DOI: 10.3969/cjz.j.issn.1002－2694.2012.06.003

EFFECTORSEARCH: Software for identifying effectors of T3SS in bacterial species

ZHANG Wen[1], Joy BERGELSON[2]

(1. *National Institute for Communicable Disease Control and Prevention, Chinese and Prevention/State Key Laboratory for Infectious Disease Prevention and Control, Beijing 102206, China*;
2. *Department of Ecology and Evolution, University of Chicago, Chicago 60637, USA*)

ABSTRACT: Many gram-negative bacteria use the type III secretion system (T3SS) to inject virulence proteins (effectors) directly into host cells. These effectors play a central role in pathogen interactions, and their molecular structures are highly variable. Previous researchers have used particular motifs common to all effectors in order to identify a full suite of candidate effectors of T3SS in the genome of *Pseudomonas syringae*. Here, we present a program called EFFECTORSEARCH that synthesizes and extends previous work by allowing users to identify candidate effectors based on any combination of the following five criteria: protein length, proximity to the hrp promoter, n-terminal region, similarity to known effectors, and dissimilarity to housekeeping genes. We demonstrate restrict attention to the candidates with the highest confidence in their assignment.

KEY WORDS: EFFECTORSEARCH; bacterial species; T3SS; prediction

This work was supported by NSF MCB-0603515 to JB and China Scholarship Council Fellowship to WZ.

Corresponding author: Joy BERGELSON. Email: jbergels@uchicago.edu

Availability: Software with manual and tutorials are freely available at http://bergelson.uchicago.edu/software

Introduction

In bacterial species of *Shigella*, *Salmonella*, *Escherichia coli*, *Burkholderia*, *Yersinia*, *Chlamydia*, and *Pseudomonas*, effectors are translocated from pathogen to host cells through a needle-like structure named type III secretion system (T3SS)[1-3], which are essential for the virulence of these gram-negative bacteria[4]. Effectors enhance pathogenicity in hosts that do not recognize them[5-6], but in hosts that do, they elicit a resistance response, the hypersensitive response (HR), that inhibits pathogen growth[7]. With the rapid increase in availability of sequenced bacterial genomes, there is tremendous opportunity to apply computational methods to study function and evolution T3SS effector, providing that effective tools for their identification.

Previous efforts to identify candidate effectors of T3SS through computational methods have focused on single motifs common among known effectors[3,8-9]. Recent work[10] has identified 93 candidate effectors in the genome of the DC3000 strain of *Pseudomonas syringae* by searching for identified open reading frames in the region of 250 bp downstream of the hypersensitive response and pathogenicity (hrp) promoter. Other work used the N-terminal region of proteins to identify 57 candidate effectors[11-12]. Interestingly, there is little overlap between candidates identified by using these 2 methods, although in each case empirical test confirmed that the method could effectively identify new effectors. Recently, two novel computational approaches were developed to detect effectors in all bacteria pathogens by analyzing the amino acid composition, G+C content and N-terminal region of the protein sequence[13-14].

In this study, we combine and extend the methods above by developing a new computational tool that is able to identify T3SS effectors quickly and effectively. Five criteria are combined in a single framework, and the user is able to weigh each criterion to enhance selectivity of the set of candidate effectors. With, EFFECTORSEARCH program, we could reproduce previous results, search broadly for candidates, and refine sets of candidates to identify true effectors of T3SS in pathogen bacterial genome, i. e. those with empirical valida-

美国国家科学基金(No. MCB 0603515)，中国国家留学基金联合资助
通讯作者：Joy BERGELSON, Email: jbergels@uchicago.edu
作者单位：1. 中国疾病预防控制中心传染病预防控制所，北京 102206；
2. 美国芝加哥大学生态与进化学系，芝加哥 60637

tion of their function.

Methods

EFFECTORSEARCH is a software package written with PERL scripts that can be run in a Linux/Unix system (Figure 1). The five criteria used to identify effectors in a target genome sequence are listed below.

Criterion 1: Protein length. The default minimum length of candidate effectors is set at 100 amino acids because previous reports [11-12,15-16] have shown that more than 96% of proven T3SS effectors contain more than 100 codons.

Criterion 2: Proximity to hrp promoter. The default setting selects candidates if the promoter and protein are located on the same contig and within 5kb of each other, and if the hrp promoter has a hidden Markov model score greater than 10.0 (see Supplemental Material) [17].

Criterion 3: N-terminal region. The default setting requires candidates with the N-terminal region matching all 3 characters in the consensus pattern for effectors (see Supplemental Material) [4,11,18-19].

Criterion 4: Similarity to known effectors. Sequence of each candidate was blasted by using blastp against functionally confirmed effectors from DC3000, B728a, 1448A and other 10 species with a T3SS system to determine if the candidate was a member of an established Effector family. If >50%, a candidate sequence could be aligned ($e<10^{-5}$) with one known effector sequence, then the candidate was considered to meet the criterion. All functionally confirmed effectors and genomes we used are listed in *blast_data* file of *EFFECTORSEARCH* package and in Supplemental Table 1.

Criterion 5: Filtering out candidate housekeeping genes. Housekeeping genes, in contrast to effectors, are likely to be shared and conserved by more than one species [20]. We used blastp to filter out candidate housekeeping genes which have homologues in 6 related *Pseudomonas* species (Supplemental Table 1) [21]. As for the previous criterion, the e-value cutoff was set to 10^{-5} and the identity was set to be > 50. All genes used in this criterion are listed in the *blast_data* file of the EFFECTORSEARCH package.

To run EFFECTORSEARCH, the user must define the criteria and their relative weights, as well as the two additional parameters that describe the target genome sequence. The first of these additional parameters, termed "Genome", should be stated "yes" if a fully sequenced genome is being provided and "no" otherwise. The second parameter is termed "Codon Table". In EFFECTORSEARCH, there are 24 tables of genetic codes by which genes are translated into proteins; the first 23 of these were downloaded from NCBI and the last one is a user-defined table. The user is able to select which table will be used to define the genetic code. The parameter Score (Score = $\sum$ (weight * number)) represents the sum of the

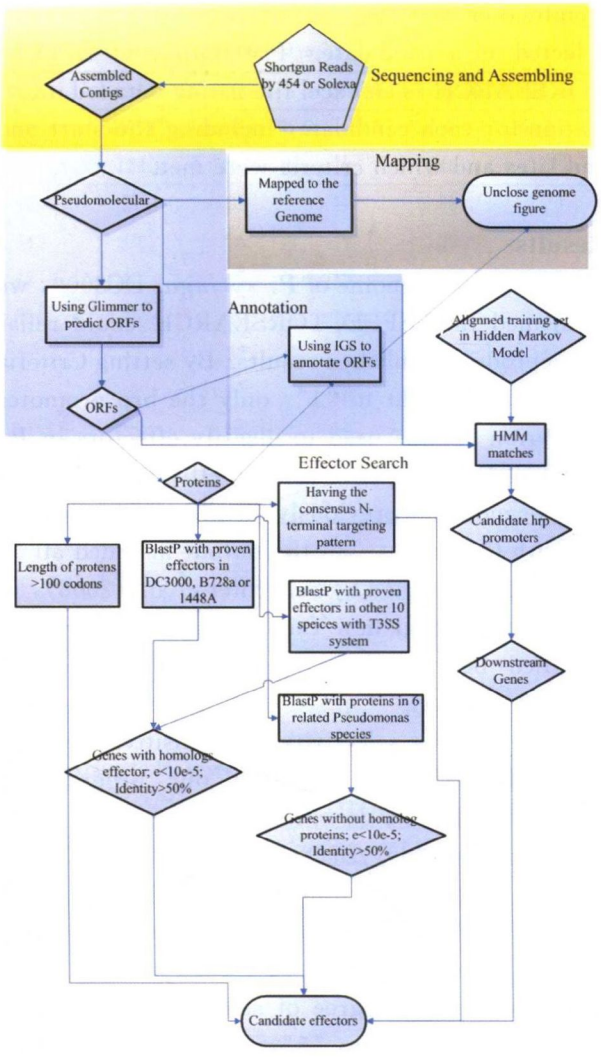

Fig. 1　Workflow for the process of sequencing, annotation and searching for candidate effectors. Major processing modules are grouped and highlighted by color. The nodes representing primary inputs, outputs, processing steps, and intermediate datasets are denoted with pentagons, ellipses, squares, and diamonds, respectively.

weighted criteria that must be met for a gene to be selected as a candidate. The output of EFFECTORSEARCH is an excel file listing detailed information for each candidate, including the start and end sites and which criteria were met.

Results

Using the genome of *P. syringae* DC3000, we confirmed that EFFECTORSEARCH could reliably reproduce published results. By setting Criteria to "2" and Weight to "1", only the hrp promoter (Criterion 2) was used to identify effectors in *P. syringae* DC3000. Our selected effectors contained the 93 candidates previously identified [10]. Similarly, with Criteria = "3" the output included all 56 candidates identified by Schechter et al. (2006), as well as additional new candidates, which were identified because we used the fully sequenced DC3000 genome.

To test the selectivity and sensitivity of EFFECTORSEARCH, we used our program to search through the genomes of three strains of *P. syringae* (DC3000, B728a and 1448A; sequences were all downloaded from NCBI) while considering all five criteria weighted equally. Thus, Criteria = "1 2 3 4 5" and Weight ="1 1 1 1 1".

The highest degree of selectivity is achieved when Score was set to "5", meaning that all five criteria should be met for a protein to be selected. Setting Score = 5, we identified 18 candidates in the genome of *P. syringae* DC3000. As expected, a larger pool of candidates was identified when selectivity was reduced (Score = 4 identifies 46 candidates, Score = 3 identifies 229 effectors). To confirm that higher values for Score have greater opportunity to identify true candidates, we calculated the fraction of candidates for which empirical validation of T3SS-dependent secretion or translocation had been published [4,12,18-19,22-26]. For Score = 5, 94.4% (17 / 18) effectors had been experimentally validated, for Score = 4, 82.6% had been experimentally validated and for Score = 3, only 20.1% had been previously validated (Figure 2a). Of course, many effectors were not tested for their secretion or translocation, and thus the percentage of true candidates will be underestimated.

Similar results were also found for two other strains of *P. syringae* (Figure 2b). When Score = 5, there were 11 candidates identified in B728a and 13 candidates in 1448A and most of these were previously shown to be true candidates (81.8% in B728a and 76.9% in 1448A) [24,27-28].

All effectors selected when setting Score = 4 were listed in Supplemental Table 2; this table provides a rich resource for future experimental research.

As shown for DC3000, longer lists of candidates were found when conditions were relaxed. When Score = 4, there were 46, 21 and 33 candidates identified in DC3000, B728a and 1448A, respectively (Table 1&2&3). While Score=3, the numbers increased to 229, 140 and 177 (Figure 2 and Figure 3). The sets of candidates identified with Score = 3 included more than 90% of candidates found in most previous papers [4,10-12,16], and 74.5% of effectors identified by Ferreira et al. (2006) (Figure 3). Only two functionally confirmed effectors (PSPTO_4590&PSPPH_5225) were not identified by EFFECTORSEARCH when setting Score = 3. Comparing with other two computational approaches (SIEVE and EffectiveT3)[13-14], EFFECTORSEARCH with Score=3 had the highest selectivity (Table 1&2&3) in three strains of *P. syringae*.

Discussion

The T3SS is one of important secretion systems necessary for the virulence of gram-negative pathogens; however the prediction of T3SS effectors is complex. Given the substantial effort needed to confirm the function of putative effectors, a computational tool that helps to prioritize empirical efforts could be quite useful. EFFECTORSEARCH works as such a computational tool to identify effectors with higher or lower selectivity based on user-define parameters. The verification on results of EFFECTORSEARCH has been proven in *P. syringae* and it is with higher selectivity than other previous programs. It then provides lists of candidates and information about which criteria effectors meet to facilitate empirical validation.

Acknowledgements

We thank Luke Barrett and Matt Horton, and Elizabeth B. Haney for contributions to edit this manuscript.

Tab. 1　Effectors identified in DC3000 of *P. syringae* by several ways

Shared Effectors[1]		F[2]	S[3]	L[4]	G[5]	C[6]	Computational Approach		EFFECTORSEARCH		
							SIEVE	EffectiveT3	Score=5	Score=4	Score=3
Num[7]		93	56	59	35	51	124	236	18	46	229
F	93	—	40	43	30	42	31	31	17	38	70
S	56	—	—	53	34	44	40	36	18	39	54
L	59	—	—	—	35	47	40	35	18	40	55
G	35	—	—	—	—	35	26	26	14	28	34
C	51	—	—	—	—	—	32	32	17	36	49
SIEVE	124	—	—	—	—	—	—	47	16	33	46
EffectiveT3	236	—	—	—	—	—	—	—	16	30	47
Proven effectors[8]	48	38	46	48	30	41	35	31	17	36	46
Selectivity[9]	79.2%	95.8%	100.0%	62.5%	85.4%	72.9%	64.6%	35.4%	75.0%	95.8%	

1: Shared Effectors represent number of effectors which were identified by both two ways.

2—6: F represents effectors found by Ferreira *et al*. (2006); S. Schechter *et al*. (2006); L. Lindeberg *et al*. (2006); G. Guttman *et al*. (2002); C. Collmer *et al*. (2002).

7: Num is the total number of effectors found by each approach.

8: Proven effectors represent effectors with empirical validation of T3SS-dependent secretion or translocation.

9: Selectivity in this table is the percentage of proven effectors.

Tab. 2　Effectors identified in B728a of *P. syringae* by several ways

Shared Effectors		L	Computational Approach		EFFECTORSEARCH		
			SIEVE	EffectiveT3	Score=5	Score=4	Score=3
Num		27	90	229	11	21	140
L	27	—	18	17	10	19	24
SIEVE	90	—	—	37	10	16	22
EffectiveT3	229	—	—	—	7	15	25
Proven Effectors	18	18	14	15	9	18	18
Selectivity		100.0%	77.8%	83.3%	50.0%	100.0%	100.0%

Tab. 3　Effectors identified in 1448A of *P. syringae* by several ways

Shared Effectors		L	M	Computational Approach		EFFECTORSEARCH		
				SIEVE	EffectiveT3	Score=5	Score=4	Score=3
Num		36	36	62	230	13	33	177
L	36	—	30	32	21	12	26	32
M	36	—	—	27	20	12	25	33
SIEVE	62	—	—	—	28	13	27	31
EffectiveT3	230	—	—	—	—	10	21	28
Proven Effectors	23	23	22	20	15	10	18	22
Selectivity		100.0%	95.7%	87.0%	65.2%	43.5%	78.3%	95.7%

Supplemental Material

Supplemental Text

More detailed information about criteria the software used is provided below.

Criterion 2: Proximity to hrp promoter. The hypersensitive response and pathogenicity (Hrp) promoter box, which is responsive to the HrpL factor and facilitates the injection of effectors into plant cells, is characterized by a functional and conserved region upstream of candidate effectors (Fouts, et al., 2002). To search for Hrp promoters in the target genomes, we first built a calibrated hidden Markov model (HMM) based on a whole-genome nucleotide monomer-based statistical background model and a combined, aligned lev-

el 2 training set (Eddy, 1998; Ferreira, et al., 2006). The background model was created by computing single nucleotide frequencies using the program compseq from the EMBOSS suite (Rice, et al., 2000). The default parameters for building this HMM model were --num 50000 --mean 500 --sd 100. Resulting HMM matches were scanned through the whole target genome to identify candidate hrp promoters with scores greater than 10.0. Two Perl scripts, *promoter_find.pl* and *promoter.pl*, in *EFFECTORSEARCH* were used to list all candidate promoters and their downstream effectors.

Criterion 3: *N-terminal region* As a result of multiple studies (Greenberg and Vinatzer, 2003; Guttman, et al., 2002; Lindeberg, et al., 2005; Petnicki-Ocwieja, et al., 2002), it is known that there is a consensus T3SS targeting pattern in the N-terminal region of effectors. This consensus pattern has the following characters: (1) >=10% Ser content in the NH2-terminal 50 amino acid region; (2) Ile, Leu, Val or Pro in the third or fourth position; (3) no Asp or Glu residues in the first 12 amino acids. By searching through all predicted ORF sequences of the whole target genome for this consensus pattern, we were able to identify candidate effectors. The default setting of EFFECTORSEARCH requires sequences identified as candidates to have the N-terminal regions matching all these three characters

Supplemental Tables

Supplemental Tab. 1: List of source species for genes used in *EFFECTORSEARCH*

Species used in Criteria 4	Kingdom	Host	TTSS	Number of Effector	Reference
Pseudomonas syringae pv. *tomato* str. DC3000	Bacteria	Plant	+	60	(Lindeberg, et al., 2005)
Pseudomonas syringae pv. *syringae* B728a	Bacteria	Plant	+	28	
Pseudomonas syringae pv. *phaseolicola* 1448A	Bacteria	Plant	+	36	
Shigella sp.	Bacteria	Human	+	14	(Ogawa, et al., 2008)
Pseudomonas aeruginosa	Bacteria	Animal and Human	+	4	(Engel and Balachandran, 2009)
Bordetella bronchiseptica	Bacteria	Mammals	+	4	(Panina, et al., 2005)
Yersinia spp.	Bacteria	Human and mammals	+	6	(Bliska, 2000)
Salmonella sp.	Bacteria	Human	+	6	(Ohl and Miller, 2001; Schlumberger and Hardt, 2006)
Enterotoxigenic Escherichia coli	Bacteria	Human	+	19	(Clarke, et al., 2003; Iguchi, et al., 2009)
Ralstonia solanacearum	Bacteria	Plant	+	74	(Poueymiro and Genin, 2009)
Xanthomonas sp.	Bacteria	Plant	+	27	(Kay and Bonas, 2009)
Erwinia amylovora	Bacteria	Plant	+	5	(Boureau, et al., 2006)
Pantoea agglomerans	Bacteria	Plant and animal	+	6	(Barash and Manulis-Sasson, 2007)
Species used in Criteria 5	Kindom	Host	TTSS	Refseq from NCBI	Reference
Pseudomonas putida	Bacteria	NA	-	NC 002947	(Nelson, et al., 2003)
Pseudomonas aeruginosa	Bacteria	Animal a nd human	+	NC 011770	(Winstanley, et al., 2009)
Pseudomonas fluorescens	Bacteria	human	-	NC 007492NC 004129	(Paulsen, et al., 2005)
Pseudomonas mendocina	Bacteria	human	+	NC 009439	NA
Pseudomonas stutzeri	Bacteria	human	-	NC 009434	(Yan, et al., 2008)
Pseudomonas entomophila	Bacteria	pathogenic for insects	-	NC 008027	(Vodovar, et al., 2006)

Supplemental Tab. 2: Candidate effectors identified by EFFECTORSEARCH when Score = 4

Locus	Strain	Num	Locus	Strain	Num
PSPTO_A0012	DC3000	5	Psyr_1193	B728a	5
PSPTO_4776	DC3000	5	Psyr_1188	B728a	5
PSPTO_4727	DC3000	5	Psyr_1183	B728a	5
PSPTO_4724	DC3000	5	Psyr_4919	B728a	5
PSPTO_4722	DC3000	5	Psyr_4269	B728a	5
PSPTO_4718	DC3000	5	Psyr_1889	B728a	5
PSPTO_4001	DC3000	5	Psyr_1192	B728a	5
PSPTO_1568	DC3000	5	Psyr_4659	B728a	4
PSPTO_1382	DC3000	5	Psyr_3839	B728a	4
PSPTO_1377	DC3000	5	Psyr_1218	B728a	4
PSPTO_1373	DC3000	5	Psyr_1186	B728a	4
PSPTO_1372	DC3000	5	Psyr_0779	B728a	4
PSPTO_0901	DC3000	5	Psyr_1224	B728a	4
PSPTO_0883	DC3000	5	Psyr_1219	B728a	4
PSPTO_0877	DC3000	5	Psyr_1189	B728a	4
PSPTO_0876	DC3000	5	Psyr_0778	B728a	4
PSPTO_0588	DC3000	5	Psyr_0738	B728a	4
PSPTO_0044	DC3000	5	PSPPH_1264	1448A	5
PSPTO_A0019	DC3000	4	PSPPH_1295	1448A	5
PSPTO_A0018	DC3000	4	PSPPH_1296	1448A	5
PSPTO_5628	DC3000	4	PSPPH_1443	1448A	5
PSPTO_5354	DC3000	4	PSPPH_4366	1448A	5
PSPTO_4732	DC3000	4	PSPPH_A0010	1448A	5
PSPTO_4726	DC3000	4	PSPPH_A0012	1448A	5
PSPTO_4720	DC3000	4	PSPPH_1272	1448A	5
PSPTO_4703	DC3000	4	PSPPH_0171	1448A	5
PSPTO_4691	DC3000	4	PSPPH_4326	1448A	5
PSPTO_4597	DC3000	4	PSPPH_4736	1448A	5
PSPTO_4594	DC3000	4	PSPPH_0767	1448A	5
PSPTO_4331	DC3000	4	PSPPH_A0120	1448A	5
PSPTO_4250	DC3000	4	PSPPH_1263	1448A	4
PSPTO_4101	DC3000	4	PSPPH_1266	1448A	4
PSPTO_3087	DC3000	4	PSPPH_1268	1448A	4
PSPTO_1406	DC3000	4	PSPPH_1273	1448A	4
PSPTO_1405	DC3000	4	PSPPH_1424	1448A	4
PSPTO_1375	DC3000	4	PSPPH_1269	1448A	4
PSPTO_1370	DC3000	4	PSPPH_1274	1448A	4
PSPTO_0905	DC3000	4	PSPPH_2198	1448A	4
PSPTO_0904	DC3000	4	PSPPH_2351	1448A	4
PSPTO_0838	DC3000	4	PSPPH_3028	1448A	4
PSPTO_0589	DC3000	4	PSPPH_3498	1448A	4
PSPTO_0502	DC3000	4	PSPPH_4540	1448A	4
PSPTO_0501	DC3000	4	PSPPH_0784	1448A	4
PSPTO_0474	DC3000	4	PSPPH_A0009	1448A	4
PSPTO_0473	DC3000	4	PSPPH_A0031	1448A	4
PSPTO_0061	DC3000	4	PSPPH_A0075	1448A	4
Psyr_1184	B728a	5	PSPPH_A0087	1448A	4
Psyr_4326	B728a	5	PSPPH_A0110	1448A	4
Psyr_3813	B728a	5	PSPPH_A0127	1448A	4
Psyr_1220	B728a	5	PSPPH_A0129	1448A	4

Note: Num is the number of criteria that were met for each candidate effector.

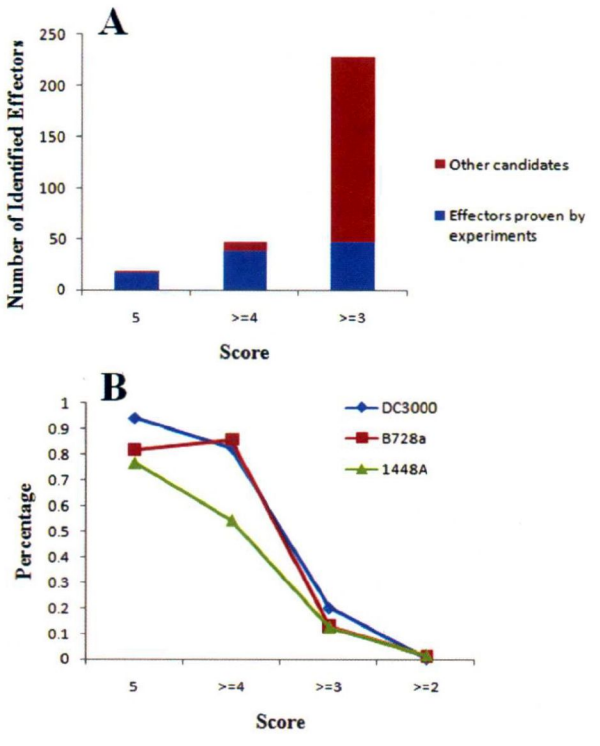

Fig. 2 Higher values of Score identified most functionally confirmed effectors as candidates but fewer total candidates, whereas lower values of Score identified more total candidates but fewer functionally confirmed effectors.

(A) The number of candidates identified in *P. syringae* DC3000 strain by EFFECTORSEARCH for a given value of the Score parameter. (B) The proportion of candidates identified by EFFECTORSEARCH are in fact functionally confirmed effectors for each of three strains of *P. syringae* for a given value of Score.

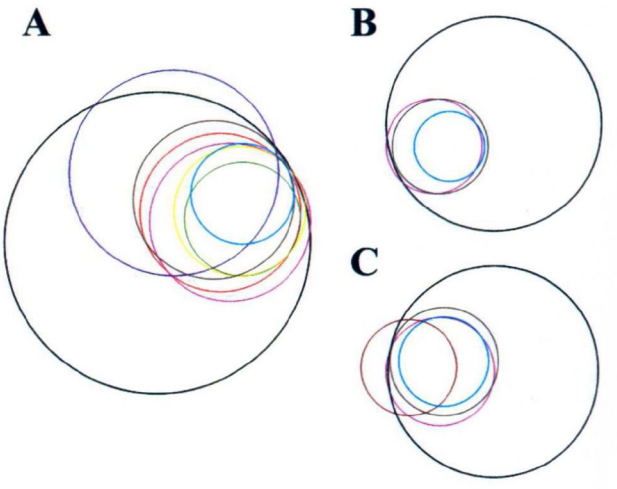

Fig. 3 *EFFECTORSEARCH* found almost all effectors identified by previous researchers using other methods and also found previously unidentified candidate effectors.

(A) Effectors identified in DC3000. (B) Effectors identified in B728a. (C) Effectors identified in 1448A.

Notes: Circles with different colors represent the number of effectors identified by using a particular method: the more effectors identified, the larger the circle. The overlap between circles represents the number of effectors identified by more than one method. Blue circles represent effectors found by Ferreira *et al*. (2006); Red circles, Schechter *et al*, (2006), Pink circles, Lindeberg *et al*. (2006); Green circles, Guttman *et al*. (2002); Yellow circles, Collmer *et al*. (2002); Brown circles, Vencato *et al*. (2006). Sky-blue, Grey, and Black circles represent effectors found by *EFFECTORSEARCH* with *Score* = 3, 4 and 5 respectively.

References

[1]Buell CR, Joardar V, Lindeberg M, et al., The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000 [J]. Proc Natl Acad Sci U S A, 2003, 100 (18): 10181-10186. DOI: 10.1073/pnas.1731982100

[2]Galan JE, Wolf-Watz H. Protein delivery into eukaryotic cells by type III secretion machines [J]. Nature, 2006, 444(7119): 567-573. DOI:10.1038/nature05272

[3]Studholme D, Ibanez SG, Maclean D, et al. A draft genome sequence and functional screen reveals the repertoire of type III secreted proteins of *Pseudomonas syringae* pathovar *tabaci* 11528 [J]. BMC Genomics, 2009, 10: 395. DOI:10.1186/1471-2164-10-395

[4]Petnicki-Ocwieja T, Schneider DJ, Tam VC, et al. Genomewide identification of proteins secreted by the Hrp type III protein secretion system of *Pseudomonas syringae* pv. *tomato* DC3000 [J]. Proc Natl Acad Sci U S A, 2002, 99(11): 7652-7657. DOI: 10.1073/pnas.112183899

[5]Ellis JG, Dodds PN, Lawrence GJ. The role of secreted proteins in diseases of plants caused by rust, powdery mildew and smut fungi [J]. Curr Opin Microbiol, 2007, 10(4): 326-331.

[6]Mudgett MB. New insights to the function of phytopathogenic bacterial type III effectors in plants [J]. Annu Rev Plant Biol, 2005, 56: 509-531. DOI: 10.1146/annurev.arplant.56.032604.144218

[7]Greenberg JT. Programmed cell death in plant-pathogen interactions [J]. Annu Rev Plant Physiol Plant Mol Biol, 1997, 48: 525-545. DOI: 10.1146/annurev.arplant.48.1.525

[8]Jehl MA, Arnold R, Rattei T. Effective-a database of predicted secreted bacterial proteins [J]. Nucl Acids Res, 2011, 39(suppl 1): D591-D595. DOI: 10.1093/nar/gkq1154

[9]McDermott JE, Corrigan A, Peterson E, et al. Computational prediction of type III and IV secreted effectors in gram-negative bacteria [J]. Infect Immun, 2011, 79(1): 23-32. DOI: 10.1128/IAI.00537-10

[10]Ferreira AO, Myers CR, Gordon JS, et al. 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 [J]. Mol Plant Microbe Interact, 2006, 19(11): 1167-1179. DOI: 10.1094/MPMI-19-1167

[11]Guttman DS, Vinatzer BA, Sarkar SF, et al. A functional

screen for the type III (Hrp) secretome of the plant pathogen *Pseudomonas syringae* [J]. Science, 2002, 295(5560): 1722-1726. DOI: 10.1126/science.295.5560.1722

[12]Schechter LM, Vencato M, Jordan KL et al. Multiple approaches to a complete inventory of *Pseudomonas syringae* pv. *tomato* DC3000 type III secretion system effector proteins [J]. Mol Plant Microbe Interact, 2006, 19(11): 1180-1192. DOI: 10.1094 / MPMI -19-1180

[13]Arnold R, Brandmaier S, Kleine F, et al. Sequence-based prediction of type III secreted proteins [J]. PLoS Pathog, 2009, 5(4): e1000376. DOI: 10.1371/journal.ppat.1000376

[14]Samudrala R, Heffron F, McDermott JE. Accurate prediction of secreted substrates and identification of a conserved putative secretion signal for type III secretion systems [J]. PLoS Pathog, 2009, 5(4): e1000375. DOI: 10.1371/journal.ppat.1000375

[15]Collmer A, Lindeberg M, Petnicki-Ocwieja T, et al. Genomic mining type III secretion system effectors in *Pseudomonas syringae* yields new picks for all TTSS prospectors [J]. Trends Microbiol, 2002, 10(10): 462-469. DOI: 10.1016/S0966-842X(02)02451-4

[16]Lindeberg M, Cartinhour S, Myers CR, et al. 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 [J]. Mol Plant Microbe Interact, 2006, 19(11): 1151-1158. DOI: 10.1094/MPMI -19-1151

[17]Fouts DE, Abramovitch RB, Alfano JR, et al. Genomewide identification of *Pseudomonas syringae* pv. *tomato* DC3000 promoters controlled by the HrpL alternative sigma factor [J]. Proc Natl Acad Sci U S A, 2002, 99(4): 2275-2280. DOI: 10.1073/pnas.032514099

[18]Lindeberg M, Stavrinides J, Chang JH, et al. Proposed guidelines for a unified nomenclature and phylogenetic analysis of type III Hop effector proteins in the plant pathogen *Pseudomonas syringae* [J]. Mol Plant Microbe Interact, 2005, 18(4): 275-282. DOI: 10.1094/MPMI -18-0275

[19]Greenberg JT, Vinatzer BA. Identifying type III effectors of plant pathogens and analyzing their interaction with plant cells [J]. Curr Opin Microbiol, 2003, 6(1): 20-28. DOI: 10.1016/S1369-5274(02)00004-8

[20]Hacker J, Carniel E. Ecological fitness, genomic islands and bacterial pathogenicity - A Darwinian view of the evolution of microbes [J]. Embo Rep, 2001, 2(5): 376-381. DOI: 10.1093/embo-reports/kve097

[21]Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool [J]. J Mol Biol, 1990, 215(3): 403-410. DOI: 10.1006/jmbi.1990.9999 pmid: 2231712

[22]Zwiesler-Vollick J, Plovanich-Jones AE, Nomura K, et al. Identification of novel hrp-regulated genes through functional genomic analysis of the *Pseudomonas syringae* pv. *tomato* DC3000 genome [J]. Mol Microbiol, 2002, 45(5): 1207-1218. DOI: 10.1046/j.1365-2958.2002.02964.x

[23]Chang JH, Urbach JM, Law TF, et al. A high-throughput, near-saturating screen for type III effector genes from *Pseudomonas syringae* [J]. Proc Natl Acad Sci U S A, 2005, 102(7): 2549-2554. DOI: 10.1073/pnas.0409660102

[24]Vinatzer BA, Jelenska J, Greenberg JT. Bioinformatics correctly identifies many type III secretion substrates in the plant pathogen *Pseudomonas syringae* and the biocontrol isolate *P. fluorescens* SBW25 [J]. Mol Plant Microbe Interact, 2005, 18(8): 877-888. DOI: 10.1094/MPMI -18-0877

[25]Badel JL, Shimizu R, Oh HS, et al. A *Pseudomonas syringae* pv. *tomato avrE1/hopM1* mutant is severely reduced in growth and lesion formation in tomato [J]. Mol Plant Microbe Interact, 2006, 19(2): 99-111. DOI: 10.1094/MPMI -19-0099

[26]Preston G, Huang HC, He SY, et al. The HrpZ proteins of *Pseudomonas syringae* pvs. *syringae*, *glycinea*, and *tomato* are encoded by an operon containing *Yersinia* ysc homologs and elicit the hypersensitive response in tomato but not soybean [J]. Mol Plant Microbe Interact, 1995, 8(5): 717-32. DOI: 10.1094/MPMI-8-0717

[27]Vencato M, Tian F, Alfano JR, et al. Bioinformatics-enabled identification of the HrpL regulon and type III secretion system effector proteins of *Pseudomonas syringae* pv. *phaseolicola* 1448A [J]. Mol Plant Microbe Interact, 2006, 19(11): 1193-1206. DOI: 10.1094/MPMI -19-1193

[28]Deng WL, Rehm AH, Charkowski AO, et al. *Pseudomonas syringae* exchangeable effector loci: sequence diversity in representative pathovars and virulence function in *P. syringae* pv. syringae B728a [J]. J Bacteriol, 2003, 185(8): 2592-2602. DOI: 10.1128/JB.185.8.2592-2602.2003

Received: 2012-05-10; Revision accepted: 2012-05-30

EFFECTORSEARCH：病原菌基因组中三型分泌系统效应蛋白的预测

张　雯[1], Joy Bergelson[2]

摘　要：病原菌为了生存和侵入寄主细胞，经过漫长的进化形成了多种可介导大分子转运的分子机制，细菌III型分泌系统(T3SS)是其中最重要的一种。T3SS又被称为“细菌注射器”，病原菌利用此系统可以将效应蛋白注射到靶细胞中。近年来发现，T3SS不仅存在于在人类、动物、甚至植物的多种致病菌中，在其它环境下的细菌中也存在。本文介绍了一种最新的生物信息学预测工具(EFFECTORSEARCH)，可用于在微生物基因组中预测T3SS的基因，并且在本文中着重介绍了其在丁香假单胞菌(*Pseudomonas syringae*)基因组中T3SS基因的预测结果。

关键词：Effectorsearch；病原菌基因组；T3SS；预测

中图分类号：R　　**文献标识码**：A　　**文章编号**：1002－2694(2012)06－0528－08