codons. *Mol Microbiol* 50: 1189–1198.
66. Cheng LW, Anderson DM, Schneewind O (1997) Two independent type III secretion mechanisms for YopE in *Yersinia enterocolitica*. *Mol Microbiol* 24: 757–765.
67. Amer AA, Åhlund MK, Bröms JE, Forsberg Å, Francis MS (2011) Impact of the N-terminal secretor domain on YopD translocator function in *Yersinia pseudotuberculosis* type III secretion. *J Bacteriol* 193: 6683–6700.
68. Tay D, Govindarajan K, Khan A, Ong T, Samad H, et al. (2010) T3SEdb: data warehousing of virulence effectors secreted by the bacterial type III secretion system. *BMC Bioinformatics* 11: S4.
69. Dawson JE, Seckute J, De S, Schueler SA, Oswald AB, et al. (2009) Elucidation of a pH-folding switch in the *Pseudomonas syringae* effector protein AvrPto. *Proc Natl Acad Sci USA* 106: 8543–8548.
70. Wehling MD, Guo M, Fu ZQ, Alfano JR (2004) The *Pseudomonas syringae* HopPtoV protein is secreted in culture and translocated into plant cells via the type III protein secretion system in a manner dependent on the ShcV type III chaperone. *J Bacteriol* 186: 3621–3630.
71. Ham JH, Bauer DW, Fouts DE, Collmer A (1998) A cloned *Ercwinia chrysanthemi* Hrp (type III protein secretion) system functions in *Escherichia coli* to deliver *Pseudomonas syringae* Avr signals to plant cells and to secrete Avr proteins in culture. *Proc Natl Acad Sci USA* 95: 10206–10211.
72. Srikanth C, Mercado-Lubo R, Hallstrom K, McCormick B (2011) *Salmonella enterica* and host-cell responses. *Cell Mol Life Sci* 68: 3687–3697.
73. Lewis JD, Guttman DS, Desveaux D (2009) The targeting of plant cellular systems by injected type III effector proteins. *Semin Cell Dev Biol* 20: 1055–1063.
74. Lee VT, Schneewind O (2002) Yop fusions to tightly folded protein domains and their effects on *Yersinia enterocolitica* type III secretion. *J Bacteriol* 184: 3740–3745.
75. Sorg JA, Miller NC, Marketon MM, Schneewind O (2005) Rejection of impassable substrates by *Yersinia* type III secretion machines. *J Bacteriol* 187: 7090–7102.
76. King EO, Ward MK, Raney DE (1954) Two simple media for the demonstration of pyocyanin and fluorescein. *J Lab Clin Med* 44: 301–307.
77. Huynh TV, Dahlbeck D, Staskawicz BJ (1989) Bacterial blight of soybean: Regulation of a pathogen gene determining host cultivar specificity. *Science* 245: 1374–1377.
78. Miller JH (1992) A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for *Escherichia coli* and related bacteria. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
79. Sambrook J, Russell D (2001) Molecular Cloning: A Laboratory Manual (Third Edition). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
80. Kvitko BH, Ramos AR, Morello JE, Oh H-S, Collmer A (2007) Identification of harpins in *Pseudomonas syringae* pv. tomato DC3000, which are functionally similar to HrpK1 in promoting translocation of type III secretion system effectors. *J Bacteriol* 189: 8059–8072.
81. Fawcett T (2006) An introduction to ROC analysis. *Pattern Recogn Lett* 27: 861–874.
82. Mason SJ, Graham NE (2002) Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Q J Roy Meteor Soc* 128: 2145–2166.
83. Cuppels DA (1986) Generation and characterization of Tn5 insertion mutations in *Pseudomonas syringae* pv. *tomato*. *Appl Environ Microbiol* 51: 323–327.
84. De Feyter R, Kado CI, Gabriel DW (1990) Small, stable shuttle vectors for use in *Xanthomonas*. *Gene* 88: 65–72.