Novel Type III Effectors in *Pseudomonas aeruginosa*

David Burstein,[a*] Shirley Satanower,[b,c] Michal Simovitch,[b,c] Yana Belnik,[b,c] Meital Zehavi,[b,c] Gal Yerushalmi,[b,c] Shay Ben-Aroya,[b] Tal Pupko,[a] Ehud Banin[b,c]

Department of Cell Research and Immunology, George S. Wise Faculty of Life Sciences, Tel-Aviv University, Tel-Aviv, Israel[a]; The Mina & Everard Goodman Faculty of Life Sciences[b] and The Institute for Nanotechnology and Advanced Materials,[c] Bar-Ilan University, Ramat Gan, Israel

* Present address: David Burstein, Department of Earth and Planetary Science, UC Berkeley, Berkeley, California, USA.

D.B., S.S., and M.S. contributed equally to this work.

ABSTRACT *Pseudomonas aeruginosa* is a Gram-negative, opportunistic pathogen that causes chronic and acute infections in immunocompromised patients. Most *P. aeruginosa* strains encode an active type III secretion system (T3SS), utilized by the bacteria to deliver effector proteins from the bacterial cell directly into the cytoplasm of the host cell. Four T3SS effectors have been discovered and extensively studied in *P. aeruginosa*: ExoT, ExoS, ExoU, and ExoY. This is especially intriguing in light of *P. aeruginosa*'s ability to infect a wide range of hosts. We therefore hypothesized that additional T3SS effectors that have not yet been discovered are encoded in the genome of *P. aeruginosa*. Here, we applied a machine learning classification algorithm to identify novel *P. aeruginosa* effectors. In this approach, various types of data are integrated to differentiate effectors from the rest of the open reading frames of the bacterial genome. Due to the lack of a sufficient learning set of positive effectors, our machine learning algorithm integrated genomic information from another *Pseudomonas* species and utilized dozens of features accounting for various aspects of the effector coding genes and their products. Twelve top-ranking predictions were experimentally tested for T3SS-specific translocation, leading to the discovery of two novel T3SS effectors. We demonstrate that these effectors are not part of the injection structural complex and report initial efforts toward their characterization.

IMPORTANCE *Pseudomonas aeruginosa* uses a type III secretion system (T3SS) to secrete toxic proteins, termed effectors, directly into the cytoplasm of the host cell. The activation of this secretion system is correlated with disease severity and patient death. Compared with many other T3SS-utilizing pathogenic bacteria, *P. aeruginosa* has a fairly limited arsenal of effectors that have been identified. This is in sharp contrast with the wide range of hosts that this bacterium can infect. The discovery of two novel effectors described here is an important step toward better understanding of the virulence and host evasion mechanisms adopted by this versatile pathogen and may provide novel approaches to treat *P. aeruginosa* infections.

*P*seudomonas aeruginosa is an extremely versatile, Gram-negative bacterial pathogen that is a major cause of health care-associated infections, especially among immunocompromised individuals (1). Most *P. aeruginosa* strains express a highly detrimental type III secretion system (T3SS), which is used to deliver toxins, termed effectors, directly into the cytoplasm of infected host cells (2). A clinical study that investigated the involvement of an active T3SS in human infections showed a high correlation between T3SS expression in clinical *P. aeruginosa* strains and patient death (3). In the lab, studies have shown that *P. aeruginosa* T3SS mutants display attenuated virulence toward various hosts (4). Extensive research into the T3SS of *P. aeruginosa* revealed only four effectors: ExoS, ExoU, ExoT, and ExoY (2). Recently, it was reported that nucleoside diphosphate kinase (NDK) is also translocated into host cells in a type III-dependent manner (5). This limited number of effectors is especially intriguing in light of *P. aeruginosa*'s ability to infect a very wide range of hosts (6). Furthermore, it has been shown that a *P. aeruginosa* strain

lacking all known effectors is still virulent during acute lung infection (7). This observation may suggest the involvement of the T3SS injection machinery itself in virulence. However, it may also point to the existence of other, yet-undiscovered, T3SS effector proteins that are encoded in the genome of *P. aeruginosa*.

Recently, we and others have used machine learning classification algorithms to identify effectors of different secretion systems (8–12). These algorithms can integrate various types of data (such as regulatory information, GC content, amino acid content, phylogenomic information) to differentiate effectors from the rest of the open reading frames (ORFs) in the bacterial genome. In our previous study (9), experimental validation proved the ability of such machine learning approaches to accurately identify type IVB effectors of the human intracellular pathogen *Legionella pneumophila*. Here we further developed and applied our machine learning tools to predict novel type III effectors in *P. aeruginosa*. Experimental validation of our predictions has led to the discovery of two novel T3SS effectors. In light of the very small number of
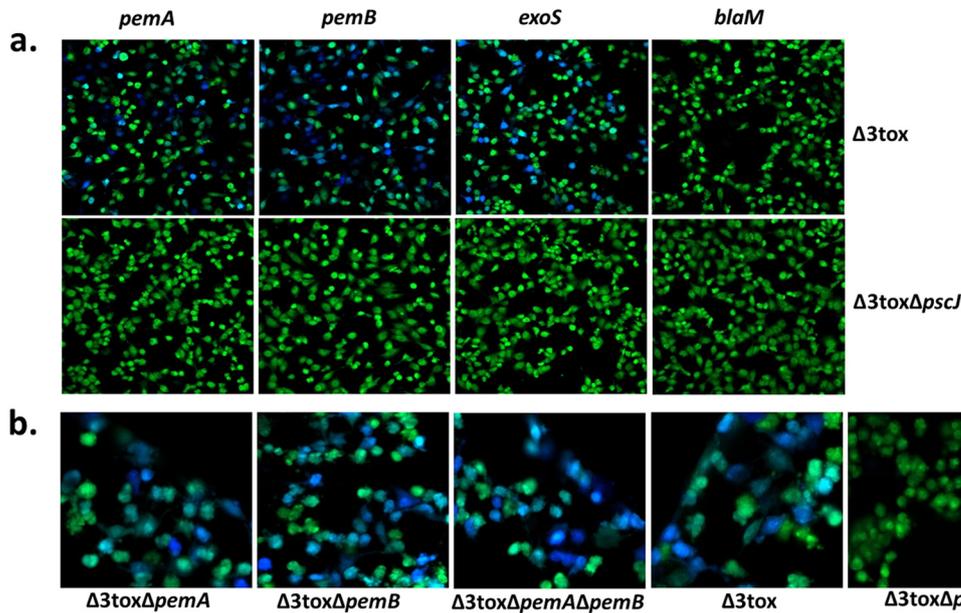
FIG 1 (a) Two new putative effectors are translocated into HeLa cells. (Top row) HeLa cells were infected with Δ3tox strains (active T3SS) or Δ3tox Δ*pscJ* strains (nonactive T3SS) expressing either unfused BlaM, ExoS, or two putative effectors (PemA and PemB) fused to BlaM. Cells stained in blue indicate successful translocation of BlaM fused to the bacterial effector. Green cells indicate no translocation (×10 magnification). (b) The two new effectors are not essential structural components of the T3SS machinery. HeLa cells were infected with a *P. aeruginosa* strain lacking the known effectors (Δ3tox), a strain which has a nonfunctional T3SS (Δ3tox Δ*pscJ*), or strains with deletions of the putative genes *pemA*, *pemB*, or both (Δ3tox Δ*pemA*, Δ3tox Δ*pemB*, or Δ3tox Δ*pemA* Δ*pemB*) expressing ExoY fused to BlaM. Cells stained in blue indicate successful translocation of BlaM fused to the bacterial effector. Green cells indicate no translocation (×20 magnification).

previously characterized effectors, the current increase in the number of known type III effectors may have a direct impact on our understanding of *P. aeruginosa* virulence mechanisms.

In this study, we searched for novel T3SS effectors in *P. aeruginosa*. We applied a set of machine learning classification algorithms to distinguish effectors from non-effectors among all ORFs in this bacterium, based on various features characterizing effectors. As four ORFs (representing the known effectors) are not enough to enable such learning, we augmented the training set with (i) additional *P. aeruginosa* proteins that are secreted via the type III apparatus, including structural proteins and putative secreted regulators, and (ii) effectors from *Pseudomonas syringae*, a phylogenetically related plant pathogen that also harbors a T3SS and has 37 validated effectors in its genome (13). The list of ORFs used as a positive training set is provided in Table S1A in the supplemental material. We trained four different machine learning classifiers, Support Vector Machine (SVM), naive Bayes, Bayesian networks, and random forest, based on a variety of features, such as regulatory information, amino acid properties, GC content, and phylogenomic information. For each ORF, our prediction score is a weighted combination of the scores yielded by the four classifiers in which the weight is assigned according to the estimated performances of the classifiers on the training set (see Fig. S1 and below). The results provide a ranked list of predicted effectors. Top putative effectors are presented in Table S2.

We experimentally tested the 12 top-ranking predictions for translocation into HeLa cells. The genes were fused to β-lactamase (BlaM) and transformed into the *P. aeruginosa* strain that does not express the three known toxins (Δ3tox strain). As a control, and to verify that the putative effectors are secreted by the T3SS, the plasmids were also transformed into the *P. aeruginosa* Δ3tox Δ*pscJ* and

Δ3tox Δ*popB* strains, which have a nonfunctional T3SS (Fig. 1a and Fig. S2A, respectively). The known effectors ExoS and ExoY fused to BlaM were used as positive controls (Fig. 1a and S2A, respectively). Using this approach, we discovered two novel effectors, PA14_16720 and PA14_44480, which are translocated in a T3SS-dependent manner (Fig. 1a and S2A). We term these effectors *pemA* and *pemB* (for *Pseudomonas* effectors discovered by machine learning), respectively. The rest of the putative effectors did not show any translocation into HeLa cells and thus are most likely not T3SS effectors (Fig. S2B). To investigate the possibility that the two novel effectors are structural components of the T3SS, we deleted the *pemA* and *pemB* genes in *P. aeruginosa* (Δ*pemA* Δ*pemB*) and tested the translocation of ExoY from this strain into HeLa cells. We found that neither of the genes is an essential component of the T3SS, as ExoY was efficiently translocated in the absence of the two effectors (Fig. 1b).

A sequence similarity search against reference sequences in the NCBI database (RefSeq) revealed that both *pemA* and *pemB* have conserved orthologs in members of the *Pseudomonas* genus (Fig. S3). We specifically determined the presence of *pemA* and *pemB* orthologs in 13 sequenced strains of *P. aeruginosa* (14). Both effectors were found in all strains. This is similar to the prevalence of *exoT* and *exoY*, which are present in 12 out of the 13 strains (Table S3). It is important to note that the two novel effectors were also located in two strains lacking homologs to essential T3SS components (*P. aeruginosa* PA7 and PA-RP73). The presence of effectors in genomes lacking an active T3SS is not unique for *pemA* and *pemB*; PA-RP73 also encodes *exoT*, *exoS*, and *exoY*. This puzzling observation may reflect an ongoing evolutionary process of T3SS being lost from these genomes; essential T3SS components are already lost, whereas the effector genes, although nonfunc-
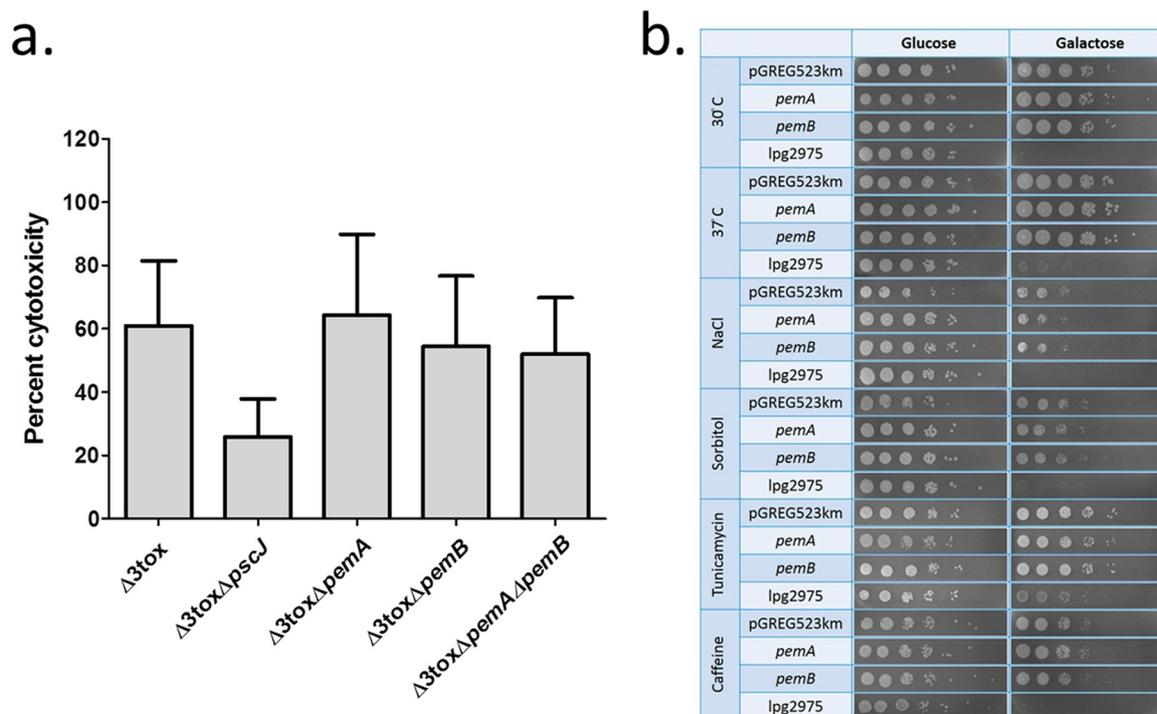
FIG 2 (a) The two new effectors are not cytotoxic to HeLa cells. HeLa cells were infected with wild-type *P. aeruginosa* or the Δ3tox or Δ3tox Δ*pscJ* strain and three strains lacking one (Δ3tox Δ*pemA* or Δ3tox Δ*pemB*) or both (Δ3tox Δ*pemA* Δ*pemB*) of the newly discovered effectors. Cytotoxicity was assessed by the CytoTox-ONE homogeneous membrane integrity assay after 4 h of infection. Three biological repeats were performed in duplicate. Expression of the effectors under these conditions was validated (see Fig. S5A in the supplemental material). None of the strains showed a significant change in cytotoxicity according to the one-way analysis of variance (ANOVA) test. (b) Expression of the two new effectors does not affect the growth of *S. cerevisiae*. *pemA* and *pemB* were expressed from the GAL10 promoter in *S. cerevisiae* BY4741. Growth was assessed by a yeast spotting assay. One molar sorbitol, 7 mM caffeine, 0.5 M NaCl, 0.12 μg/ml tunicamycin, and growth at 37°C were used to assess the growth effects of the effectors under various stress conditions. *S. cerevisiae* expressing the *Legionella* toxic effector lpg2975 and the empty vector pGREG523-Km were used as controls. Three biological repeats were performed. Effector expression was validated by Western blotting (Fig. S6B).

tional, still reside in the genome. Alternatively, these effectors may have additional functions not related to the secretion system.

Notably, none of the *pemA* or *pemB* orthologs that we found were annotated. No distant homologs of these genes were found following PSI-BLAST (15) and HHpred (16) searches, suggesting that these are *Pseudomonas*-specific genes. Inspection of the genomic context of the newly discovered effectors suggested that *pemA* is part of an operon together with PA14_16730, a hypothetical protein whose gene has a start codon 18 bp downstream of *pemA*. This operon structure is conserved across various *Pseudomonas* species (including *P. aeruginosa*, *P. syringae*, and *P. fluorescens*). The *pemB* gene resides between genes encoding coproporphyrinogen III oxidase (*hemN*, part of the heme pathway) and *anr*, encoding an anaerobic regulatory protein. It is in an orientation opposite to that of these genes and probably not related to them in function. We identified in PemB a conserved coiled-coil domain (located at amino acids 78 to 140 [Fig. S3]) and an ExsA binding site in its regulatory region. ExsA is the principal regulator of *P. aeruginosa* type III secretion gene expression. The ExsA footprint that we identified is highly similar to that of the ExsA binding site in the known T3SS-associated genes *popN*, *exoT*, *exoU*, and *exoY* (Fig. S4). The presence of a coiled-coil domain and the ExsA binding motif further suggests the involvement of this effector in T3SS-associated activity (2, 17).

All previously known effectors of *P. aeruginosa*, ExoS, ExoT, ExoU, and ExoY, elicit a toxic response in *Saccharomyces cerevisiae*

(7, 18–20), and ExoS and ExoU are also toxic to cultured human cells (5, 7). We hence tested whether *pemA* and *pemB* affect host cell viability in a human cell line and in *S. cerevisiae*. We infected HeLa cells with *P. aeruginosa* Δ3tox lacking either *pemA*, *pemB*, or both and assessed host cell viability using a homogeneous membrane integrity assay. Neither gene was cytotoxic to HeLa cells (Fig. 2a). Cytotoxicity toward HeLa cells was also not affected when *pemA* and *pemB* were ectopically overexpressed (Fig. S6). In addition, the expression of the two newly identified effector genes in *S. cerevisiae* did not affect host cell viability (Fig. 2b). To investigate the involvement of certain stress pathways in *S. cerevisiae*'s response to the two bacterial effectors, we conducted growth assays in the presence of various known stress agents, including sorbitol, NaCl, caffeine, tunicamycin, and a temperature of 37°C (21). The results, presented in Fig. 2b, show that the presence of stress agents also did not elicit a toxic response in *S. cerevisiae* expressing the *pemA* and *pemB* genes. The lack of a toxicity phenotype in our tests may indicate that the toxicity of the two genes is restricted to a small group of specific host cell types or triggered by yet-undefined signals. Alternatively, these effectors may not have a toxic effect; rather, they may interact with the host's innate immunity or regulate specific host pathways for the benefit of the bacteria. Clearly, our findings open the way to further research aimed at deciphering the functions and mechanisms of these newly discovered effectors.

Genome sequences. The genomic sequence of *P. aeruginosa* PA14 (NCBI accession number NC_008463) and the sequences of the *P. syringae* DC3000 chromosome and plasmids (NCBI accession numbers NC_004578, NC_004632, and NC_004633) were downloaded from the NCBI genome FTP site.

Classifier training sets. The classification of ORFs as either effectors or non-effectors requires a training set that serves to identify properties that characterize effectors and differentiate them from the rest of the ORFs encoded by the bacteria. In *P. aeruginosa* PA14, only three effectors have been identified to date. Such a limited set of effectors is insufficient for reliable classification. Therefore, we added to the training set additional *P. aeruginosa* proteins that are secreted via the T3SS, as well as known effectors of *P. syringae*, a closely related species of *Pseudomonas* that also utilizes a T3SS to translocate effectors. In total, the effector training set consisted of 50 effectors, which are specified in Table S1A in the supplemental material. For the noneffector set, we used all the ORFs of *P. aeruginosa* and *P. syringae* that were not part of the effector set. This ensures that no bias is introduced while choosing the negative set. Few of the ORFs in the negative set are expected to be effectors; hence, the performance estimations are conservative (a lower bound of the actual performances). An ORF is considered a putative effector if the classification scheme challenges its assignment to the non-effector set, predicting that it has a high likelihood of being part of the effector set. Highly similar ORFs used in the training set might introduce biases into the classification and estimation procedures. As we used ORFs from two closely related bacteria, the training set is expected to contain a high number of orthologous proteins in addition to the paralogs existing in each of the species. To reduce the redundancy, we filtered out ORFs with a BLAST E-value lower than 0.01 in favor of an ORF already in the training set (when applicable, we filtered out *P. syringae* rather than *P. aeruginosa* ORFs). Classifiers use different attributes or features that were measured for each instance of the training set to perform the classification. Various features were measured for each ORF in the data set. Table S1B specifies the full list of 166 features used in this study and the importance score of each feature, as measured by the random forest classifier.

Classification scheme. The classification scheme is an extension of our previously developed methodology described in detail in reference 9. The same classification algorithms were used, except for the multilayer perceptron classifier, which was replaced with the random forest classifier (22), which consistently displayed superior performance. The classification performance of each classifier was calculated by a mean area under the precision-recall curve (AUPRC) over 10-fold cross-validation. It should be noted that the data serving as a test set were excluded from the feature selection stage to ensure that the evaluation process was performed with unseen data. The final prediction scores were achieved by computing a weighted mean of the results with the four classifiers, with their performances as weights. The AUPRC values were calculated with AUCCalculator 0.2 (23). For random forests, we used the R implementation (24), and the rest of the machine learning analysis was implemented in Java by using the open-source package Weka, version 3.7.0 (25).

Strains, plasmids, and growth medium. The list of the bacterial/yeast strains and plasmids used in this study are listed in Table S4 in the supplemental material. For infection, overnight cultures of *P. aeruginosa* were grown in MINS medium (25 mM

KH$_2$PO$_4$, 95 mM NH$_4$Cl, 50 mM monosodium glutamate, 110 mM disodium succinate, 10 mM trisodium nitrilotriacetic acid, 2.5% glycerol, 5 mM MgSO$_4$, 18 μM FeSO$_4$) at 37°C overnight, diluted the next morning, and grown to exponential phase. HeLa cells were cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal calf serum (FCS). Where indicated, modified Casamino Acids-DMEM (cDMEM) [0.25 μM Fe(NO$_3$)$_3$, 1.4 mM CaCl$_2$, 5.4 mM KCl, 0.8 mM MgSO$_4$, 110 mM NaCl, 1 mM Na$_2$HPO$_4$, 44 mM NaHCO$_3$, 0.45% glucose, 0.1 M HEPES, 0.1% Casamino Acids] was used instead of DMEM. *S. cerevisiae* strains were grown at 30°C, unless specified otherwise. Yeast-peptone-dextrose (YPD) medium was used for nonselective growth. Synthetic defined medium-His (SD-His) supplemented with 2% glucose or 2% galactose was used for selective growth. The following stressors were used for growth experiments: 1 M sorbitol, 7 mM caffeine, 0.5 M NaCl, 0.12 μg/ml tunicamycin, and a temperature of 37°C.

Construction of *P. aeruginosa* strains. To construct *P. aeruginosa* strains expressing the putative effectors fused to BlaM, the putative effector genes were PCR amplified without the stop codon (primer pairs are listed in Table S4B), and the products were digested with the XbaI and EcoRI restriction enzymes. The cut products were ligated into pMS50 and fused to the *blaM* gene. Ligation products were electroporated into DH5α and grown on LB containing 100 μg/ml gentamicin for selection. The sequences of the plasmids from positive transformants were validated, and the plasmids were electroporated into the PAO1f Δ3tox and PAO1f Δ3tox Δ*pscJ* strains. BlaM expression was verified by Western blotting using anti-BlaM. pMS50 was constructed by inserting the *blaM* gene from pCX341 into pJN105 between the XbaI and SacI restriction sites. For construction of deletion mutants of *pemA* and *pemB*, overlap extension PCR using the primers specified in Table S4B was used in order to generate a fragment containing the upstream and downstream regions of each gene. By Gateway cloning, each fragment was cloned into the allelic-exchange vector DB3.1 pEX18GmGW with BP-Clonase (Invitrogen). Each deletion was introduced into PA14 by using biparental mating. Deletions were generated by a standard method for two-step allelic exchange and were confirmed by PCR and Southern blotting. To construct *P. aeruginosa* strains expressing the putative effectors fused to FLAG (for overexpression HeLa toxicity assays), the putative effector genes were PCR amplified (primer pairs are listed in Table S4B) and the products were digested with the XbaI and EcoRI restriction enzymes. The *exoU* gene was PCR amplified, and the product was digested with the NheI and SacI restriction enzymes. The cut product was ligated into pJN105 and fused to a FLAG tag on the N terminus. Ligation products were electroporated into DH5α and grown on LB containing 100 μg/ml gentamicin for selection. The sequences of the plasmids from positive transformants were validated, and the plasmids were electroporated into the relevant parental strain. FLAG expression was verified by Western blotting using anti-FLAG.

Construction of *S. cerevisiae* strains. To construct *S. cerevisiae* strains expressing the *pemA*, *pemB*, and lpg2975 genes from the pGREG523-Km expression vector, the genes were PCR amplified, digested, and ligated into pUC18. Ligation products were electroporated into DH5α cells and grown on LB containing 100 μg/ml ampicillin for selection. The sequences of the plasmids from the positive transformants were validated. The genes were then cut from pUC18 and transferred to pGREG523-Km between EcoRI

and BamHI restriction sites by using a standard restriction-ligation protocol. The vectors were then transferred to BY4741 by a standard yeast transformation protocol. Protein expression was verified with anti-myc.

Translocation assay. A translocation assay was conducted as described in reference 26 with the following exceptions. HeLa cells were seeded at 10^6 cells/well in a 96-well plate. cDMEM was used instead of Hanks' balanced salt solution. Cells were viewed and photographed on a Zeiss confocal microscope in real time.

Cytotoxicity assay. HeLa cells were seeded at 10^6 cells/well in a 96-well plate. HeLa cells were infected at an MOI of 100 for 4 h and washed. Following infection, the wells were washed of bacteria with cDMEM and cytotoxicity was quantified by Promega's CytoTox-ONE homogeneous membrane integrity assay according to the manufacturer's instructions. In the *S. cerevisiae* spotting assay, *S. cerevisiae* BY4741 strains ectopically expressing *pemA* (pSS1586) and *pemB* (pYB1704) were grown overnight on selective medium supplemented with 2% glucose. BY4741 harboring the empty vector pGREG523-Km was used as a control. BY4741 expressing the *Legionella* effector protein lpg2975 (27) was used as a positive control. The next day, cultures were diluted 1:10 seven times. Five microliters from each dilution was pipetted on SD-His plates and SC-His 2% galactose plates. Plates were incubated at 30°C or 37°C for 48 h.

Western blot analysis. Protein extracts were separated by SDS-PAGE and transferred onto nitrocellulose membranes. The membranes were then probed with antibodies against FLAG or MYC. Proteins of interest were detected with horseradish peroxidase (HRP)-conjugated anti-mouse IgG antibody (for FLAG) or HRP-conjugated anti-rabbit IgG antibody (for MYC) and visualized by enhanced chemiluminescence in a Western blotting substrate.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00161-15/-/DCSupplemental.

Figure S1, DOCX file, 0.3 MB.
Figure S2, DOCX file, 0.6 MB.
Figure S3, PDF file, 0.8 MB.
Figure S4, DOCX file, 0.1 MB.
Figure S5, DOCX file, 1.1 MB.
Figure S6, DOCX file, 0.03 MB.
Table S1, DOCX file, 0.04 MB.
Table S2, DOCX file, 0.01 MB.
Table S3, DOCX file, 0.02 MB.
Table S4, DOCX file, 0.03 MB.

REFERENCES

1. **Coggan KA, Wolfgang MC.** 2012. Global regulatory pathways and cross-talk control *Pseudomonas aeruginosa* environmental lifestyle and virulence phenotype. Curr Issues Mol Biol **14:**47–70.
2. **Hauser AR.** 2009. The type III secretion system of *Pseudomonas aeruginosa*: infection by injection. Nat Rev Microbiol **7:**654–665. http://dx.doi.org/10.1038/nrmicro2199.
3. **Roy-Burman A, Savel RH, Racine S, Swanson BL, Revadigar NS, Fujimoto J, Sawa T, Frank DW, Wiener-Kronish JP.** 2001. Type III protein secretion is associated with death in lower respiratory and systemic *Pseudomonas aeruginosa* infections. J Infect Dis **183:**1767–1774. http://dx.doi.org/10.1086/320737.
4. **Miyata S, Casey M, Frank DW, Ausubel FM, Drenkard E.** 2003. Use of the *Galleria mellonella* caterpillar as a model host to study the role of the type III secretion system in *Pseudomonas aeruginosa* pathogenesis. Infect Immun **71:**2404–2413. http://dx.doi.org/10.1128/IAI.71.5.2404-2413.2003.
5. **Neeld D, Jin Y, Bichsel C, Jia J, Guo J, Bai F, Wu W, Ha UH, Terada N, Jin S.** 2014. *Pseudomonas aeruginosa* injects NDK into host cells through type III secretion system. Microbiology **160:**1417–1426. http://dx.doi.org/10.1099/mic.0.078139-0.
6. **Ramos J, Filloux A.** 2004. Pseudomonas. Kluwer Academic/Plenum Publishers, New York, NY.
7. **Galle M, Jin S, Bogaert P, Haegman M, Vandenabeele P, Beyaert R.** 2012. The *Pseudomonas aeruginosa* type III secretion system has an exotoxin S/T/Y independent pathogenic role during acute lung infection. PLoS One **7:**e41547. http://dx.doi.org/10.1371/journal.pone.0041547.
8. **Arnold R, Brandmaier S, Kleine F, Tischler P, Heinz E, Behrens S, Niinikoski A, Mewes HW, Horn M, Rattei T.** 2009. Sequence-based prediction of type III secreted proteins. PLoS Pathog **5:**e1000376. http://dx.doi.org/10.1371/journal.ppat.1000376.
9. **Burstein D, Zusman T, Degtyar E, Viner R, Segal G, Pupko T.** 2009. Genome-scale identification of *Legionella pneumophila* effectors using a machine learning approach. PLoS Pathog **5:**e1000508. http://dx.doi.org/10.1371/journal.ppat.1000508.
10. **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. http://dx.doi.org/10.1371/journal.ppat.1000375.
11. **Yang Y, Zhao J, Morgan RL, Ma W, Jiang T.** 2010. Computational prediction of type III secreted proteins from gram-negative bacteria. BMC Bioinformatics **11**(Suppl 1):S47. http://dx.doi.org/10.1186/1471-2105-11-S1-S47.
12. **McDermott JE, Corrigan A, Peterson E, Oehmen C, Niemann G, Cambronne ED, Sharp D, Adkins JN, Samudrala R, Heffron F.** 2011. Computational prediction of type III and IV secreted effectors in Gram-negative bacteria. Infect Immun **79:**23–32. http://dx.doi.org/10.1128/IAI.00537-10.
13. **Chang JH, Urbach JM, Law TF, Arnold LW, Hu A, Gombar S, Grant SR, Ausubel FM, Dangl JL.** 2005. A high-throughput, near-saturating screen for type III effector genes from *Pseudomonas syringae*. Proc Natl Acad Sci U S A **102:**2549–2554. http://dx.doi.org/10.1073/pnas.0409660102.
14. **Winsor GL, Lam DK, Fleming L, Lo R, Whiteside MD, Yu NY, Hancock RE, Brinkman FS.** 2011. Pseudomonas Genome Database: improved comparative analysis and population genomics capability for *Pseudomonas* genomes. Nucleic Acids Res **39:**D596–D600. http://dx.doi.org/10.1093/nar/gkq869.
15. **Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.** 1997. Gapped BLAST and psi-blast: a new generation of protein database search programs. Nucleic Acids Res **25:**3389–3402. http://dx.doi.org/10.1093/nar/25.17.3389.
16. **Remmert M, Biegert A, Hauser A, Söding J.** 2012. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Methods **9:**173–175. http://dx.doi.org/10.1038/nmeth.1818.
17. **Delahay RM, Frankel G.** 2002. Coiled-coil proteins associated with type III secretion systems: a versatile domain revisited. Mol Microbiol **45:**905–916. http://dx.doi.org/10.1046/j.1365-2958.2002.03083.x.
18. **Arnoldo A, Curak J, Kittanakom S, Chevelev I, Lee VT, Sahebol-Amri M, Koscik B, Ljuma L, Roy PJ, Bedalov A, Giaever G, Nislow C, Merrill AR, Lory S, Stagljar I.** 2008. Identification of small molecule inhibitors of *Pseudomonas aeruginosa* exoenzyme S using a yeast phenotypic screen. PLoS Genet **4:**e1000005. http://dx.doi.org/10.1371/journal.pgen.1000005.
19. **Rabin SD, Hauser AR.** 2003. *Pseudomonas aeruginosa* ExoU, a toxin transported by the type III secretion system, kills *Saccharomyces cerevisiae*. Infect Immun **71:**4144–4150. http://dx.doi.org/10.1128/IAI.71.7.4144-4150.2003.
20. **Garrity-Ryan L, Shafikhani S, Balachandran P, Nguyen L, Oza J, Jakobsen T, Sargent J, Fang X, Cordwell S, Matthay MA, Engel JN.** 2004. The ADP ribosyltransferase domain of *Pseudomonas aeruginosa* ExoT contrib-

utes to its biological activities. Infect Immun **72:**546–558. http://dx.doi.org/10.1128/IAI.72.1.546-558.2004.

21. **Salomon D, Bosis E, Dar D, Nachman I, Sessa G.** 2012. Expression of *Pseudomonas syringae* type III effectors in yeast under stress conditions reveals that HopX1 attenuates activation of the high osmolarity glycerol MAP kinase pathway. Microbiology **158:**2859–2869. http://dx.doi.org/10.1099/mic.0.062513-0.

22. **Breiman L.** 2001. Random forests. Mach Learn **45:**5–32. http://dx.doi.org/10.1023/A:1010933404324.

23. **Jesse D, Mark G.** 2006. The relationship between precision-recall and ROC curves, p 233–240. Proceedings of the 23rd International Conference on Machine Learning. ACM, Pittsburgh, PA.

24. **Liaw A, Wiener M.** 2002. Classification and regression by randomForest. R News **2:**18–22.

25. **Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH.** 2009. The Weka data mining software: and update. SIGKDD Explor Newsl **11:**10–18. http://dx.doi.org/10.1145/1656274.1656278.

26. **Diaz MH, Hauser AR.** 2010. *Pseudomonas aeruginosa* cytotoxin ExoU is injected into phagocytic cells during acute pneumonia. Infect Immun **78:**1447–1456. http://dx.doi.org/10.1128/IAI.01134-09.

27. **Nevo O, Zusman T, Rasis M, Lifshitz Z, Segal G.** 2014. Identification of *Legionella pneumophila* effectors regulated by the LetAS-RsmYZ-CsrA regulatory cascade, many of which modulate vesicular trafficking. J Bacteriol **196:**681–692. http://dx.doi.org/10.1128/JB.01175-13.