

Review

Open Access

Gene Ontology annotation highlights shared and divergent pathogenic strategies of type III effector proteins deployed by the plant pathogen *Pseudomonas syringae* pv *tomato* DC3000 and animal pathogenic *Escherichia coli* strains

Magdalen Lindeberg^{†1}, Bryan S Biehl^{†2}, Jeremy D Glasner², Nicole T Perna², Alan Collmer¹ and Candace W Collmer^{*1,3}

Address: ¹Department of Plant Pathology, Cornell University, Ithaca, NY 14850, USA, ²Department of Genetics, University of Wisconsin, Madison WI, USA and ³Department of Biological and Chemical Sciences, Wells College, Aurora, NY 13026, USA

Email: Magdalen Lindeberg - ML16@cornell.edu; Bryan S Biehl - bsbiehl@wisc.edu; Jeremy D Glasner - jeremy@genome.wisc.edu; Nicole T Perna - ntperna@wisc.edu; Alan Collmer - arc2@cornell.edu; Candace W Collmer* - ccollmer@wells.edu

* Corresponding author †Equal contributors

Published: 19 February 2009

BMC Microbiology 2009, 9(Suppl 1):S4 doi:10.1186/1471-2180-9-S1-S4

This article is available from: <http://www.biomedcentral.com/1471-2180/9/S1/S4>

© 2009 Lindeberg et al; licensee BioMed Central Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Genome-informed identification and characterization of Type III effector repertoires in various bacterial strains and species is revealing important insights into the critical roles that these proteins play in the pathogenic strategies of diverse bacteria. However, non-systematic discipline-specific approaches to their annotation impede analysis of the accumulating wealth of data and inhibit easy communication of findings among researchers working on different experimental systems. The development of Gene Ontology (GO) terms to capture biological processes occurring during the interaction between organisms creates a common language that facilitates cross-genome analyses. The application of these terms to annotate type III effector genes in different bacterial species – the plant pathogen *Pseudomonas syringae* pv *tomato* DC3000 and animal pathogenic strains of *Escherichia coli* – illustrates how GO can effectively describe fundamental similarities and differences among different gene products deployed as part of diverse pathogenic strategies. In depth descriptions of the GO annotations for *P. syringae* pv *tomato* DC3000 effector AvrPtoB and the *E. coli* effector Tir are described, with special emphasis given to GO capability for capturing information about interacting proteins and taxa. GO-highlighted similarities in biological process and molecular function for effectors from additional pathosystems are also discussed.

Background

Bacterial pathogens subvert defenses and generate favorable niches in diverse eukaryotes through an array of extracellular factors. Most of these factors are proteins transported out of the bacterial cell by one of several secretion pathways, numerically distinguished as type I through type VI. Proteins secreted by the type III secretion

system (T3SS) are critical to pathogens of diverse hosts; the best studied of these bacteria include enteric pathogens of animals, *E. coli*, *Salmonella*, and *Yersinia*, and plant pathogens in the genera *Pseudomonas*, *Xanthomonas*, and *Ralstonia* [1,2]. Related secretion systems are utilized during beneficial types of symbiosis, with the best studied

being the T3SS in the nitrogen-fixing legume symbionts, *Rhizobium* [3,4].

The ever-growing number of complete genome sequences has prompted intensive genome-informed investigations of Type III effector repertoires in various bacterial strains and species. However, a long history of non-systematic, discipline-specific approaches to the annotation of Type III effectors (and virulence genes in general) has created a significant impediment to rapid analysis of this wealth of new data. Effector gene names typically vary, with two or more three-letter gene names often used for effectors from the same species (e.g., *sip* and *sop* in *Salmonella* or *avr* and *hop* in *P. syringae*) and annotation of product names being similarly idiosyncratic. Attempts have been made to systematize three letter gene name assignments for select groups of organisms [5,6], but even with a rigorously applied nomenclature, only limited information can be captured in this way. Furthermore, the high rates of horizontal transfer observed for effector genes [5], coupled with an accelerated rate of evolution for some [7,8], can confound recognition of related effectors through standard analysis of homology. As a result, researchers interested in broad comparisons of gene products and pathosystems must invest significant amounts of time piecing together their own summary from the primary literature.

The Gene Ontology – a universal language for capturing effector characteristics

The Gene Ontology (GO) was originally developed by researchers working on eukaryotic model systems as a controlled vocabulary for describing processes common to diverse organisms [9]. The three ontologies that make up GO are designed to capture information on the cellular location, molecular functions, and biological processes that characterize individual gene products. GO terms are organized into a tree-like hierarchy where more general, high level terms are the parents of more specific child terms. Gene products can be annotated to as many terms as are applicable, with the level of specificity dependent on the extent of characterization. Evidence codes attached to individual annotations document the nature of supporting evidence and identifiers for the associated publications provide direct links to the source of the data.

The Plant-associated Microbe Gene Ontology (PAMGO) project <http://pamgo.vbi.vt.edu/> was initiated for the purpose of creating GO terms that specifically capture cellular locations and biological processes relevant to interactions between organisms. Of the more than 700 new GO terms created as part of this project; most are found under the "interspecies interaction between organisms" parent in the Biological Process Ontology. Term development has been accompanied by focused efforts on the part of

PAMGO members to comprehensively annotate effectors in selected bacterial pathogens – specifically, the plant pathogen *Pseudomonas syringae* pv *tomato* DC3000 (*Pto* DC3000) and numerous enterics including the plant pathogen *Dickeya dadantii* and animal pathogenic strains of *E. coli*.

Pto DC3000 and *E. coli* 0157:H7 represent useful case studies for initiation of a global effector annotation project. Both pathogens require a wide range of T3SS-dependent effectors to establish infection within their respective hosts. Furthermore, as pathogens of hosts in both the plant and animal kingdoms, they illustrate the utility of GO's multi-level structure for conceptualizing shared and divergent aspects of their pathogenic strategies.

Pseudomonas syringae pv. *tomato* DC3000

Pto DC3000 is a pathogen of tomato and *Arabidopsis*, was the first *P. syringae* strain sequenced to completion, and is a model for the study of bacterial-plant interactions [10]. T3SS effector proteins, identified on the basis of their regulation by the HrpL alternative sigma factor and their passage out of the bacterial cell via the T3SS, have long been known to play a critical role in pathogenicity and host-range determination of *P. syringae* pathovars. Indeed, cataloguing their complete repertoire represented one of the chief motivations for sequencing the *Pto* DC3000 genome. More than 50 effector families, defined by phylogenetic grouping [11], have been identified among the *P. syringae* pathovars, with over 36 families found in *Pto* DC3000. The majority of these were identified using a combination of BLAST analysis of predicted genes against previously identified effectors and iterative pattern-based searches using the conserved HrpL binding site and N-terminal sequence patterns associated with T3SS targeting [11].

Since their initial identification as substrates of the T3SS, research on the *Pto* DC3000 effectors has yielded new insights into their molecular functions, cellular destinations within the host, and the biological processes in which they participate. To date, over 300 Gene Ontology annotations have been generated for 36 effector genes as part of the PAMGO project, with the vast majority of annotations concerning processes that occur during the interaction between microbes and their host organisms. An example of the range of processes in which the effectors engage is illustrated by the *Pto* DC3000 effector protein AvrPtoB (HopAB2) (see the table in Figure 1). Like many other effectors, AvrPtoB is annotated to terms in all three ontologies. Within the Biological Process Ontology, terms range from the more general such as "GO:0009405 pathogenesis" and "GO:0044412 growth or development of symbiont within the host", applicable to a wide range

Biological Process Ontology	GO ID	GO term name	Publication ^b	Evidence Code ^c	with/from ID ^d	host taxon ^e				
Parent terms ^a										
GO:0044403 : symbiosis, encompassing mutualism through parasitism	GO:0051701 : interaction with host	pathogenesis	PMID:12505984	IDA	-	4081				
			PMID:14756767	IDA	-	4097				
			PMID:14756767	IDA	-	3702				
	GO:0044003 : modification by symbiont of host morphology or physiology	GO:0052049	interaction with host via protein secreted by type III secretion system	PMID:12062102	IDA	-	4081			
				PMID:16678099	IDA	-	3702			
	GO:0052031 : modulation by symbiont of host defense response	GO:0044412	growth or development of symbiont within host	PMID:15672817	IDA	-	4081			
				PMID:17637671	IDA	-	4084			
		GO:0034053 : modulation by symbiont of host defense-related programmed cell death	GO:0075346	modification by symbiont of host protein by ubiquitination	PMID:12062102	IDA	-	4081		
					GO:0034054	negative regulation by symbiont of host defense-related PCD	PMID:12505984	IDA	-	4100
							PMID:14756767	IDA	-	3702
		GO:0052553 : modulation by symbiont of host immune response	GO:0052034	negative regulation by symbiont of pathogen-associated molecular pattern-induced host innate immunity	PMID:16373536	IMP	-	4081		
					PMID:14756767	IDA	-	4097		
		GO:0052031 : modulation by symbiont of host defense response	GO:0033660	negative regulation by symbiont of host resistance gene-dependent defense response	PMID:16678099	IDA	-	3702		
					PMID:12505984	IDA	-	4100		
		GO:0052031 : modulation by symbiont of host defense response	GO:0075132	negative regulation by symbiont of host protein kinase-mediated signal transduction	PMID:12505984	IDA	-	4081		
PMID:12505984	IDA				-	4081				
GO:0052031 : modulation by symbiont of host defense response	GO:0033663	negative regulation by symbiont of host defense-related protein level	PMID:16678099	IDA	-	3702				
			PMID:17637671	IDA	-	4081				
GO:0052031 : modulation by symbiont of host defense response	GO:0052021	modulation by symbiont of ethylene levels in host	PMID:14756767	IDA	-	4097				
			PMID:16167902	IEP, IMP	-	4081				
GO:0052031 : modulation by symbiont of host defense response	GO:0033668	negative regulation by symbiont of host apoptosis	PMID:12505984	IDA	-	4100				
			PMID:12505984	IDA	-	4932				
GO:0052031 : modulation by symbiont of host defense response	GO:0052041	negative regulation by symbiont of host programmed cell death	PMID:12505984	IDA	-	4932				
			PMID:12505984	IDA	-	4932				
Cellular Component Ontology	GO:0030430	host cell cytoplasm	PMID:16367952	IDA	-	4100				
Molecular Function Ontology	GO:0019901	protein kinase binding	PMID:12062102	IPI	UniProt: Q40234	-				
			PMID:17637671	IPI	UniProt: Q40126	-				
	GO:0005515	protein binding	PMID:17637671	IPI	UniProt: Q9LLD3	-				
	GO:004842	ubiquitin-protein ligase activity	PMID:16373536	IDA	-	-				
	GO:0031624	ubiquitin conjugating enzyme binding	PMID:16373536	IMP	-	-				
PMID:16373536			ISS	UniProt: P62877, Q8VZ40	-					

Figure 1
Gene Ontology annotation for the Pto DC3000 Type III effector AvrPtoB. ^aIndicates the nearest common parent term in the GO term hierarchy. Terms sharing the specified parent are delimited by dashed lines. ^bIndicates the publication supporting annotation of AvrPtoB to the specified GO term. ^cIndicates the nature of the evidence supporting the annotation; IDA, inferred from direct evidence; IEP, inferred from expression profile; IPI, inferred from physical interaction; IMP, inferred from mutant phenotype; ISS, inferred from sequence similarity. ^dIndicates the Uniprot accession number of the interacting protein, where inferred from physical evidence, or of the similar protein, where inferred from sequence similarity. ^eIndicates the taxon ID of the host where biological processes occurred in relation to a host organism.

of virulence factors in diverse pathogens, to more specific terms such as "GO:0052049 interaction with host via protein secreted by type III secretion system" that specifically identifies the Type III effectors.

AvrPtoB and other *P. syringae* effectors play significant roles in modulating the host defense response (GO:0052031), and as part of PAMGO term development, an extensive tree of child terms was created to capture the variety of processes contributing to this phenomenon. Though many of these terms are, at least for the present, used only for bacteria-plant interactions, it is critical to the utility of GO that terms be defined using language that is meaningful across many pathosystems. The phrase "hypersensitive response" serves as a useful example. While this term is commonly used among plant pathologists to refer to rapid defense-associated plant cell death at the site of infection, to researchers in animal systems it can have very different allergy-related or behavioral connotations. Therefore, newly developed PAMGO terms avoid using "hypersensitive response" in the term name and instead use term names such as "GO:0034053 modulation by symbiont of host defense-related programmed cell death" to annotate such bacterial effector activity. At the same time, the previously existing GO term "GO:0009626 hypersensitive response" was modified at the request of PAMGO collaborators to "plant-type hypersensitive response," thus clearly matching the new term name with the existing GO definition, which specified plant cells. It is important to note that an investigator searching GO terms for "hypersensitive response" would be pointed to both terms named above by means of the synonym field attached to each GO term. Finally, these two terms illustrate the difference between terms appropriate for annotating genes in the microbe, or symbiont (GO:0034053) versus those in the plant (GO:0009626) that are involved in the process described.

AvrPtoB is annotated to several child terms of "GO:0052031 modulation by symbiont of host defense response" including: "GO:0034054 negative regulation by symbiont of host defense-related programmed cell death [PCD]", "GO:0034055 positive regulation by symbiont of host defense-related programmed cell death", "GO:0033660 negative regulation by symbiont of host resistance gene-dependent defense response", "GO:0075132 negative regulation by symbiont of host protein kinase-mediated signal transduction", and "GO:0052034 negative regulation by symbiont of pathogen-associated molecular pattern-induced host innate immunity". At first glance, these annotations may appear contradictory – after all, how can the same gene product be annotated to both "GO:0034055 positive regulation by symbiont of host defense-related PCD" and "GO:0034054 negative regulation by symbiont of host

defense-related PCD"? In this case, the answer lies in the secondary or dual taxon field incorporated into the GO database as part of the PAMGO project. This field functions to indicate the identities of the organisms between which the interaction is occurring. Thus, closer examination reveals that "GO:0034055 positive regulation by symbiont of host defense-related PCD" applies to AvrPtoB in the *Pto* DC3000 interaction with *S. lycopersicum* (tomato) while "GO:0034054 negative regulation by symbiont of host defense-related PCD" is used to annotate the interaction between *Pto* DC3000 and *Nicotiana benthamiana* (tobacco). In fact, annotation to "GO:0034054 negative regulation by symbiont of host defense-related PCD" is shown in triplicate to reflect interactions of *Pto* DC3000 in three separate hosts – *Nicotiana benthamiana*, *Nicotiana tabacum* cv. *Xanthi*, and *Arabidopsis thaliana*. Where additional clarification of strains and genotypes of interacting organisms is required, users can refer to the associated publications found in the reference field of the GO annotation.

In addition to annotations in the Biological Process ontology, annotations to the Cellular Component and Molecular Function ontologies are also shown. As one of the most thoroughly characterized of the *Pto* DC3000 effectors, AvrPtoB has several Molecular Function annotations that provide insight on the specific enzymatic and binding capabilities by which AvrPtoB accomplishes the processes described above. Molecular Function annotations include: "GO:0019901 protein kinase binding", "GO:0004842 ubiquitin-protein ligase activity", and "GO:0031624 ubiquitin conjugating enzyme binding". Just as documenting the taxa of interacting organisms is critical to the usefulness of biological process terms, so documentation of interacting proteins significantly enhances the value of Molecular Function terms. For example, the interacting proteins for "GO:0019901 protein kinase binding" are Q40234 (*Pto*) and Q40126 (*Fen*), and for "GO:0005515 protein binding", Q9LLD3 (*PtoC*). Most of the evidence codes used for AvrPtoB indicate experimental evidence for the assigned annotations, including IDA (inferred from direct assay), IMP (inferred from mutant phenotype), and IPI (inferred from physical interaction). In contrast, the evidence code ISS (inferred from sequence or structural similarity) indicates that the annotation is based on similarity of the given gene product to an experimentally characterized homolog. Annotations made on the basis of sequence or structural similarity require that the ID of the protein from which the annotation is inferred be included in the with/from column. Unlike AvrPtoB, for which the ISS code is used only once to capture its structural similarity to known E3 ubiquitin ligases (UniProt: P62877, Q8VZ40), GO annotations for effectors in some other *P. syringae* strains rely more extensively on sequence similarity. In such cases

where experimental evidence is lacking, sequence similarity to Pto DC3000 effectors can be used to guide GO annotation of those effectors. (Some important considerations relevant to propagating GO annotations based on sequence similarity are described in the following section.) When sequence similarity is absent, GO annotations can provide clues to candidate functions or biological processes in newly identified gene products based on annotations previously made for other experimentally characterized gene products. For example, once a newly described gene product is found to be secreted and thus annotated to "GO:0052049 interaction with host via protein secreted by type III secretion system", other processes associated with this annotation in other experimentally characterized effectors become candidates for testing. These might include "GO:0044412 growth or development of symbiont within host", "GO:0034055 positive regulation by symbiont of host defense-related PCD", or "GO:0052034 negative regulation by symbiont of pathogen-associated molecular pattern-induced host innate immunity".

Escherichia coli

Like *P. syringae*, many strains of *E. coli* rely on effectors to establish a pathogenic relationship with their host and are the focus of intense interest owing to their ability to cause serious disease in humans. Numerous genomes have recently been sequenced from pathogenic and non-pathogenic *E. coli* strains, and no one strain serves as a general model for the diverse pathogenic strategies found within this species. Consequently, PAMGO consortium members working on the Enterobacteriaceae, in contrast to those working on *P. syringae*, have focused on automated propagation of annotations from a handful of experimentally characterized effectors to homologs in numerous complete and draft genomes of *E. coli* and other enteric bacteria.

E. coli O157:H7 provides a representative example of the current state of experimentally-based annotation of enteric effectors and also reveals the limitations of transitive or similarity-based propagation of annotations. A total of 39 type III effectors have been identified in *E. coli* O157:H7 strain Sakai through a combination of bioinformatics, "secretome" analysis, and translocation assays with Cya fusions [5]. However, the absence of a convenient model host system for *E. coli* O157:H7 has impeded *in vivo* characterization of host interaction phenotypes on the level conducted in *P. syringae*, with the result that annotation of *E. coli* O157:H7 effectors has relied more extensively on inferences made from sequence similarity.

Among the O157:H7 effectors studied in greater depth is the Translocated Intimin Receptor protein (Tir), which plays a key role in bacterial attachment to host cells [12-

15]. Given that attachment has proven a tractable process for studying in cell culture models, it is possible to assign GO annotations to Tir with a specificity comparable to that of AvrPtoB. Tir is secreted into host cells via the LEE (locus of enterocyte effacement) T3SS and then trafficked to the host cell plasma membrane (GO:0020002), where it binds the (also LEE-encoded) intimin protein on the bacterial cell surface. This binding activity is captured by the combination of Molecular Function ontology terms "GO:0051635 bacterial cell surface binding" and "GO:0005515 protein binding" using the IPI evidence code and "with" qualifier to specify the interacting partner as intimin. The role of Tir in bacterial attachment is reflected by the Biological Process term "GO:0044406 adhesion to host", with subsequent effects of Tir on the cascade of host signaling, described by "GO:0052027 modulation by symbiont of defense-related host MAP kinase-mediated signal transduction pathway" and major host cytoskeletal remodeling captured by "GO:0052039 modification by symbiont of host cytoskeleton" easily accommodated by child terms of "GO:0051701 interaction with host". The term "GO:0052057 modification by symbiont of host morphology or physiology via protein secreted by type III secretion system" links Tir to other T3SS effectors, while "GO:0009405 pathogenesis" indicates its role in disease.

The Tir effector also highlights some of the challenges inherent to cross-genome term assignments. Given that transitive annotation based on sequence and structural similarity forms the basis of most annotations in the GO database, discussion of the limitations of such annotations is warranted. Specifically, similar gene products involved in the interaction between organisms can have very different properties depending on both their source organism and the host with which they are interacting. For example, Tir has been shown to have different molecular functions depending on whether it is produced by enterohemorrhagic (EHEC) O157:H7 or enteropathogenic (EPEC) strains of *E. coli*. In EPEC, but not EHEC, tyrosine phosphorylation of Tir by the bacterium plays a key role in initiating the host cell signaling cascades and the cytoskeletal rearrangements involved with formation of a pedestal-like structure immediately under the adherent bacterial cell [13,14]. It is therefore imperative that the annotations assigned to EPEC-derived Tir not be propagated to Tir from EHEC strains. GO provides the option of using the qualifier "NOT" together with an annotation such as "GO:0019901 protein kinase binding" to indicate that the *E. coli* O157:H7 Tir protein is not phosphorylated. However, many GO annotation repositories, including the UW ASAP database of enterobacterial genomes [16]), do not display this qualifier by default, with the result that the "NOT" qualifier is used infrequently. A more in depth discussion of the "NOT" quali-

fier and differences in its use among databases is described by Yon Rhee et al (2008) [17].

In other cases, the properties of effectors and other host interaction factors are simply uncharacterized in particular strains or during interactions with particular hosts. In databases where the host taxon is not readily displayed for annotations to terms in the "interaction between organisms" tree or where the host is specified but with an ISS evidence code, users should consider the possibility that the annotation may not be accurate for all source strains and hosts. When involved in generating annotations based on sequence or structural similarity, users should consider avoiding propagation of those most likely to vary based on source and host. Within the ASAP database, annotations likely to be host-dependent are not routinely propagated with the automated annotation systems used to annotate rapidly accumulating sequence data from "next-generation" sequencing technologies, and transitive annotation of effectors is limited to the general term "GO:0052049 interaction with host via protein secreted by type III secretion system".

Effector repertoire comparison

Although the approaches used in effector characterization and annotation differ between *P. syringae* and *E. coli*, comparison of the assigned terms illustrates how GO can be used to conceptualize the fundamental similarities and differences that exist among different gene products and pathogenic strategies. As previously mentioned, terms such as "GO:0009405 pathogenesis", "GO:0044412 growth or development of symbiont within host", and "GO:0052049 interaction with host via protein secreted by type III secretion system" are broadly applicable to a wide array of effectors in diverse pathosystems. In contrast, other terms are highly specific to effectors from particular pathosystems, revealing fundamental differences in the processes by which Type III effectors influence the bacterial-host interaction. For example, critical stages of adhesion to the host (GO:0044406), are mediated by Type III effectors in *E. coli* and other animal-associated pathogens [18]. In contrast, host adhesion in *P. syringae* is accomplished by fimbriae and exopolysaccharides [19,20] genes which can be further distinguished in the GO with the terms "GO:0043683 type IV pilus biogenesis" and "GO:0045226 extracellular polysaccharide biosynthetic process", respectively. Conversely, "GO:0001907 killing by symbiont of host cells", whether by the natural progression of necrotic disease or by induction of defense-related programmed cell death (captured with the more specific term GO:0052044), is a hallmark of *P. syringae* effector action [21] that is mediated by toxins independent of the T3SS in *E. coli* and other animal pathogens. Examples include cholera toxin deployed by *Vibrio cholera* and pertussis toxin of *Bordetella pertussis*, the

secretion properties of which are described with the terms "GO:0052051 interaction with host via protein secreted by type II secretion system" and "GO:0052050 interaction with host via substance secreted by type IV secretion system", respectively. These examples illustrate the value of annotating to multiple terms, where appropriate, so as to maximally capture both shared and divergent properties exhibited by different virulence factors.

Beyond these broad similarities and differences, shared processes and activities at surprisingly specific levels can also be found. For example, selected *Pto* DC3000 and *E. coli* 0157:H7 effectors modulate host innate immunity (expressed with GO:0052167 and its child terms), with some specifically demonstrated to negatively regulate host innate immunity induced by pathogen-associated molecular patterns (captured with GO:0052034).

A further illustration of GO-highlighted similarities is shown for a select group of effectors from multiple pathosystems in the table in Figure 2. In both plant and animal systems, complex signaling pathways mediate the response to detected pathogens, with elements of the intervening signaling pathways representing the most common targets for effector-mediated suppression of the immune response. This property is reflected by annotation of AvrPtoB as well as effectors AvrPto, HopAO1, and HopAI1 (*P. syringae*); IpaH9.8, OspF (*Shigella*); SspH1 (*Salmonella*); and YopP/J (*Yersinia*) to the term "GO:0052027 modulation by symbiont of host signal transduction pathway". For some effectors from both plant and animal pathosystems, the nature of this process has been more intensively characterized, supporting annotation to more specific child terms such as "GO:0052078 negative regulation by symbiont of defense-related host MAP kinase-mediated signal transduction pathway" and "GO:0052034 negative regulation by symbiont of pathogen-associated molecular pattern-induced host innate immunity". In other cases, the effectors in question await in depth evaluation.

Additional information as to how effectors accomplish modulation of host signaling is captured with terms in the Molecular Function Ontology, and once again common strategies are evident among effectors in diverse systems. For AvrPtoB, IpaH9.8, and SspH1, interference with signaling pathways is mediated by ubiquitination of host kinases, activities captured with the molecular function terms "GO:0019901 protein kinase binding" and "GO:0004842 ubiquitin-protein ligase activity" [22-24]. AvrPtoB ubiquitinates host kinases involved in resistance-gene mediated host immunity [24], while SspH1 ubiquitinates PKN1 [22,25], a host kinase integral to innate immune response signaling pathways. The ability of

Biological Process Ontology

		AvrPtoB	AvrPto	HopAl1	IpaH9.8	SspH1	YopP/J	OspF
GO:0052049	interaction with host via protein secreted by type III secretion system	Black	Black	Black	Black	Black	Black	Black
GO:0009405	pathogenesis	Black	Black	Black	Black	Black	Black	Black
GO:0052027	modulation by symbiont of host signal transduction pathway	Black	Black	Black	Black	Black	Black	Black
GO:0052078	negative regulation by symbiont of defense-related host MAP kinase-mediated signal transduction pathway	Green	Green	Black	Black	Orange	Black	Black
GO:0052034	negative regulation by symbiont of pathogen-associated molecular pattern-induced host innate immunity	Black	Black	Black	Orange	Orange	Black	Black

Molecular Function Ontology

GO:0019901	protein kinase binding	Black	Black	Black	Black	Black	Black	Black
GO:0004842	ubiquitin-protein ligase activity	Black	Green	Green	Black	Black	Orange	Orange
GO:0034598	phosphothreonine lyase activity	Green	Green	Black	Orange	Orange	Orange	Black
GO:0016407	acetyltransferase activity	Green	Green	Green	Orange	Orange	Black	Orange

Figure 2
Comparative Gene Ontology annotation for selected Type III effectors from Pto DC3000 and animal pathogenic genera. Black indicates the identity of effectors annotated to the specified GO term; green, effectors from plant pathogenic bacteria; orange, effectors from animal pathogenic bacteria.

IpaH9.8 to suppress innate immunity also appears linked to ubiquitination – in this case of MAP kinases [22].

Other effectors alter immune signaling pathways using alternative enzymatic activities. These include inactivation of MAP kinase signaling by "GO:0034598 phosphothreonine lyase activity", documented for both OspF [26] and HopAl1 [27], and targeting of MAP kinases by acetyltransferase-activity (GO:0016407) observed for YopP/J [28]. AvrPtoB, in addition to its ubiquitination-dependent suppression of resistance gene-mediated host immunity, suppresses innate immunity through E3-ligase independent targeting of kinase signaling [29]. The specific molecular functions by which this suppression is accomplished have yet to be characterized.

Putting Gene Ontology to work for you

The *Pto* DC3000 and *E. coli* annotations have been deposited in the GO database where searches can be conducted for GO terms or particular gene products. At that site and through linked resources, users can search for gene products annotated to multiple, user-selected GO terms of interest or analyze microarray data for enrichment of

genes annotated to particular terms. The value of the Gene Ontology project is directly proportional to the number and quality of the annotations being contributed, and though intensive efforts have been directed toward annotation of virulence-related gene products in *Pto* DC3000 and *E. coli*, ongoing annotation of these and effectors deployed by a wider range of plant and animal pathogens would greatly enhance the insights that could be gained. Among the steps being taken to facilitate ongoing annotation are proposals that journals request suggested GO terms from manuscript authors akin to requests for keywords. Guidelines are also being developed to aid researchers in identification of appropriate terms for specific protein families. For example, a tutorial on GO term assignment for plant pathogenic effectors is presently available through the *Pseudomonas-Plant Interaction* website <http://www.pseudomonas-syringae.org>.

For users of GO annotation, it is important to remember that annotations exist within a diffuse network of databases that include the primary, organism-specific databases where annotations are chiefly generated, the GO database itself, and additional sites to which annotation

data are downloaded. Significant inconsistencies can and do occur among databases resulting from differences in annotation format, as previously discussed with regard to the "NOT" qualifier, as well as from differences in the frequency of data exchange among databases. In some instances, the differences among databases simply reflect the length of time it takes for changes instituted by the GO Consortium to propagate through the many databases using GO. For example, the dual taxon field pioneered by PAMGO has only recently been added to TIGR-CMR, the database through which *P. syringae* annotations are forwarded to GO. For these reasons, users are encouraged to identify the sources and version numbers of the annotations they are using and include this information in publications making use of these data.

GO annotation represents a vitally important tool for organizing the wealth of biological data that has accompanied the emergence of genomics and high-throughput expression analysis. Through development of terms capturing the interaction between organisms, the PAMGO consortium has added the important domain of interorganismal interactions to the range of processes encompassed by GO, applicable to research on both pathogenic interactions and beneficial symbioses. Creation of the secondary taxon field has additionally provided a means of capturing nuances of interaction observed upon interaction with different hosts. As exemplified by ongoing annotation of effectors in *P. syringae* and *E. coli*, application of these terms to gene products deployed by different organisms interacting with diverse hosts represents a powerful tool for identification of fundamental parallels underlying outwardly dissimilar interactions.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

The authors would like to thank the editors at The Gene Ontology Consortium, in particular Jane Lomax and Amelia Ireland and the members of the PAMGO Consortium, for their collaboration in developing many PAMGO terms. This work was supported by the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service, grant number 2005-35600-16370 and by the U.S. National Science Foundation, grant number EF-0523736.

This article has been published as part of *BMC Microbiology* Volume 9 Supplement 1, 2009: The PAMGO Consortium: Unifying Themes In Microbe-Host Associations Identified Through The Gene Ontology. The full contents of the supplement are available online at <http://www.biomedcentral.com/1471-2180/9?issue=S1>.

References

- Coburn B, Sekirov I, Finlay BB: **Type III Secretion Systems and Disease.** *Clin Microbiol Rev* 2007, **20(4)**:535-549.
- Zhou J-M, Chai J: **Plant pathogenic bacterial type III effectors subdue host responses.** *Current Opinion in Microbiology* 2008, **11(2)**:179-185.
- Marie C, Broughton WJ, Deakin WJ: **Rhizobium type III secretion systems: legume charmers or alarmers?** *Curr Opin Plant Biol* 2001, **4(4)**:336-342.
- Skorpil P, Saad MM, Boukli NM, Kobayashi H, Ares-Orpel F, Broughton WJ, Deakin WJ: **NopP, a phosphorylated effector of Rhizobium sp. strain NGR234, is a major determinant of nodulation of the tropical legumes Flemingia congesta and Tephrosia vogelii.** *Molecular Microbiology* 2005, **57(5)**:1304-1317.
- Tobe T, Beatson SA, Taniguchi H, Abe H, Bailey CM, Fivian A, Younis R, Matthews S, Marches O, Frankel G, et al.: **An extensive repertoire of type III secretion effectors in Escherichia coli O157 and the role of lambdaoid phages in their dissemination.** *PNAS* 2006, **103(40)**:14941-14946.
- Lindeberg M, Stavrinides J, Chang JH, Alfano JR, Collmer A, Dangl JL, Greenberg JT, Mansfield JW, Guttman DS: **Proposed guidelines for a unified nomenclature and phylogenetic analysis of type III hop effector proteins in the plant pathogen Pseudomonas syringae.** *Mol Plant Microbe Interact* 2005, **18**:275-282.
- Ma W, Dong FF, Stavrinides J, Guttman DS: **Type III effector diversification via both pathoadaptation and horizontal transfer in response to a coevolutionary arms race.** *PLoS Genet* 2006, **2(12)**:e209.
- Stavrinides J, Ma W, Guttman DS: **Terminal Reassortment Drives the Quantum Evolution of Type III Effectors in Bacterial Pathogens.** *PLoS Pathogens* 2006, **2(10)**:e104.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al.: **Gene Ontology: tool for the unification of biology.** *Nat Genet* 2000, **25(1)**:25-29.
- Buell CR, Joardar V, Lindeberg M, Selengut J, Paulsen IT, Gwinn ML, Dodson RJ, Deboy RT, Durkin AS, Kolonay JF, et al.: **The complete genome sequence of the Arabidopsis and tomato pathogen Pseudomonas syringae pv. tomato DC3000.** *Proc Natl Acad Sci USA* 2003, **100(18)**:10181-10186.
- Lindeberg M, Cartinhour S, Myers CR, Schechter LM, Schneider DJ, Collmer A: **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* 2006, **19(11)**:1151-1158.
- DeVinney R, Stein M, Reinscheid D, Abe A, Ruschkowski S, Finlay BB: **Enterohemorrhagic Escherichia coli O157:H7 produces Tir, which is translocated to the host cell membrane but is not tyrosine phosphorylated.** *Infect Immun* 1999, **67(5)**:2389-2398.
- Goosney DL, DeVinney R, Finlay BB: **Recruitment of cytoskeletal and signaling proteins to enteropathogenic and enterohemorrhagic Escherichia coli pedestals.** *Infect Immun* 2001, **69(5)**:3315-3322.
- Kenny B, Warawa J: **Enteropathogenic Escherichia coli (EPEC) Tir receptor molecule does not undergo full modification when introduced into host cells by EPEC-independent mechanisms.** *Infect Immun* 2001, **69(3)**:1444-1453.
- Allen-Vercoe E, Waddell B, Livingstone S, Deans J, DeVinney R: **Enteropathogenic Escherichia coli Tir translocation and pedestal formation requires membrane cholesterol in the absence of bundle-forming pili.** *Cell Microbiol* 2006, **8(4)**:613-624.
- Glasner JD, Plunkett G 3rd, Anderson BD, Baumler DJ, Biehl BS, Burland V, Cabot EL, Darling AE, Mau B, Neeno-Eckwall EC, et al.: **Enteropathogen Resource Integration Center (ERIC): bioinformatics support for research on biodefense-relevant enterobacteria.** *Nucleic Acids Res* 2008:D519-523.
- Yon Rhee S, Wood V, Dolinski K, Draghici S: **Use and misuse of the gene ontology annotations.** *Nat Rev Genet* 2008, **9(7)**:509-515.
- Coburn B, Sekirov I, Finlay BB: **Type III secretion systems and disease.** *Clin Microbiol Rev* 2007, **20(4)**:535-549.
- Ude S, Arnold DL, Moon CD, Timms-Wilson T, Spiers AJ: **Biofilm formation and cellulose expression among diverse environmental Pseudomonas isolates.** *Environmental Microbiology* 2006, **8(11)**:1997-2011.
- Roine E, Raineri DM, Romantschuk M, Wilson M, Nunn DN: **Characterization of type IV pilus genes in Pseudomonas syringae pv. tomato DC3000.** *Mol Plant-Microbe Interact* 1998, **11**:1048-1056.
- Greenberg JT: **Programmed cell death: A way of life for plants.** *Proc Natl Acad Sci U S A* 1996, **93(22)**:12094-12097.

22. Rohde JR, Breitskreutz A, Chenal A, Sansonetti PJ, Parsot C: **Type III Secretion Effectors of the IpaH Family Are E3 Ubiquitin Ligases.** *Cell Host & Microbe* 2007, **1(1)**:77-83.
23. Abramovitch RB, Janjusevic R, Stebbins CE, Martin GB: **Type III effector AvrPtoB requires intrinsic E3 ubiquitin ligase activity to suppress plant cell death and immunity.** *Proc Natl Acad Sci USA* 2006.
24. Rosebrock TR, Zeng L, Brady JJ, Abramovitch RB, Xiao F, Martin GB: **A bacterial E3 ubiquitin ligase targets a host protein kinase to disrupt plant immunity.** *Nature* 2007, **448(7151)**:370-374.
25. Haraga A, Miller SI: **A Salmonella type III secretion effector interacts with the mammalian serine/threonine protein kinase PKNI.** *Cellular Microbiology* 2006, **8(5)**:837-846.
26. Arbibe L, Kim DW, Batsche E, Pedron T, Mateescu B, Muchardt C, Parsot C, Sansonetti PJ: **An injected bacterial effector targets chromatin access for transcription factor NF- κ B to alter transcription of host genes involved in immune responses.** *Nat Immunol* 2007, **8(1)**:47-56.
27. Zhang J, Shao F, Li Y, Cui H, Chen L, Li H, Zou Y, Long C, Lan L, Chai J, et al.: **A Pseudomonas syringae effector inactivates MAPKs to suppress PAMP-induced immunity in plants.** *Cell Host & Microbe* 2007, **1(3)**:175-185.
28. Sweet CR, Conlon J, Golenbock DT, Goguen J, Silverman N: **YopJ targets TRAF proteins to inhibit TLR-mediated NF- κ B, MAPK and IRF3 signal transduction.** *Cellular Microbiology* 2007, **9(11)**:2700-2715.
29. He P, Shan L, Lin N-C, Martin GB, Kemmerling B, Nurnberger T, Sheen J: **Specific Bacterial Suppressors of MAMP Signaling Upstream of MAPKKK in Arabidopsis Innate Immunity.** *Cell* 2006, **125(3)**:563-575.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

