

RESEARCH ARTICLE

Mining Host-Pathogen Protein Interactions to Characterize *Burkholderia mallei* Infectivity Mechanisms

Vesna Memišević¹, Nela Zavaljevski¹, Seesandra V. Rajagopala², Keehwan Kwon², Rembert Pieper², David DeShazer³, Jaques Reifman^{1*}, Anders Wallqvist¹

1 Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, Fort Detrick, Maryland, United States of America, **2** J. Craig Venter Institute, Rockville, Maryland, United States of America, **3** Bacteriology Division, U.S. Army Medical Research Institute of Infectious Diseases, Fort Detrick, Maryland, United States of America

* jaques.reifman.civ@mail.mil



OPEN ACCESS

Citation: Memišević V, Zavaljevski N, Rajagopala SV, Kwon K, Pieper R, DeShazer D, et al. (2015) Mining Host-Pathogen Protein Interactions to Characterize *Burkholderia mallei* Infectivity Mechanisms. PLoS Comput Biol 11(3): e1004088. doi:10.1371/journal.pcbi.1004088

Editor: Alice Carolyn McHardy, Heinrich Heine University, GERMANY

Received: August 14, 2014

Accepted: December 15, 2014

Published: March 4, 2015

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files, or available at <http://www.bhsai.org/downloads/hpia>.

Funding: This work was supported by the U.S. Defense Threat Reduction Agency (www.dtra.mil; award number CBS.MEDBIO.02.10.BH.021) and by the U.S. Army Medical Research and Materiel Command (mrmc.amedd.army.mil) as part of the U.S. Army's Network Science Initiative. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Burkholderia pathogenicity relies on protein virulence factors to control and promote bacterial internalization, survival, and replication within eukaryotic host cells. We recently used yeast two-hybrid (Y2H) screening to identify a small set of novel *Burkholderia* proteins that were shown to attenuate disease progression in an aerosol infection animal model using the virulent *Burkholderia mallei* ATCC 23344 strain. Here, we performed an extended analysis of primarily nine *B. mallei* virulence factors and their interactions with human proteins to map out how the bacteria can influence and alter host processes and pathways. Specifically, we employed topological analyses to assess the connectivity patterns of targeted host proteins, identify modules of pathogen-interacting host proteins linked to processes promoting infectivity, and evaluate the effect of crosstalk among the identified host protein modules. Overall, our analysis showed that the targeted host proteins generally had a large number of interacting partners and interacted with other host proteins that were also targeted by *B. mallei* proteins. We also introduced a novel Host-Pathogen Interaction Alignment (HPIA) algorithm and used it to explore similarities between host-pathogen interactions of *B. mallei*, *Yersinia pestis*, and *Salmonella enterica*. We inferred putative roles of *B. mallei* proteins based on the roles of their aligned *Y. pestis* and *S. enterica* partners and showed that up to 73% of the predicted roles matched existing annotations. A key insight into *Burkholderia* pathogenicity derived from these analyses of Y2H host-pathogen interactions is the identification of eukaryotic-specific targeted cellular mechanisms, including the ubiquitination degradation system and the use of the focal adhesion pathway as a fulcrum for transmitting mechanical forces and regulatory signals. This provides the mechanisms to modulate and adapt the host-cell environment for the successful establishment of host infections and intracellular spread.

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

Author Summary

Burkholderia species need to manipulate many host processes and pathways in order to establish a successful intracellular infection in eukaryotic host organisms. *Burkholderia mallei* uses secreted virulence factor proteins as a means to execute host-pathogen interactions and promote pathogenesis. While validated virulence factor proteins have been shown to attenuate infection in animal models, their actual roles in modifying and influencing host processes are not well understood. Here, we used host-pathogen protein-protein interactions derived from yeast two-hybrid screens to study nine known *B. mallei* virulence factors and map out potential virulence mechanisms. From the data, we derived both general and specific insights into *Burkholderia* host-pathogen infectivity pathways. We showed that *B. mallei* virulence factors tended to target multifunctional host proteins, proteins that interacted with each other, and host proteins with a large number of interacting partners. We also identified similarities between host-pathogen interactions of *B. mallei*, *Yersinia pestis*, and *Salmonella enterica* using a novel host-pathogen interactions alignment algorithm. Importantly, our data are compatible with a framework in which multiple *B. mallei* virulence factors broadly influence key host processes related to ubiquitin-mediated proteolysis and focal adhesion. This provides *B. mallei* the means to modulate and adapt the host-cell environment to advance infection.

Introduction

Burkholderia mallei is the causative agent of glanders, a highly contagious disease that primarily affects horses, mules, and donkeys, but is also transmittable to other mammals through direct contact with infected animals [1]. This host-adapted bacterium is equipped with an extensive set of mechanisms for invasion and modulation of eukaryotic host-cell environments. Key mechanisms of *B. mallei* pathogenicity are encoded in virulence factors (proteins required for virulence) that control and promote pathogenic internalization, survival, and replication within host cells [2, 3]. While a number of *B. mallei* proteins associated with pathogenicity have been characterized and mapped to adhesion, endosomal escape and evasion of host-cell autophagy, actin-based motility, multi-nucleated giant cell formation, replication, and cell-to-cell spread [3–7], the identities of their host targets are largely unknown, and the underlying mechanisms by which the bacterial proteins affect these processes are poorly understood.

In our previous study [8], we used a combined computational and experimental strategy to systematically identify and characterize the interactions between *B. mallei* virulence factors and their host targets. We employed several bioinformatics approaches to identify and select a small number of putative and known virulence factors, and used yeast two-hybrid (Y2H) assays to identify their interacting protein partners in human and murine hosts. The analysis of these host-*B. mallei* protein-protein interactions (PPIs) allowed us to identify three novel *B. mallei* ATCC 23344 virulence factors and show that they attenuated *B. mallei* virulence in mouse aerosol challenge experiments. Although our PPI data contained extensive interactions between multiple host proteins and *B. mallei* proteins, we did not fully explore these data to more generally characterize *B. mallei* virulence mechanisms. Here, we performed a systematic analysis of these interactions to investigate the mechanisms by which *B. mallei* virulence factors interact with host proteins to establish infection, evade host immune responses, and spread within the host. We evaluated whether the virulence factors target specific (non-random) host proteins and processes and whether they jointly affect entry into and survival within the host

cells. Furthermore, we evaluated whether we could detect commonalities in Gram-negative bacterial host-pathogen interactions among *B. mallei*, *Yersinia pestis*, and *Salmonella enterica* virulence factors.

A number of studies have used small- or large-scale experiments to analyze Gram-negative bacteria and their host interactions [9–13]. Although the identified interactions represent only a fraction of all possible interactions between host and pathogen proteins (ranging from less than 10 interactions to a few thousand PPIs), they have proved to be a valuable source of information about bacterial pathogenicity mechanisms. Analyses of these host-pathogen PPI datasets showed that virulence-associated pathogen proteins preferentially target host proteins involved in biological processes essential for cell vitality, e.g., signaling, cell cycle, or immune response [9–13]. Additionally, other studies demonstrated that similarities in host-pathogen PPIs can be used to predict novel host proteins that are targeted by bacterial proteins [14–16].

Our analysis showed that *B. mallei* virulence factors targeted host proteins that had a large number of interacting partners and were closely connected to each other. In addition, the analysis revealed specific host processes relevant to *B. mallei* virulence factors' pathogenicity, e.g., signaling and communication, protein modification and regulation, and cytoskeleton organization, and suggested that virulence factors preferentially targeted multifunctional host proteins, thereby affecting multiple host cellular processes simultaneously. When we used all of our interaction data, including host interactions with putative but not validated *B. mallei* virulence factors, we identified additional host processes and molecular pathways that were previously experimentally associated with *B. mallei* pathogenicity [2, 17–21]. Moreover, our evaluation of the relationship between targeted host proteins involved in different processes and pathways supported a previously observed mechanism for bacterial interference with eukaryotic hosts: virulence factors can focus interference by targeting key host proteins whose effect can propagate through and influence multiple host processes and pathways [2, 17, 18]. Additionally, we introduced a novel Host-Pathogen Interaction Alignment (HPIA) algorithm and used it to explore similarities between host-pathogen interactions of *B. mallei*, *Y. pestis* [13], and *Salmonella enterica* subsp. *enterica* serovar Typhimurium [12]. Using the HPIA algorithm, we identified a statistically significant number of functionally similar host-pathogen interactions between these three PPI datasets. We inferred putative roles for *B. mallei* proteins based on the role of their aligned *Y. pestis* and *S. enterica* partners and showed that up to 73% of the putatively annotated *B. mallei* protein roles matched their existing annotations.

Our findings show that host-pathogen interactions represent a rich source of information about molecular mechanisms of pathogenicity. A key insight from these analyses into *Burkholderia* pathogenicity is the concerted targeting of the ubiquitination degradation system and use of the focal adhesion pathway as a fulcrum for signaling and changing cell morphology. These mechanisms provide *B. mallei* with the ability to modulate and adapt the host-cell environment to establish intracellular host infections.

Results/Discussion

We created an inclusive set of human-*B. mallei* PPIs by merging human-*B. mallei* and orthologous murine-*B. mallei* protein interaction data identified in our previous Y2H screens [8]. The resulting dataset consisted of 1,235 unique interactions between 21 *B. mallei* and 828 human proteins. Fig. 1 shows these interactions and their Y2H-library origins. It also shows that the majority of the *B. mallei* proteins interacted with unique host proteins, i.e., 615 (74%) of host proteins interacted with a single *B. mallei* protein. Importantly, the bulk of the host-*B. mallei*

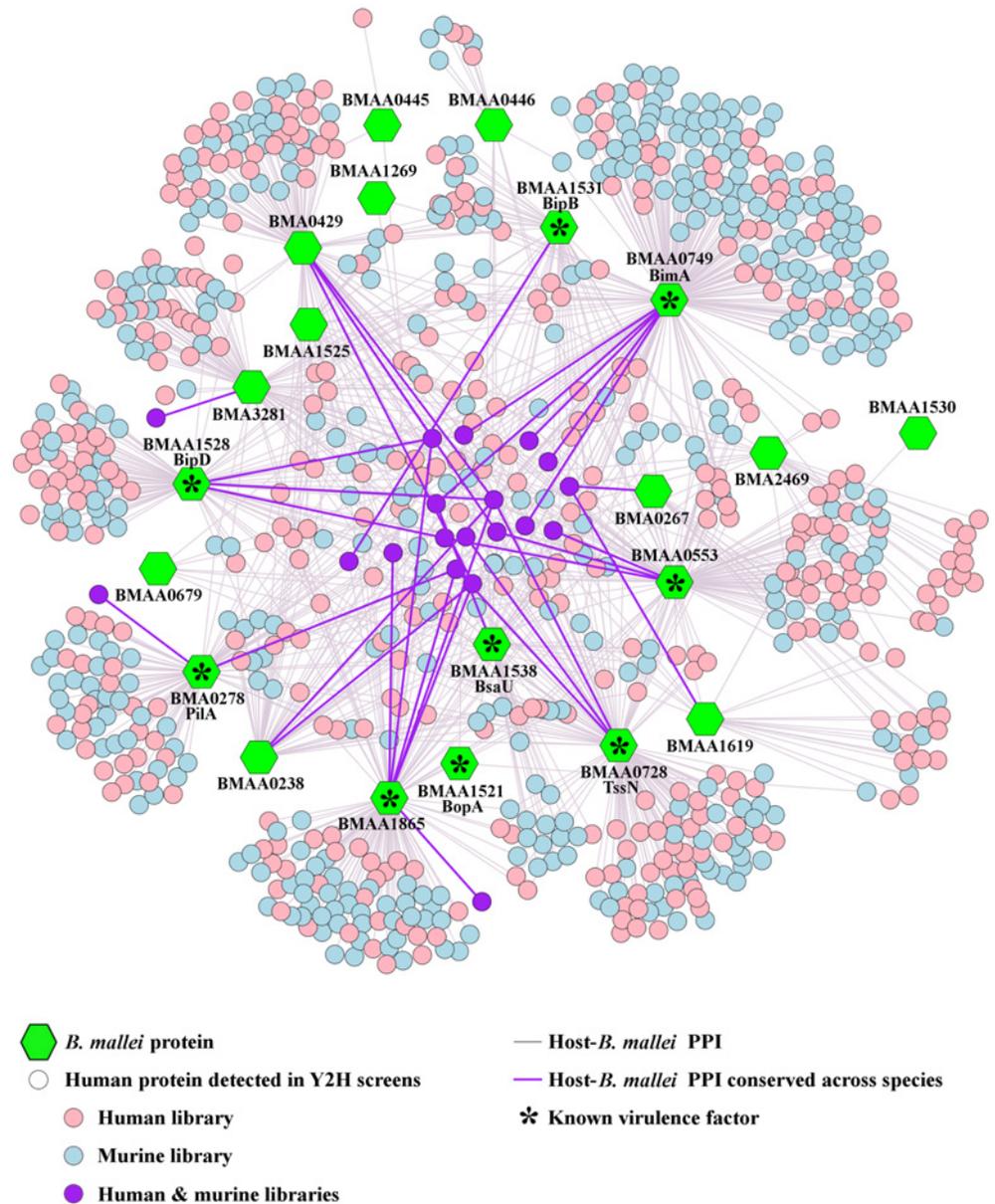


Fig 1. *B. mallei*-human host-pathogen protein interactions. The set of human-*B. mallei* protein-protein interactions (PPIs) was created by merging human-*B. mallei* and orthologous murine-*B. mallei* interaction data. The set consists of 1,235 unique interactions (gray and purple lines) between 21 *B. mallei* (green hexagons) and 828 human proteins (pink, blue, and purple circles). Known virulence factors are also indicated in the graph.

doi:10.1371/journal.pcbi.1004088.g001

interactions (72% or 890 interactions) involved nine known *B. mallei* virulence factors: PilA, BimA, BopA, BipD, BipB, BsaU, BMAA1865, TssN, and BMAA0553 (Table 1) [22]. These nine *B. mallei* virulence factors interacted with 663 human proteins (80% of all identified host proteins), implying that the captured data were largely reflective of host-pathogen virulence mechanisms. We start our analysis by assessing the characteristics of the host proteins targeted by these virulence factors.

Table 1. Known *B. mallei* virulence factors that have been shown to attenuate the disease in animal models.

Protein	Secretion System	Description	Role in Virulence	Animal model	Ref.
BMA0278 (PilA)	2	Type IV pilin	Bacterial adhesion	BALB/c	[62]
BMAA0749 (BimA)*	5	Trimeric autotransporter adhesin	Actin-based motility and intracellular/intercellular spread	Murine	[4]
BMAA1521 (BopA)*	3	Secreted effector/translocon	Intracellular survival	BALB/c	[63]
BMAA1528 (BipD)*	3	Transcriptional regulation; Needle tip protein	Invasion of non-phagocytic cells	BALB/c, C57BL/6	[64]
BMAA1531 (BipB)*	3	Secreted effector/translocon	Multi-nucleated giant cell formation	BALB/c	[65]
BMAA1538 (BsaU)*	3	Secretion apparatus; Needle-length control	Bacterial escape from endocytic vesicles	BALB/c	[66]
BMAA1865	3	N/A	Modulation of host ubiquitination; Phagosome escape	BALB/c	[8]
BMAA0728 (TssN)	6	N/A	Interference with host actin cytoskeleton rearrangement	BALB/c	[8]
BMAA0553	2	Ser/Thr protein phosphatase	Interference with host actin cytoskeleton rearrangement	BALB/c	[8]

*Proteins that have been experimentally linked to a particular secretion system.

doi:10.1371/journal.pcbi.1004088.t001

Characteristics of host proteins interacting with known *B. mallei* virulence factors

B. mallei virulence factors are associated with multiple pathogenic mechanisms of action (Table 1) [3–7], but their direct molecular interactions are not well delineated. First, we applied functional enrichment analyses based on Gene Ontology (GO) annotation data [23] to assess the characteristics of the human proteins targeted by the nine virulence factors. Table 2 shows that these virulence factors interacted with a statistically significant number of human proteins that were associated with 1) protein ubiquitination and ubiquitin ligase activity, 2) vesicle organization, and 3) protein complexes located in the cytoskeleton, in lysosomes, and in the nuclear lumen. These results were consistent with the experimentally observed pathogen interference with host cytoskeleton organization and ubiquitination levels [2, 3, 19–21, 24].

Next, we examined the gross topological properties of the network of interactions formed by the *B. mallei*-targeted host proteins and their interacting host partners, regardless of whether these proteins did or did not interact with the *B. mallei* proteins. We mapped the identified host proteins interacting with *B. mallei* onto a human PPI network [25] consisting of 76,043 physical PPIs among 11,688 proteins. Of the 663 human proteins interacting with the nine *B. mallei* virulence factors, approximately 75% (498) were present in our human PPI network. This set contained proteins that had, on average, a significantly larger number of interacting partners per protein (19.5 vs. 13.0) than would be expected from a corresponding random selection of proteins from the entire human PPI network (Table 3). Among the highest-interacting host proteins targeted by the virulence factors, we found the adapter protein YWHAG (14-3-3 protein gamma) with 376 interactions. This protein, an interacting partner of BimA, has been implicated in the regulation of a large spectrum of signaling pathways [26]. Further topological measures associated with the set of 498 proteins, such as their clustering coefficient (a measure of interactions among nearest neighbors), were not different from the random selection (Table 3). We observed small effects on the length of the shortest path between any two proteins in the set, but it was unclear how to associate these topological parameters with *B. mallei* virulence.

Table 2. Enrichment of Gene Ontology (GO) terms for human proteins interacting with *B. mallei* virulence factors.

Type	Term		Number of proteins	p-value	
	ID	Description		Original	FDR
GO Biological Processes	GO:0043161	Proteasomal ubiquitin-dependent protein catabolic process	21	4.8·10 ⁻⁶	0.00
	GO:0006457	Protein folding	18	1.2·10 ⁻⁴	0.03
	GO:0000209	Protein polyubiquitination	14	2.0·10 ⁻⁴	0.04
	GO:0016050	Vesicle organization	11	2.5·10 ⁻⁴	0.05
GO Molecular Functions	GO:0004842	Ubiquitin-protein ligase activity	18	3.1·10 ⁻⁴	0.03
	GO:0003729	mRNA binding	10	5.1·10 ⁻⁴	0.05
	GO:0031072	Heat shock protein binding	10	6.3·10 ⁻⁴	0.05
GO Cellular Localizations	GO:0000151	Ubiquitin ligase complex	13	3.7·10 ⁻⁴	0.01
	GO:0015629	Actin cytoskeleton	24	4.8·10 ⁻⁴	0.01
	GO:0030529	Ribonucleoprotein complex	30	7.8·10 ⁻⁴	0.02
	GO:0043220	Schmidt-Lanterman incisure	3	8.5·10 ⁻⁴	0.02
	GO:0034663	Endoplasmic reticulum chaperone complex	2	1.3·10 ⁻³	0.03
	GO:0030135	Coated vesicle	19	1.6·10 ⁻³	0.03
	GO:0031371	Ubiquitin conjugating enzyme complex	3	2.3·10 ⁻³	0.04
	GO:0005764	Lysosome	19	2.5·10 ⁻³	0.04
	GO:0032838	Cell projection cytoplasm	4	2.7·10 ⁻³	0.04
	GO:0030131	Clathrin adaptor complex	5	3.4·10 ⁻³	0.05
	GO:0000803	Sex chromosome	4	3.4·10 ⁻³	0.05
	GO:0070971	Endoplasmic reticulum exit site	2	3.8·10 ⁻³	0.05
	GO:0031981	Nuclear lumen	80	4.3·10 ⁻³	0.05
	GO:0030128	Clathrin coat of endocytic vesicle	3	4.6·10 ⁻³	0.05

FDR: False discovery rate calculated using Benjamini and Hochberg multiple test correction [45].

doi:10.1371/journal.pcbi.1004088.t002

Table 3. Topological properties of human proteins interacting with *B. mallei*. We evaluated the following properties of the host proteins that interacted with *B. mallei* proteins based on the human protein-protein interaction (PPI) network [25]: the number of these host proteins in the human PPI network (N_p); the average number of interacting partners (in the human PPI network) of each host protein (D); the clustering coefficient, i.e., the number of interactions among the nearest neighbors (C); the average shortest path between any two proteins in the set (SP); the average number of interacting partners in the human PPI network where both partners interact with *B. mallei* proteins (D_i); and the number of host proteins in the largest connected component (N_p^{LCC}). The top three rows show the results for the host proteins present in the PPI that interacted with the nine known virulence factors, whereas the three lower rows correspond to host proteins that interacted with all 21 tested *B. mallei* proteins from the yeast two-hybrid screening (known and putative virulence factors). The results for the randomly selected (498 or 619) human proteins from the entire human PPI network (All PPIs) were generated through 10³ random repetitions to create averages and standard deviations. The indicated p-values correspond to the probability of the observed properties being different from the randomly selected set from all PPIs.

		N_p	D (SD)	C (SD)	SP (SD)	D_i (SD)	N_p^{LCC} (SD)
Known virulence factors	PPIs	498	19.5	0.15	3.41	0.65	202
	All PPIs	498	13.0 (1)	0.17 (0.01)	3.70 (0.04)	0.28 (0.05)	80 (29)
	p-value	-	8.8·10 ⁻⁸	0.33	2.0·10 ⁻¹¹	1.6·10 ⁻⁴¹	3.3·10 ⁻⁵
Known and putative virulence factors	PPIs	619	19.5	0.15	3.40	0.85	284
	All PPIs	619	13.0 (1)	0.17 (0.01)	3.70 (0.04)	0.35 (0.06)	136 (35)
	p-value	-	4.3·10 ⁻¹⁰	0.27	1.0·10 ⁻¹⁴	1.9·10 ⁻²⁴	2.2·10 ⁻⁵

SD: standard deviation.

doi:10.1371/journal.pcbi.1004088.t003

Next, we examined human protein interactions where *both* proteins individually interacted with one or more *B. mallei* proteins. For the 498-protein set, we found 202 unique proteins that participated in 325 human protein-protein interactions. In comparison with randomly selected proteins, the *B. mallei*-targeted proteins were engaged in a significantly larger number of these interactions (0.65 vs. 0.28 on a per-protein basis). A further examination of sets of connected human proteins that also interacted with the virulence factors, revealed the presence of a single, large connected component, i.e., a sub-network in which a path connects any two proteins to each other. This *largest connected component* was composed of 202 proteins and contained the majority (95%) of the 325 interactions between the human proteins interacting with *B. mallei* (Table 3). The other 11 connected components consisted of five or fewer proteins, an observation that was not statistically significant from a random selection of proteins (data not shown). We found a five-fold increase, from 0.28 to 1.53 ($0.95 \times 325/202$), in the number of human PPIs for each protein in the largest connected component that were all targeted by the virulence factors, compared to a random selection of proteins. These results suggest that a property of the *B. mallei* virulence phenotype is to target well-connected host proteins in a unique set of interconnected host proteins. Next, we used this property to expand on our initial set of GO annotations to better characterize *B. mallei* infectivity and pathogenesis.

B. mallei virulence factors target interactions among host proteins

The analysis of the interactions between the virulence factors and host proteins showed that the targeted human proteins were highly likely to interact among themselves. We hypothesized that interactions among these host proteins are equally important targets as the proteins themselves and could be used to shed light on how virulence factors exert their influence. As detailed in Materials and Methods, we used the largest connected component identified above to detect 93 sets of human PPIs in which, in each set, all human proteins interacted with at least one of the nine known *B. mallei* virulence factors and had the same GO biological process annotations; we denoted these sets *interaction modules*.

Table 4 shows that these interaction modules were associated with biological processes related to ligase activity, ubiquitination, protein modification, transcription and translation, immune response, signaling, cytoskeleton organization, development, and mRNA processing. Overall, the identified biological processes were similar to the ones identified when interactions among host proteins were not taken into account; however, they provided an improved annotation granularity. For example, the interaction modules allowed us to identify a biological process termed “positive regulation of protein ubiquitination” instead of just “protein ubiquitination.” Importantly, the analysis provided evidence of a much larger effort to target intracellular host signaling processes, in particular those related to the immune response. Fig. 2 shows the subset of 116 proteins and 163 interactions from the largest connected component that were part of the 93 identified interaction modules and the location of six interaction modules. Each of the interaction modules constituting ubiquitination and ligase activity, transcriptional regulation, immune response, cytoskeleton organization, and mRNA processing, consisted of proteins and interactions that were closely grouped together in the largest connected component (Fig. 2A–E). Fig. 2 also shows that some human proteins are a part of multiple interaction modules, suggesting that *B. mallei* interacts with multifunctional or “moonlighting” host proteins [27]. Multifunctional proteins have been associated with such neurological disorders as Alzheimer’s and Parkinson’s diseases [28], as well as with bacterial virulence in *Helicobacter pylori*, *Mycobacterium tuberculosis*, and *Streptococcus pneumoniae* [29]. Given the multifaceted role of these proteins in enzymatic catalysis, signal transduction, transcriptional regulation, apoptosis, motility, and growth [30, 31], interactions with them suggest an avenue for *B. mallei* to simultaneously interfere with multiple

Table 4. Enrichment of Gene Ontology (GO) biological processes in host subnetworks. LCC represents the number of proteins in the largest connected component annotated with a given term; LIM represents the number of proteins in the largest interaction module for a given term; p_{GO} denotes the probability of the same number of proteins as the LCC being annotated with a given GO term solely through a random selection; p_{Rp} denotes the probability that a given number of proteins as the LIM are annotated with a given GO term solely through random selection; p_{Rn} represents the probability that a given number of proteins as the LIM are annotated with a given GO term solely through random selection in a random network that has the same degree distribution as our human network. All p -values were assessed using the Benjamini-Hochberg method to meet a maximum false discovery rate threshold of 5% [45]. The table contains only the lowest-level GO terms; the complete data are available in [S1 Table](#).

Category	Term		Size		p-value		
	ID	Description	LCC	LIM	p_{GO}	p_{Rp}	p_{Rn}
Protein modification	GO:0051443	Positive regulation of ubiquitin-protein ligase activity	5	3	$3.6 \cdot 10^{-3}$	$8.2 \cdot 10^{-3}$	$4.5 \cdot 10^{-5}$
	GO:0051351	Positive regulation of ligase activity	5	3	$4.1 \cdot 10^{-3}$	$8.2 \cdot 10^{-3}$	$4.5 \cdot 10^{-5}$
	GO:0000209	Protein polyubiquitination	10	8	0.0	0.0	0.0
	GO:0016574	Histone ubiquitination	3	3	$5.8 \cdot 10^{-3}$	$6.0 \cdot 10^{-4}$	$3.7 \cdot 10^{-7}$
	GO:0051248	Negative regulation of protein metabolic process	15	10	$7.0 \cdot 10^{-4}$	$1.0 \cdot 10^{-4}$	$2.0 \cdot 10^{-4}$
Cell cycle	GO:0051320	S phase	6	3	$1.0 \cdot 10^{-2}$	$1.0 \cdot 10^{-2}$	$1.7 \cdot 10^{-4}$
	GO:0051437	Positive regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	4	3	$1.0 \cdot 10^{-2}$	$7.4 \cdot 10^{-3}$	$2.4 \cdot 10^{-5}$
mRNA processing	GO:0051028	mRNA transport	5	3	$9.5 \cdot 10^{-3}$	$3.2 \cdot 10^{-3}$	$6.4 \cdot 10^{-5}$
	GO:0000398	mRNA splicing, via spliceosome	8	6	$7.3 \cdot 10^{-3}$	$1.0 \cdot 10^{-4}$	$1.0 \cdot 10^{-6}$
Transcription and translation	GO:0006355	Regulation of transcription, DNA-dependent	45	15	$7.4 \cdot 10^{-3}$	$6.0 \cdot 10^{-4}$	$4.3 \cdot 10^{-4}$
	GO:0006413	Translational initiation	8	5	$9.0 \cdot 10^{-4}$	0.0	0.0
Signaling and immune response	GO:0007154	Cell communication	71	44	$5.7 \cdot 10^{-3}$	0.0	$4.4 \cdot 10^{-5}$
	GO:0023052	Signaling	71	44	$2.7 \cdot 10^{-3}$	0.0	$4.4 \cdot 10^{-5}$
	GO:0035556	Intracellular signal transduction	34	17	$7.3 \cdot 10^{-3}$	0.0	$1.1 \cdot 10^{-4}$
	GO:0007167	Enzyme linked receptor protein signaling pathway	21	8	$1.6 \cdot 10^{-3}$	$1.6 \cdot 10^{-3}$	$6.8 \cdot 10^{-4}$
	GO:0050852	T cell receptor signaling pathway	7	4	$4.0 \cdot 10^{-4}$	$1.1 \cdot 10^{-3}$	$3.3 \cdot 10^{-4}$
	GO:0016032	Viral reproduction	13	4	$2.0 \cdot 10^{-3}$	$5.0 \cdot 10^{-3}$	$2.1 \cdot 10^{-2}$
	GO:0050688	Regulation of defense response to virus	5	3	$1.8 \cdot 10^{-3}$	$7.0 \cdot 10^{-4}$	$2.0 \cdot 10^{-4}$
	GO:0019221	Cytokine-mediated signaling pathway	12	3	$1.8 \cdot 10^{-3}$	$1.0 \cdot 10^{-2}$	$6.3 \cdot 10^{-4}$
Development	GO:0048731	System development	52	28	$1.0 \cdot 10^{-2}$	0.0	$6.1 \cdot 10^{-4}$
	GO:0048812	Neuron projection morphogenesis	14	5	$1.0 \cdot 10^{-2}$	$1.4 \cdot 10^{-4}$	$4.7 \cdot 10^{-4}$
Other	GO:0016311	Dephosphorylation	9	4	$1.6 \cdot 10^{-3}$	0.0	$2.0 \cdot 10^{-5}$
	GO:0006457	Protein folding	8	3	$3.6 \cdot 10^{-3}$	$3.7 \cdot 10^{-3}$	$1.6 \cdot 10^{-5}$
	GO:0016192	Vesicle-mediated transport	20	5	$1.0 \cdot 10^{-2}$	$3.8 \cdot 10^{-3}$	$3.0 \cdot 10^{-5}$
	GO:0007010	Cytoskeleton organization	20	9	$3.1 \cdot 10^{-3}$	0.0	$4.8 \cdot 10^{-6}$

doi:10.1371/journal.pcbi.1004088.t004

host-cellular processes to facilitate invasion and survival. In particular, [Fig. 2F](#) shows the largest interaction module associated with biological processes linked to multifunctional proteins. This interaction module contained 54 interactions among 44 human proteins associated with various types of regulation (regulation of gene expression, cytokinesis, or apoptosis), signal transduction (GTPase mediated signal transduction and Janus kinase/signal transduction), and response triggering (immune response and response to stress). Additionally, this module contained host-interacting partners of eight out of the nine *B. mallei* virulence factors from our set, lacking only BopA. These results suggest that *B. mallei* virulence factors target multifunctional host proteins to simultaneously interfere with multiple host processes required for normal cellular function.

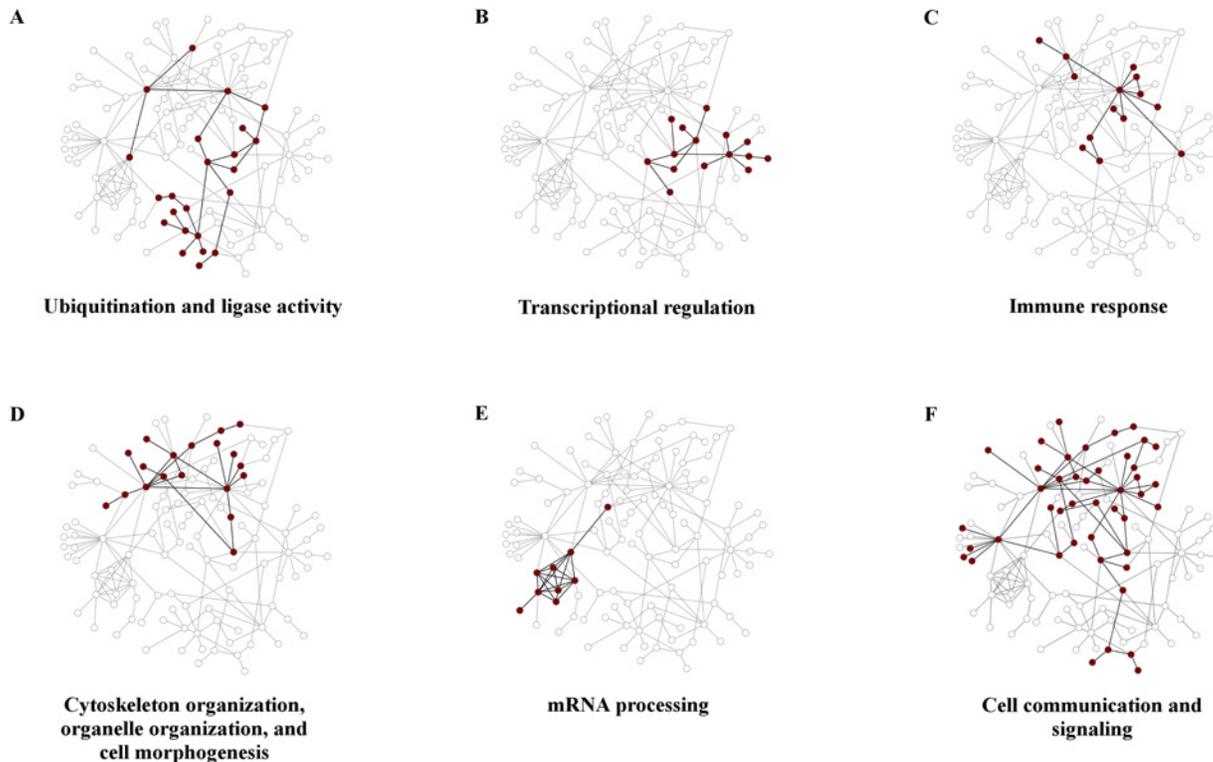


Fig 2. Clustering of human proteins targeted by *B. mallei* known virulence factors. The graphic shows 116 proteins of the largest connected component of the human PPI network that belong to one or more statistically significant interaction modules. Note that each of these human proteins also interacted with one or more known *B. mallei* virulence factors. As exemplified by the annotated interaction modules in A-F, the known virulence factors targeted human proteins that were highly interacting among themselves and belonged to the same biological process. For a list of host proteins that compose each interaction module, see [S2 Table](#).

doi:10.1371/journal.pcbi.1004088.g002

Putative *B. mallei* virulence factors improve characterization of *B. mallei* targets

Given that our host-*B. mallei* interaction dataset contained a number of putative virulence factors, we also evaluated the effect of adding these virulence factors into our analysis to characterize host targets. Similarly to the above analyses, we first evaluated the prevalent characteristics of human proteins using GO annotation [23]. The identified molecular annotations largely matched those identified for known virulence factors only, but also included additional GO terms, such as terms related to RNA metabolic processes (S3 Table). Table 3 shows that the analysis of topological properties of host proteins interacting with known and putative virulence factors displayed the same trends observed in the analysis of the interacting partners of known virulence factors. Next, we evaluated the extent to which known and putative virulence factors also targeted connected subsets of host PPIs. We identified 75 statistically significant interaction modules whose GO biological process annotations largely overlapped with the ones identified for interacting partners of known virulence factors only. Although the number of statistically significant interaction modules was smaller than above (an increase in the number of host proteins dilutes the enrichment), the addition of new host proteins increased the size (in terms of proteins and interactions) of previously identified interaction modules (S4 Table). This suggests that with the increase of protein annotation or with the identification of additional host-*B. mallei* PPIs, we will be able to identify larger and more complete host interaction modules targeted by *B. mallei* virulence factors.

Consequently, we used all the interactions shown in Fig. 1 to identify biological pathways targeted by all tested *B. mallei* proteins using the Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation database [32]. We identified two statistically significantly enriched host pathways: bacterial invasion of epithelial cells and focal adhesion. Fig. 3 shows that the proteins targeted in the focal adhesion pathway appeared to be coordinated for pathway activation and largely interacted with each other (yellow boxes). The majority of these molecular interactions belonged to a connected sub-pathway located at the beginning of the pathway (the probability of observing such connectivity at random is $< 10^{-6}$), and they provided a link between membrane receptors and signaling events that led to reorganization of the actin cytoskeleton.

Human-*B. mallei* interactions and their effect on the crosstalk between different biological processes

One of the most prominently recurring results across all of our analyses was the link between *B. mallei* pathogenicity and host cytoskeleton organization. It has been shown that a number of bacterial pathogens, including *Yersinia*, *Salmonella*, *Shigella*, *Listeria*, and *Burkholderia*, interfere with host signaling pathways to stimulate the host's cytoskeleton rearrangement [2, 33]. These changes in signaling lead to changes in the host-cell shape and facilitate bacterial internalization and cell-to-cell spread [33]. Fig. 4 shows host proteins that interacted with known and putative *B. mallei* virulence factors that can be directly associated with cytoskeleton organization. The largest statistically significant interaction module, represented by red stars,

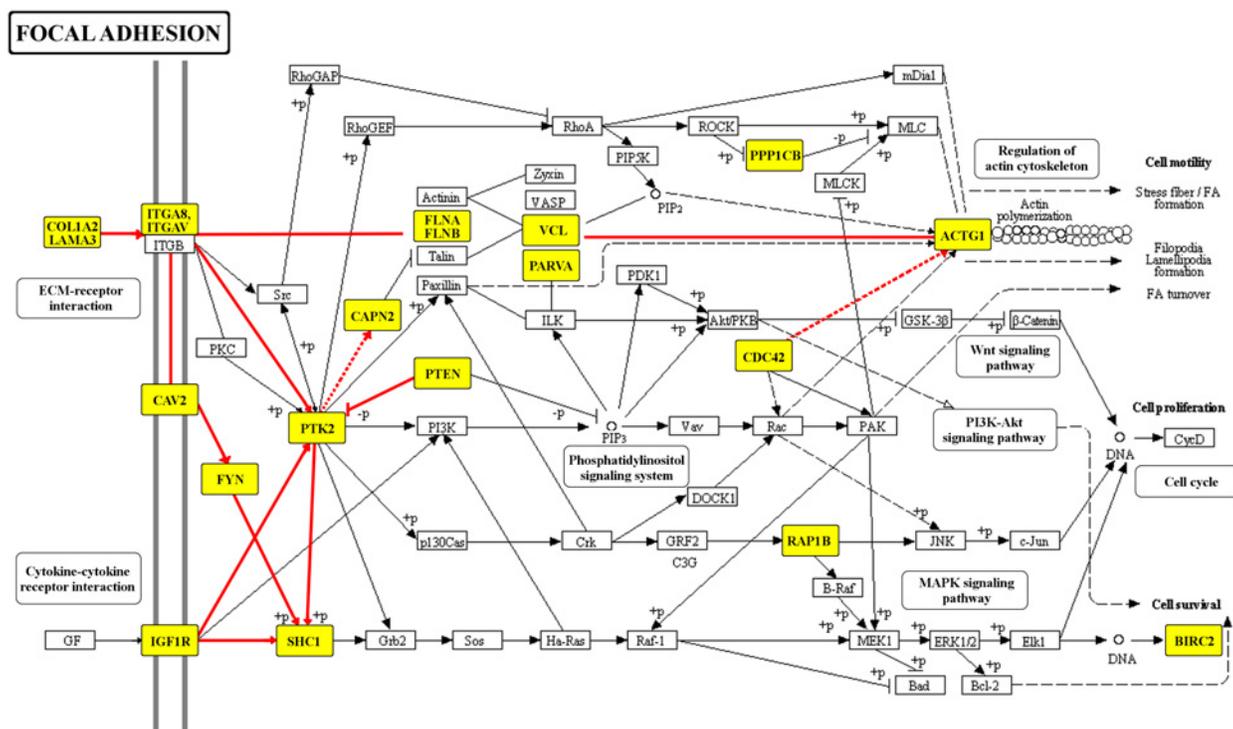


Fig 3. Focal adhesion pathway as a virulence factor target. We identified the Kyoto Encyclopedia of Genes and Genomes (KEGG) focal adhesion pathway as enriched with multiple virulence factor targets. The majority (17 of 20) of host proteins interacting with *B. mallei* virulence factors in this pathway belong to a connected sub-pathway of proteins (yellow boxes and red lines), mainly grouped at the beginning of the pathway. This observation implied that receptors and signaling molecules were likely *B. mallei* targets, corroborating previous observations that pathogens tend to interfere with host processes related to cell communication and actin cytoskeleton organization. The pathway diagram for the focal adhesion pathway was adapted from the KEGG pathway map [32] with permission from the KEGG database administrators.

doi:10.1371/journal.pcbi.1004088.g003

contained proteins previously identified as bacterial targets vital for host actin cytoskeleton re-arrangement, e.g., membrane-associated small GTPases (CDC42 and RALA), Filamin-A (FLNA), and Rho GDP-dissociation inhibitor (ARHGDI1) [2, 33]. The remaining cytoskeleton-related host proteins, represented as dark red circles, participated in smaller cytoskeleton organization interaction modules that were, on average, less than two proteins (three interactions or edges) away from the largest module.

Given that connector proteins (represented as white and yellow circles in Fig. 4) between the cytoskeleton organization interaction modules were annotated with biological processes different from cytoskeleton organization, as well as the multifunctional nature of some cytoskeletal reorganization proteins, we examined the occurrence of shared proteins that interacted with multiple pathways, i.e., pathway crosstalk. Initially, out of all human proteins interacting with the examined *B. mallei* proteins, we evaluated the relationships among proteins involved in the focal adhesion pathway that participated in or interacted with human proteins associated with cytoskeleton organization. The shaded area in Fig. 4 shows that six of 12 proteins were present in both systems. Fig. 5 shows an extension of this analysis that includes *B. mallei* interacting host proteins that are components of eight other molecular pathways that shared proteins with the focal adhesion pathway. Fig. 5A (left) shows that among these nine pathways, the number of pathways that shared one or more proteins was low. However, the number of PPIs connecting the proteins from one pathway with proteins from another pathway was markedly higher [Fig. 5A (right)]. The large number of signaling pathways affected via the

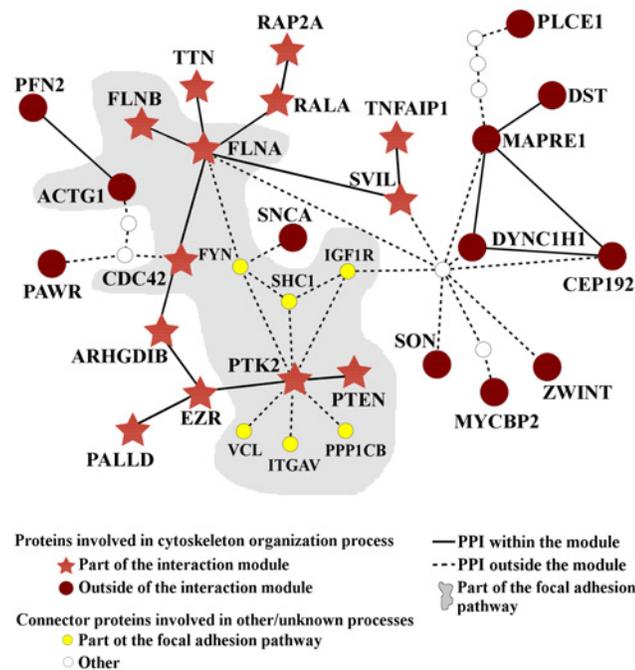


Fig 4. Actin cytoskeleton organization as a virulence factor target. Human proteins targeted by *B. mallei* proteins formed an interaction module that was primarily linked to cytoskeleton organization and focal adhesion. Twenty-five of these proteins were involved in cytoskeleton organization processes; 13 of them (red stars) interacted with each other (forming an interaction module), and the remaining 12 proteins (dark red circles) were on average < 2 nodes (< 3 edges) away from the interaction module. The figure also shows the overlap between the cytoskeleton organization interaction module and the focal adhesion pathway (shaded area), where connecting protein interactions appear as smaller circles and dashed lines. Note that all human proteins shown interacted with one or more *B. mallei* proteins.

doi:10.1371/journal.pcbi.1004088.g004

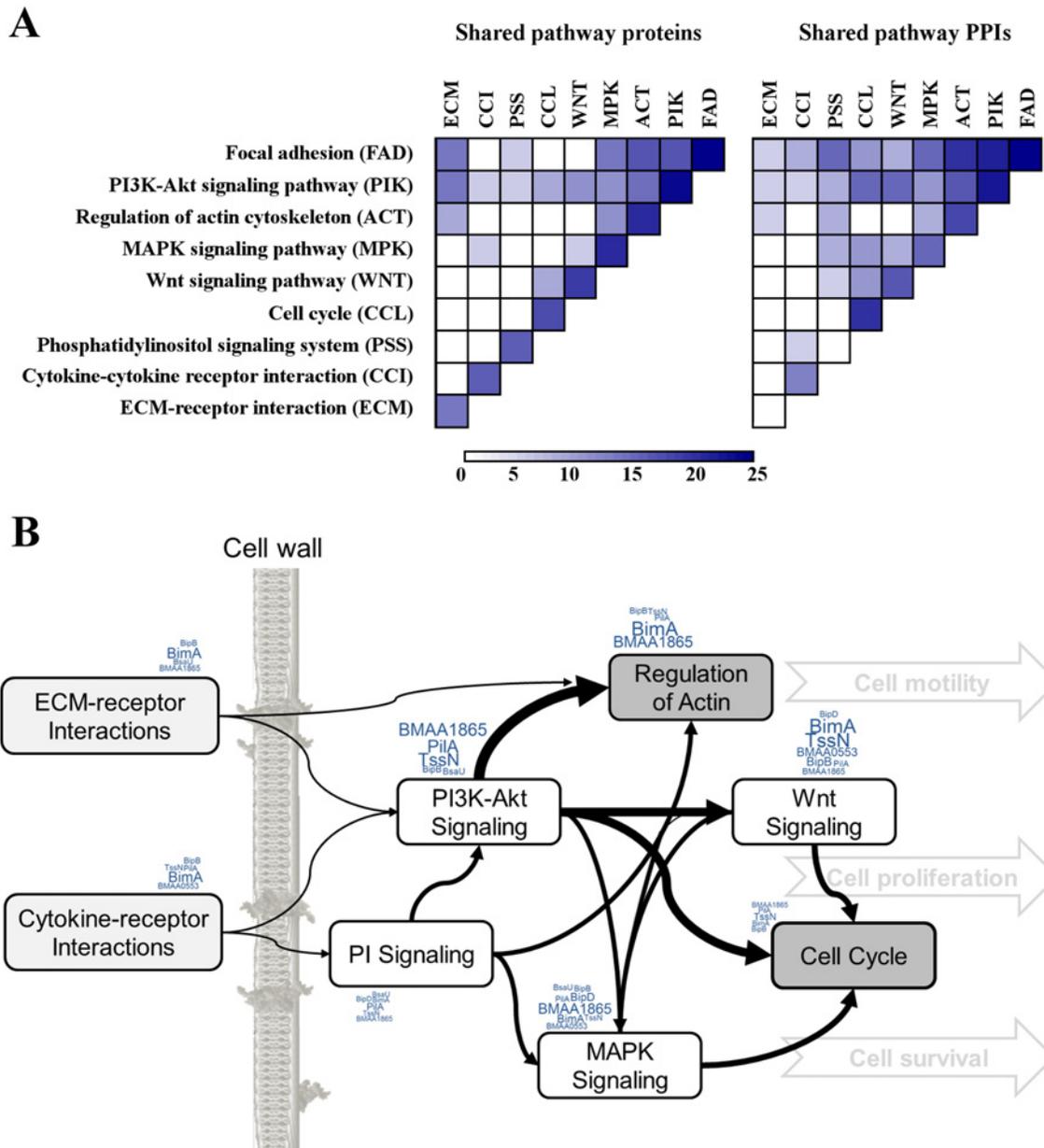


Fig 5. Crosstalk between host pathways targeted by *B. mallei* virulence factors. A) The number of shared proteins and shared pathway protein-protein interactions (PPIs) among human proteins interacting with *B. mallei* that appeared in the focal adhesion pathway and in up to eight other molecular pathways that shared proteins (partially overlapped) with this pathway. The number of shared proteins across pathways was smaller than the number of shared pathway PPIs. B) The location and number of crosstalk interactions affected by *B. mallei* centered around the focal adhesion pathway and appear as arrows with line thicknesses proportional to the number of shared PPIs. The identity and number of virulence factors that target each pathway are illustrated using a word cloud. By preferentially targeting signaling pathways, the effect of one protein modulated through interaction with a virulence factor could propagate and disproportionately influence a larger number of biological processes.

doi:10.1371/journal.pcbi.1004088.g005

focal adhesion pathway magnified the effects of these cross-pathway interactions. Fig 5B illustrates the propagation and number of cross-pathway interactions that were mediated via the focal adhesion pathway and shows the known virulence factors (Table 1) that can be associated with each pathway. Thus, virulence factors affected biological processes and molecular pathways associated with multiple interconnecting host processes, providing an explanation of how

interference with the function of a single protein propagated to and influenced multiple host processes and pathways.

Using multiple host-pathogen interaction networks to predict the role of pathogen proteins

Our statistical analyses show that the aggregated host-pathogen interaction data could identify host molecular mechanisms targeted by *B. mallei*. However, detecting specific mechanisms of action for *each* pathogen protein based on enrichment analysis of large-scale Y2H protein interaction data is not trivial. This partly stems from experimental, biological, and statistical considerations. For example, the Y2H methodology is biased for certain types of interactions in a non-native environment [34], binding events may or may not be biologically relevant, and statistical testing is hampered by small effect sizes and small statistical power. Conversely, this and previous studies have shown that multiple pathogens tend to target the same host proteins, biological processes, and pathways [2, 9, 10, 18]. Hence, one could potentially use common pathogenic mechanisms to more robustly characterize bacterial proteins and their host targets.

We explored a focused set of human-pathogen interactions derived from putative virulence factor proteins from *S. enterica* and *Y. pestis*; the majority of these proteins are associated with a Type 3 Secretion System (T3SS). These datasets contained 62 host-pathogen interactions for 21 *S. enterica* proteins [12] and 223 interactions for 69 *Y. pestis* proteins [13]. An initial orthology-based approach to retrieve annotations proved too restrictive and did not generate any novel insights into *B. mallei* virulence. Instead, we used an alternative network alignment-based methodology optimized for inter-species alignment, i.e., we differentiated between host and pathogen proteins and avoided mapping host proteins to pathogen proteins and vice versa. As detailed in Materials and Methods, we introduced a novel alignment algorithm (HPIA) designed specifically for the alignment of cross-species interactions. We used the HPIA algorithm to identify similarities between host-pathogen interactions using the *B. mallei*, *S. enterica*, and *Y. pestis* PPI datasets, based on a combined similarity measure that included topological similarity, sequence similarity, and functional similarity. Table 5 lists the *B. mallei* proteins, their aligned protein partners, and inferred function(s) derived from the alignment.

Table 5 shows *B. mallei* proteins with a known/assumed function in pathogenicity and the corresponding functionality, predicted based on the function of their aligned *Y. pestis* and *S. enterica* partners. We identified similarities between T3SS proteins involved in bacterial internalization for all three pathogens, including the orthologs BipB-SipB and BipC-SipC. Additionally, the *B. mallei* PilA protein was aligned to *Y. pestis* fimbrial protein FimA6, another cell adhesion protein. Furthermore, although the *S. enterica* and *Y. pestis* interaction datasets included mainly T3SS proteins, the two *B. mallei* Vgr proteins associated with bacterial survival and replication via Type 6 Secretion Systems (T6SS) were aligned to *S. enterica* and *Y. pestis* proteins also known to promote bacterial survival and replication. Thus, while the aligned proteins may have different roles within each pathogen, it is possible that they interact with a similar type of host proteins, causing the alignment algorithm to capture these similarities. Overall, the alignment-based inferred roles for six out of the 11 (55%) annotated *B. mallei* proteins matched their existing annotation and their corresponding secretion system assignment (Table 5, fourth and second columns, respectively). If we only consider matching functionality and ignore the Vgr association to T6SS, the inferred functions matched the existing annotation in eight of 11 (73%) cases.

For the *B. mallei* virulence factors without known functions in pathogenicity, listed in the lower part of Table 5, the functional mappings from *S. enterica* and *Y. pestis* provided an indication of their mechanistic role in virulence. Of special importance were the three

Table 5. Putative role of *B. mallei* proteins inferred from host-pathogen interaction network alignment results.

<i>B. mallei</i> proteins				Proteins aligned to a given <i>B. mallei</i> protein		Putative (alignment-predicted) role of a given <i>B. mallei</i> protein
Locus tag (name)	Secretion system association	Description	Known role	<i>S. enterica</i>	<i>Y. pestis</i>	
BMA0278 (PilA) [†]	Type 2	Type IV pilin	Cell adhesion	AvrA	FimA6	- Cell adhesion - Promotion of bacterial survival
BMAA0445 (VgrG) ^{‡, *}	Type 6	Rhs element Vgr protein	Promoting bacterial survival and replication	SifA	YPO2940	- Cell adhesion - Promotion of bacterial survival
BMAA0446 (VgrG) ^{‡, *}	Type 6	Rhs element Vgr protein	Promoting bacterial survival and replication	SseG	PilF	- Replication niche establishment
BMAA0749 (BimA) [*]	Type 5	Hemagglutinin domain protein	Actin based motility	SopB	OmpA	- Promotion of bacterial survival - Bacterial internalization
BMAA1269 (VgrG)	Type 6	Rhs element Vgr protein	Promoting bacterial survival and replication	SseJ	LcrD	- Regulation of T3SS secretion - Promotion of bacterial survival
BMAA1521 (BopA) ^{†, *}	Type 3	Effector protein	Bacterial internalization and promoting bacterial survival	SptP	YscK	- Bacterial internalization
BMAA1528 (BipD) ^{†, *}	Type 3	Translocator protein	Bacterial internalization	SipC	YopT	- Bacterial internalization - Interference with host cytoskeleton
BMAA1530 (BipC) ^{†, *}	Type 3	Effector protein	Bacterial internalization	SipC	LcrV	- Regulation of T3SS activation - Bacterial internalization - Interference with host cytoskeleton
BMAA1531 (BipB) ^{†, *}	Type 3	Translocator protein	Bacterial internalization	SipB	YscS	- Bacterial internalization
BMAA1538 (BsaU) ^{†, *}	Type 3	Type 3 secretion protein (needle assembly)	Bacterial internalization	SopE	YscL	- Bacterial internalization - Interference with host cytoskeleton
BMA0429 (Cmk)	Type 3	Cytidylate kinase	Kinase activity; ATP binding	SpiC	YscN	- Regulation of T3SS secretion
BMAA1525 (BapB) [*]	Type 3	Type 3 secretion protein	N/A	SspH2	TyeA	- Regulation of T3SS secretion - Interference with host ubiquitination
BMAA1865	Type 3	Hypothetical protein	N/A	SopE2	YopE	- Bacterial internalization - Interference with host cytoskeleton
BMAA0728 (TssN)	Type 6	Hypothetical protein	N/A	SseL	YpkA	- Interference with host signaling - Interference with host ubiquitination
BMA0267	-	Pseudogene	N/A	SipA	YPO4044	- Bacterial internalization
BMAA0553	Type 2	Ser/Thr protein phosphatase	N/A	SseI	YPO2113	- Promotion of bacterial survival
BMAA0679	-	Chemotaxis protein CheC	N/A	SspH1	YscY	- Bacterial internalization - Interference with host ubiquitination
BMA2469 (Tkt)	Type 3	Transketolase	N/A	SipA	YscX	- Bacterial internalization - Interference with host cytoskeleton
BMA3281 (FliF)	Type 3	Flagellar M-ring protein	N/A	PipB2	YPMT1.42ac	- Promotion of bacterial survival
BMAA0238	Type 2	Hypothetical protein	N/A	SlrP	YopN	- Regulation of T3SS secretion - Interference with host ubiquitination
BMAA1619	Type 3	Hypothetical protein	N/A	SpvB	HofG7	- Promotion of bacterial survival

For detailed information about the inferred roles, see [S5 Table](#).

[†]Proteins that matched their existing annotation and secretion system (if known).

[‡]Proteins that matched their existing annotation but not the secretion system.

^{*}Proteins that have been experimentally linked to a particular secretion system.

doi:10.1371/journal.pcbi.1004088.t005

novel virulence factors we identified from the Y2H data and experimentally verified in a *B. mallei* ATCC 23344 aerosol mouse infection model: BMAA1865, BMAA0728, and BMAA0553. The HPIA algorithm identified similarities between BMAA1865, *S. enterica* protein SopE2, and *Y. pestis* protein YopE, based on their interactions with host proteins involved in actin-cytoskeleton rearrangement processes. The alignment also identified similarities between *B. mallei* protein BMAA0728 and *S. enterica* protein SseL based on their interactions with host proteins involved in ubiquitination. Furthermore, the HPIA algorithm identified similarities between *B. mallei* protein BMAA0553 and *S. enterica* protein SseI associated with regulation of the host cytoskeleton and inhibition of cell motility. These results imply that the hypothetical protein BMAA1865 has a role in the host actin-cytoskeleton manipulation, that BMAA0728 has a role in (de)ubiquitination, and that serine/threonine phosphatase BMAA0553 has a role in cytoskeleton regulation. These are the same roles we previously proposed for these three proteins based on the literature review of pathogenic mechanisms [8].

The alignment also identified a putative role of another protein of interest from our previous study, cytidylate kinase BMA0429. The host-pathogen PPI data linked this protein to multiple processes related to pathogenicity. We were not able to test its pathogenicity in an animal model, because this protein appeared to be essential. However, the alignment results imply that this protein had a role in the regulation of T3SS secretion, as it is mapped to two T3SS regulators that are more likely to be localized in the bacterial cytoplasm than to be translocated into the infected cell: SpiC in *S. enterica* and YscN in *Y. pestis* [35–38].

Multiple *B. mallei* virulence factors target eukaryotic-specific host-cell processes

A key insight into the virulence mechanisms that we could derive from the Y2H interactions was that *B. mallei* targeted eukaryotic-specific cellular mechanisms, such as ubiquitination and focal adhesion. Thus, the specific virulence adaptations retained in the evolution of *B. mallei* as an obligate mammalian pathogen include targeting the ubiquitination degradation/signaling system and using the focal adhesion pathway as a fulcrum for transmitting mechanical forces and regulatory signals. This provides the mechanisms to modulate and adapt the host-cell environment for the successful establishment of host infections.

Based on our analysis of their host protein interactions and targeted pathways, the nine known virulence factors shared many common points of attack on the host cell's physiology. Expanding on the cross-talk analysis shown in Fig. 5B, we created an interconnected host-pathway map. We inferred the connection between two pathways from the number of inter-pathway PPIs where both host proteins are interacting with at least one *B. mallei* protein. The larger the number of such host protein pairs, the larger the potential influence *B. mallei* have on the cross-talk between the involved pathways. Fig. 6 shows the extent of this influence between host pathways and the central role the focal adhesion pathway plays in propagating cell signaling and affecting key host cellular processes relevant to *B. mallei* pathogenesis. The pathways implicated in Fig. 5B directly related to focal adhesion are marked with stars in Fig. 6. These, in turn, are interconnected with a large number of signaling pathways (Fig. 6—grey background) that ultimately control cell cycle, morphology, and growth. A number of known disease processes, marked in red symbols, are also interconnected or directly connected to this signaling network. We hypothesized that this virulence factor host-pathogen network is central in controlling key cellular mechanisms that allow *B. mallei* to adapt the host cell environment and ensure robust infection.

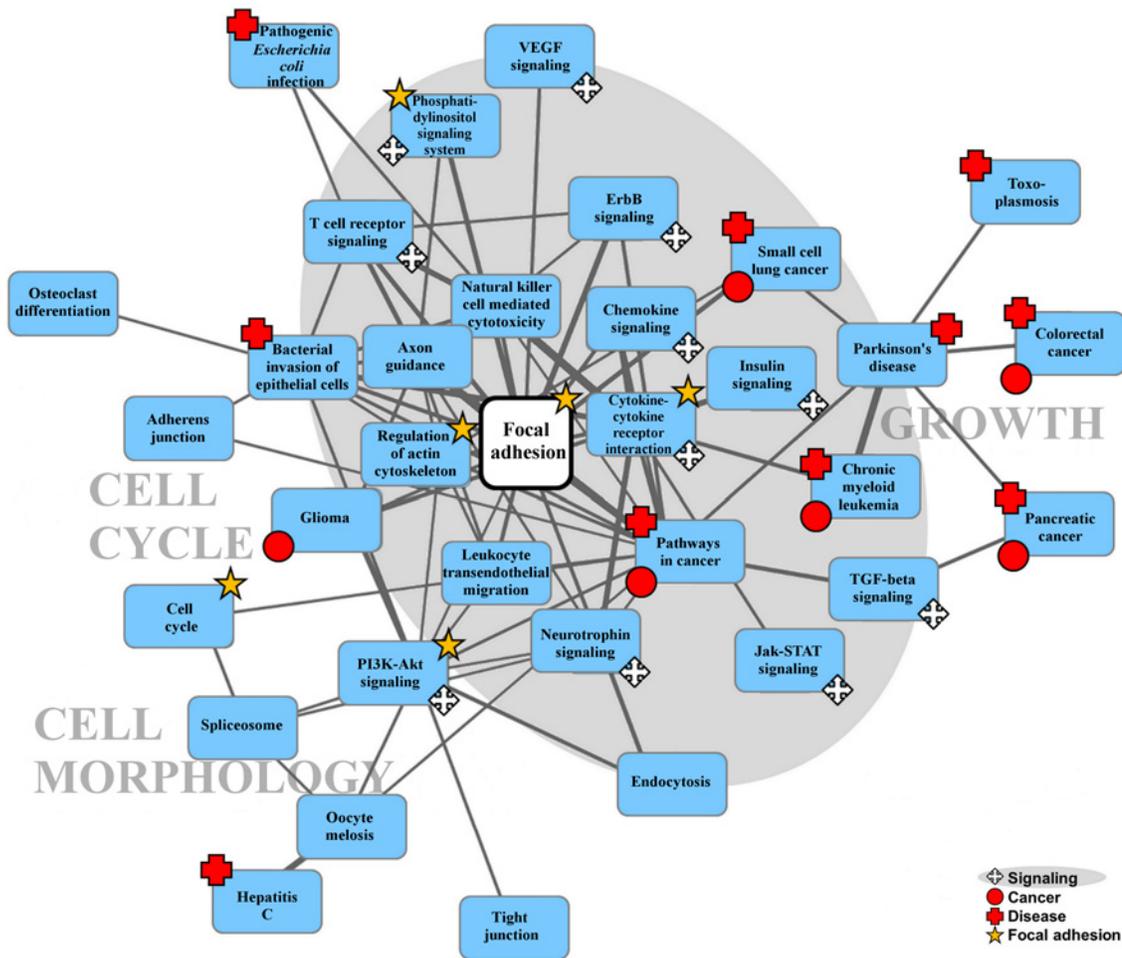


Fig 6. Focal adhesion as a central hub for targeting host cells. The number of shared protein-protein interactions (PPIs) targeted by *B. mallei* virulence factors is shown as lines proportional to the number of PPIs (only connections with 10 or more interactions are illustrated). Physiological, cancer, and disease pathways were all interconnected via signaling pathways that could be affected through the focal adhesion pathway.

doi:10.1371/journal.pcbi.1004088.g006

Summary

Given the association of the selected pathogen proteins to secretion systems, the underlying Y2H methodology, and our analysis methodology, our detection capabilities were geared to finding host pathways and biological processes targeted by *B. mallei* via virulence factors. The limitation of this approach is that while a host-pathogen protein interaction may occur, as determined via Y2H experimentation, this type of data does not allow us to resolve when, where, or why such interactions are important. Furthermore, even though the strict statistical threshold at a false discovery rate (FDR) $\leq 5\%$ minimizes the chances of identifying random data correlations, it does not test our hypothesis that the pathway is involved in *B. mallei* virulence. Conversely, the strength of our analysis is that the identified host interactions are dominated by known and validated virulence factors, allowing us to create new hypotheses around the biological interpretation of pathway interaction patterns.

Our results showed that host-pathogen PPIs represent a rich source of information about molecular mechanisms of pathogenicity, and that these interactions can be used to identify and characterize host molecular pathways and processes targeted by pathogens. Specifically, our

topological analysis of human-*B. mallei* protein interactions showed that known and putative *B. mallei* virulence factors tend to target multifunctional host proteins, host proteins that interact with each other, and host proteins with a large number of interacting partners. Additionally, the analysis identified a number of host processes and pathways relevant to *B. mallei* pathogenicity, many of which have been linked to bacterial pathogenicity in previous experimental studies, e.g., signaling and communication, protein modification and regulation, cytoskeleton organization, and focal adhesion. Furthermore, the topological analysis suggested that *B. mallei* virulence factors target host molecular processes through interference with their direct and indirect host-interacting partners, implying that the process of pathogenic internalization and intracellular survival requires the modulation of multiple host cellular processes.

We further introduced the novel HPIA algorithm that can be used to identify common sets of host-pathogen interactions by aligning (mapping) host-to-host and pathogen-to-pathogen proteins from two interaction datasets. We used the HPIA algorithm to compare human-*B. mallei* interactions to those of human-*Y. pestis* and human-*S. enterica* and identified a statistically significant number of aligned interactions. We also showed that the resulting alignments could be used to predict roles of *B. mallei* proteins based on the roles of their aligned *Y. pestis* and *S. enterica* partners.

Finally, given that nine of 21 proteins in our dataset are known virulence factors, we could hypothesize on why and how *B. mallei* uses these proteins to overcome multiple defense systems and orchestrate a robust infection process in mammalian hosts. Ultimately, the bacterial host-virulence program is derived from a survival strategy developed in the rhizosphere, i.e., in a generally competitive environment containing multiple, diverse species. Using multiple virulence factors to target eukaryotic-specific mechanisms common to eukaryotic rhizosphere species, *B. mallei* broadly influences key processes in ubiquitination and cell signaling to modulate and adapt the host-cell environment for its benefit.

Materials and Methods

Human-*B. mallei* protein interaction set

To create a comprehensive set of human-*B. mallei* PPIs, we merged human-*B. mallei* and murine-*B. mallei* PPI datasets identified in [8]. These datasets contained 586 interactions between 409 human and 21 *B. mallei* proteins, and 797 interactions between 574 murine and 25 *B. mallei* proteins; 19 *B. mallei* proteins appeared in both sets, including nine known *B. mallei* virulence factors (Table 1). When creating the merged set, we considered only a subset of murine-*B. mallei* PPIs in which the *B. mallei* proteins also interacted with human proteins and, thus, had shown the ability to bind to human proteins. The merging procedure consisted of four steps. In the first step, we identified *B. mallei* proteins that interacted with both hosts (19 *B. mallei* proteins). Then, we found human orthologs for each of the 419 (73%) murine proteins that interacted with the *B. mallei* proteins identified in step 1. In the third step, we assessed whether the human orthologs constituted unique proteins, i.e., whether they were not a part of the experimentally detected human-*B. mallei* interactions. If not, we added this interaction into the orthologous human-*B. mallei* dataset. The resulting orthologous dataset consisted of 649 interactions between 419 human proteins and 19 *B. mallei* proteins, corresponding to 82% of the murine-*B. mallei* PPIs. Finally, we merged the experimental human-*B. mallei* dataset with the orthologous human-*B. mallei* dataset to create a merged set of human-*B. mallei* PPIs. The resulting merged dataset consisted of 1,235 unique interactions between 21 *B. mallei* and 828 human proteins (S1 Data and S6 Table). Approximately 72% (890) of these represent interactions among the nine *B. mallei* known virulence factors (Table 1) and 663 unique human proteins. All

proteins were annotated by their official gene symbols as defined in the HUGO Gene Nomenclature Committee database [39].

We used the National Center for Biotechnology Information HomoloGene database of homologs (<http://www.ncbi.nlm.nih.gov/homologene>) to identify human-murine orthologs [40].

Topological properties of human proteins interacting with *B. mallei* in the human PPI network

We calculated the following topological properties for a set of human proteins interacting with *B. mallei*: 1) the number of human proteins interacting with *B. mallei* proteins (N_p); 2) the average number of their interacting partners in the human PPI network (D); 3) the clustering coefficient, i.e., the number of interactions among the nearest neighbors (C); the average shortest path between any two proteins in the set (SP); the average number of interacting partners in the human PPI network where both partners interact with *B. mallei* proteins (D_i); and the number of host proteins in the largest connected component (N_p^{LCC}). All calculations were performed in R using the igraph package [41]. We evaluated whether the observed values for each of the five properties were statistically significant as follows. From the human interactome, we randomly selected the same number of proteins as the number of proteins interacting with *B. mallei* virulence factors. Next, we calculated each of the five topological properties for this random set of proteins, repeating the procedure 10^3 times. This procedure yielded 10^3 values for each property, which followed a Normal distribution (Normality was evaluated using the quantile-quantile plots and the Kolmogorov-Smirnov test [42], where we found that there was not enough evidence in the data to suggest that the distributions were not Normal). Then, for each property, we evaluated a relationship between the observed value for proteins interacting with *B. mallei* and the values obtained for random protein sets using a Z-score. Finally, we computed the p -values corresponding to the resulting Z-scores.

Gene set functional enrichment analyses

We performed GO and KEGG enrichment analyses in R using the Bioconductor packages BioMart and KEGGgraph, respectively [43, 44]. As the universe of human proteins, we used all constituent proteins from the human PPI network. As GO terms are specified at multiple levels of detail, we used a complete GO tree annotation, excluding the root and the top two levels of GO terms. GO annotation was obtained from BioMart [43]. For the KEGG enrichment analysis, as the universe of human proteins, we used the human proteins available in KEGGgraph that participated in at least one KEGG pathway [44]. All obtained p -values were assessed using the Benjamini and Hochberg multiple test correction [45]. We retained only annotations that were enriched at an FDR control level of 0.05, i.e., there is a less than 5% chance that the obtained p -values are not statistically significant.

We performed two types of enrichment analysis: standard enrichment analysis and network-based enrichment analysis. In the standard enrichment analysis, we computed the probability of observing the number of proteins annotated with a given term using the hypergeometric distribution. In the network-based enrichment analysis, we first identified the largest connected component in the human interactome that consisted of human proteins interacting with at least one *B. mallei* virulence factor (denoted as LCC). Then, we counted the number of proteins n_t in the LCC that were annotated with a GO biological process term t . Additionally, for each term t , we identified connected (sub)networks of LCC in which all proteins were annotated with t ; these we termed *interaction modules*. We denoted such interaction modules as IM_t and the number of proteins in these modules as m_t .

We evaluated whether the observed IM_t interaction modules were statistically significant as follows. First, we evaluated whether the observed interacting proteins and their corresponding biological processes in the *LCC* were statistically significant compared to an equal number of random proteins. For each GO term t , we counted the number of proteins in the random set that were annotated with t (denoted as r_t). Additionally, for each t , we identified interaction modules in which all proteins from the random set were annotated with t . We denoted the number of proteins in such interaction modules as s_t . We repeated this procedure 10^4 times. Given the obtained values, we defined the probability p_t of observing n_t proteins annotated with t as

$$p_t = \frac{1}{10^4} \cdot \sum_{i=1}^{10^4} k_i, \text{ where } k_i = \begin{cases} 1 & \text{if } (r_t)_i \geq n_t \\ 0 & \text{otherwise} \end{cases}, \quad (1)$$

and the probability p_{rp} of observing an interaction module IM_b given a random set of host proteins, as

$$p_{rp} = \frac{1}{10^4} \cdot \sum_{i=1}^{10^4} k_i, \text{ where } k_i = \begin{cases} 1 & \text{if } (s_t)_i \geq m_t \\ 0 & \text{otherwise} \end{cases}. \quad (2)$$

Second, we evaluated whether the observed interacting proteins and their corresponding biological processes were statistically significant compared to random interactions. We randomly rewired the human protein interaction network, while preserving the same degree distribution as observed in the original network. Next, we mapped human proteins from the *LCC* on the rewired network and, for each term t , identified interaction modules in which all proteins were annotated with t . We denoted the number of proteins in such modules as w_t . We repeated this procedure 10^4 times. Finally, we calculated the probability p_{rw} of observing an interaction module, IM_t , given a random set of host interactions, as

$$p_{rw} = \frac{1}{10^4} \cdot \sum_{i=1}^{10^4} k_i, \text{ where } k_i = \begin{cases} 1 & \text{if } (w_t)_i \geq m_t \\ 0 & \text{otherwise} \end{cases}. \quad (3)$$

IM_t interaction modules with $p_t \leq 0.01$, $p_{rp} \leq 0.01$, and $p_{rw} \leq 0.01$ contain a statistically significant number of human proteins and interactions among them, and are statistically significantly enriched in a biological process t .

Human-*S. enterica* and human-*Y. pestis* protein interactions set

The human-*S. enterica* subsp. *enterica* serovar Typhimurium dataset consisted of 62 interactions between 21 *S. enterica* virulence-associated proteins and 51 human proteins identified in several small-scale experiments [12]. The majority of *S. enterica* proteins from this set were associated with the bacterial T3SS. The human-*Y. pestis* PPI dataset consisted of a union of 204 interactions identified by Y2H screens and 23 interactions identified in several small-scale experiments [13]. The combined human-*Y. pestis* PPI dataset contained 223 unique interactions between 69 *Y. pestis* virulence-associated proteins and 125 human proteins. The majority of *Y. pestis* proteins were also associated with the bacterial T3SS. For the basic comparison of the human-*B. mallei*, human-*S. enterica*, and human-*Y. pestis* PPI networks' characteristics, see [S7 Table](#).

The HPIA algorithm

Network alignment algorithms [46–49] have been used previously to successfully identify both conserved PPIs [50–55] and phylogenetic relationships between species [51–53]. Although the existing network alignment algorithms can be applied to host-pathogen PPIs, these algorithms are not optimized for inter-species alignment, i.e., they cannot differentiate between different types of proteins, such as host and pathogen proteins, and consequently may map host proteins to pathogen proteins and vice versa. The primary motivation for designing a network alignment algorithm was to be able to use previous data and insights from other host-pathogen interaction studies for interpreting our *B. mallei* host interaction data. If conserved network motifs and interaction exist, we can use this information to infer/predict more complex roles for proteins than just transferring sequence-based annotation information [47, 56]. Thus, we designed the HPIA algorithm specifically for the alignment of host-pathogen interactions to augment the sparsely annotated *B. mallei* protein data.

We have taken a number of considerations into account in designing the algorithm based on the nature of the biological problem at hand. First, due to the non-exhaustive nature of Y2H experimentation, the underlying interaction data are not complete [8]. Second, the selected pathogen species are not identical to each other, i.e., proteins and biological processes have evolved differently between the species. Because of this sparse and diverse nature of the pathogen data at hand, we cannot *a priori* expect to obtain satisfactory alignments based on simply mapping human proteins to each other. In this sense, a “perfect” alignment is never attainable, and instead we must rely on approximate alignments with desirable properties, such as biological consistency and sequence similarity. Hence, we developed an alternate approach for aligning bipartite graphs for which one set of nodes (pathogen) are less-well characterized than the other nodes (human).

For the limited number of host-pathogen interactions we have available for comparisons, our algorithm attempts to identify interactions “conserved” on the functional level rather than at the exact protein level. Thus, we can exploit the fact that all three host-pathogen PPI networks contained interactions with human proteins that participated in similar biological processes. In effect, this allowed us to extend the known annotations from the other networks to the previously uncharacterized *B. mallei* virulence factors.

Furthermore, the algorithm guarantees that host proteins will be aligned only to host proteins and that pathogen proteins will be aligned only to pathogen proteins.

Notation. Let $G_1(U_1, V_1, E_1)$ and $G_2(U_2, V_2, E_2)$ be two bipartite graphs (networks), where U_1 and V_1 are two disjoint sets of nodes in G_1 , U_2 and V_2 are two disjoint sets of nodes in G_2 , and E_1 and E_2 are sets of edges of G_1 and G_2 such that every edge in G_1 connects a node in U_1 to one node in V_1 , and every edge in G_2 connects a node in U_2 to one node in V_2 (i.e., no two nodes within the same set are adjacent). Without loss of generality, we can assume that $|U_1| < |U_2|$ and $|V_1| < |V_2|$ (hence, $G_1 < G_2$). The HPIA algorithm is a global network alignment algorithm that uniquely matches each node from U_1 to exactly one node in U_2 , and each node from V_1 to exactly one node in V_2 . Formally, the alignment of G_1 to G_2 can be represented as a set of two ordered pairs $\{(u_1, u_2), (v_1, v_2)\}$, where $u_1 \in U_1$, $u_2 \in U_2$, $v_1 \in V_1$, and $v_2 \in V_2$, and no two ordered pairs share a node.

For the host-pathogen interaction networks (G_1 and G_2), sets U_1 and U_2 correspond to pathogen proteins, sets V_1 and V_2 correspond to host proteins, and sets E_1 and E_2 correspond to interactions between host and pathogen proteins. Thus, our host-pathogen interaction set corresponds to the host-pathogen network, nodes correspond to proteins, and edges correspond to host-pathogen interactions.

Algorithm description. The HPIA algorithm is a seed-and-extend algorithm that consists of the following three steps: 1) pre-processing, 2) identification of local alignment, and 3) identification of global alignment. In this context, we referred to “local alignment” as a smaller area of local network similarity between two networks where not all nodes need to be included, and “global” where all nodes from the smaller network must be aligned to nodes from the larger network [28]. Both mappings are 1-to-1, i.e., one node from one network can be aligned to only one node of the other network. In the first step, the algorithm reads in host-pathogen networks and node annotations, and it calculates similarities between nodes based on either the provided annotation (for node sets with available annotation) or the default topological similarity (for node sets without available annotation). This step also includes handling user-specified seed nodes (nodes that should be mapped to each other). The HPIA algorithm allows a user to provide a set of node pairs (seeds) and/or to employ the algorithm’s feature to automatically search for seed pairs (this search is based on the sequence similarity or equivalent protein names). The HPIA algorithm can treat seed pairs in two ways: 1) as a suggestion for node alignment, i.e., nodes that should be aligned to each other if other alignment constraints are satisfied (see below), or 2) as a requirement for the alignment, i.e., seed pairs that have to be aligned to each other. The first step also includes initialization of the aligned pairs list as empty.

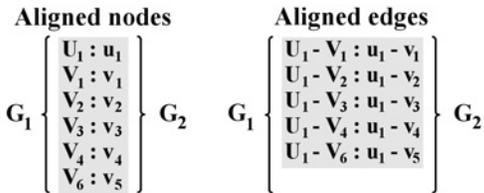
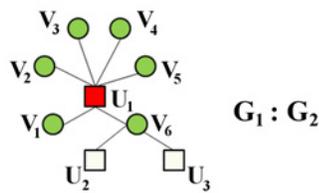
In the second step, the HPIA algorithm first identifies a pair of seed nodes (s_1, s_2), where $s_1 \in G_1$ and $s_2 \in G_2$, based on the node similarity measures from one of the following three sets (in order of preference): 1) a set of aligned protein pairs in which both proteins are adjacent to at least one unaligned protein, 2) a set of user-suggested seed pairs, and 3) a set of unaligned proteins. All three of these sets contain proteins from the host and pathogen sets and, thus, a pair of seed nodes can come from either the host or pathogen sets of proteins. However, if there are seed node candidates from both sets, the HPIA algorithm preferentially selects a pathogen set. Once selected, the seed pair (s_1, s_2) is added to the list of aligned pairs. Next, the HPIA algorithm expands around the seed nodes by greedily aligning their direct neighbors s_{1i} and s_{2j} ($s_{1i} \in N[s_1]$ and $s_{2j} \in N[s_2]$), based on the given node similarity measure (see below). The HPIA algorithm repeats step two while there exists at least one unaligned pair of host proteins or pathogen proteins adjacent to at least one other unaligned protein. When there are no such pairs left, HPIA proceeds to the third step.

In the third step, the HPIA algorithm greedily aligns all of the remaining unaligned pathogen nodes in G_1 to unaligned pathogen nodes G_2 and all of the remaining unaligned host nodes in G_1 to unaligned host nodes G_2 , solely based on node similarity measure. Each pair of nodes is aligned one at a time based on the given node similarity (see below); network connectivity information is not taken into account explicitly (only as a part of the node similarity measure). The results of the alignment are lists of aligned nodes and edges and the following alignment statistics: the total number and percentage of aligned nodes, the number and percentage of aligned pathogen and host nodes, and the number and percentage of aligned edges.

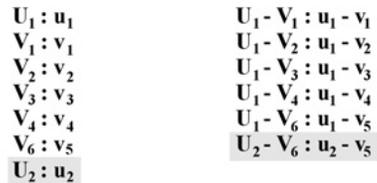
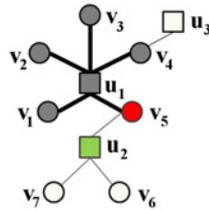
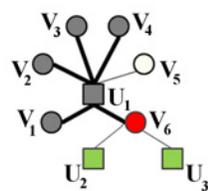
The HPIA algorithm allows a user to provide a set of node pairs (seeds) and to use an automatic search for seed pairs (this search is based on the sequence similarity or equivalent protein names) by specifying an optional parameter, “additionalSeeds.” The HPIA algorithm can treat these seed pairs in two ways. If the “relaxSeeds” option is given, the nodes will be aligned to each other only if other alignment constraints are satisfied as detailed below. If the “relaxSeeds” option is not given, the seed pairs are forced to be aligned to each other. We recommend that the “relaxSeeds” parameter be turned on if the “additionalSeeds” parameter is used.

All ties in the algorithm are broken randomly. The implementation of the HPIA algorithm is also presented as pseudocode in [S1 Text](#). [Fig. 7](#) shows a high-level description of the HPIA alignment algorithm.

A: Initial seed and extend phase



B: Iterative seed and extend phase



C: Global alignment greedy phase

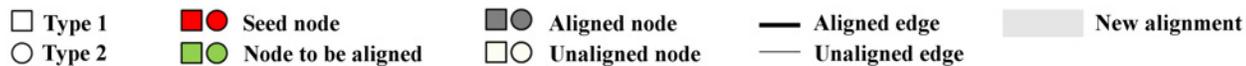
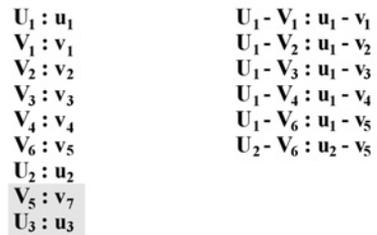
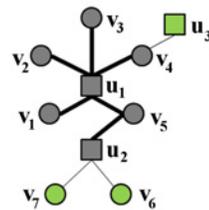
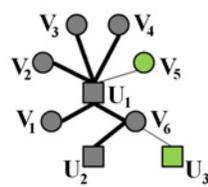


Fig 7. Host-Pathogen Interactions Alignment (HPIA) algorithm. The HPIA algorithm is a seed-and-extend algorithm that aligns two bipartite graphs, e.g., two different host-pathogen protein interaction networks. A) Given an initial pair of seed nodes (red nodes U_1 and u_1) from two graphs G_1 (left) and G_2 (right), the algorithm first aligns seed nodes to each other. Then, it aligns the neighbors of the seed nodes from the first graph (green nodes V_1 - V_6) to the neighbors of the seed nodes in the second graph (green nodes v_1 - v_5) based on the node similarity measure (as defined in Equations 6, 8, and 9). This procedure results in six aligned nodes and five aligned edges. B) The algorithm iteratively selects new seeds and extends around them, e.g., it selects nodes V_6 from G_1 and v_5 from G_2 as new seed nodes and, based on the node similarity measure, aligns their unaligned neighbors U_2 to u_2 , creating an additional aligned edge (U_2 - V_6 to u_2 - v_5). C) When the algorithm cannot find any seed nodes of the same type that have unaligned neighbors, it greedily aligns all of the remaining unaligned nodes based on their type and the node similarity measure. Some nodes may remain unaligned if the graphs' sizes vary, e.g., when there is no match for v_6 from G_2 in G_1 . The HPIA algorithm generates a list of aligned nodes and a list of aligned edges inferred from the aligned nodes.

doi:10.1371/journal.pcbi.1004088.g007

Node similarity measures. The HPIA algorithm uses one or more topologically and biologically based protein similarity measures to identify conserved interactions. Similarity between two proteins can always be calculated based on at least one of the metrics and, thus, the algorithm always has a metric to match one protein to another. If no node annotation is provided, the HPIA algorithm uses the default topological similarity measure, $S_{DT}(n_1, n_2)$, to calculate the similarity between nodes $n_1 \in G_1$ and $n_2 \in G_2$:

$$S_{DT}(n_1, n_2) = \alpha \cdot \frac{\min[\deg(n_1), \deg(n_2)]}{\max[\deg(n_1), \deg(n_2)]} + (1 - \alpha) \cdot \frac{\min[nd(n_1), nd(n_2)]}{\max[nd(n_1), nd(n_2)]} \quad (4)$$

where $\deg(n)$ denotes the degree of node n and $nd(n)$ denotes the neighborhood density of n , defined as

$$nd(n) = \sum_{n_k \in N[n]} \deg(n_k) \quad (5)$$

where $N[n] = N(n) \cup \{n\}$ represents the closed neighborhood of node n , i.e., the node n and the set of its adjacent nodes (for a PPI network, this corresponds to a protein and all of its interacting partners). α is a parameter in $[0, 1]$ that controls the contribution of the degree of a node to its similarity function. We empirically selected $\alpha = 0.7$, as we wanted to weight the number of direct host-pathogen interactions higher than the number of host-pathogen-host or pathogen-host-pathogen interactions.

If node annotation is provided, the HPIA algorithm defines a similarity between pathogen proteins $u_1 \in U_1$ and $u_2 \in U_2$ as

$$S(u_1, u_2) = S_{GDV-P}(u_1, u_2) + S_{EB}(u_1, u_2) + S_{GO}(u_1, u_2) \quad (6)$$

and a similarity between host proteins $v_1 \in V_1$ and $v_2 \in V_2$ as

$$S(v_1, v_2) = S_{GDV-P}(v_1, v_2) + S_{GDV-H}(v_1, v_2) + S_{EB}(v_1, v_2) + S_{GO}(v_1, v_2), \quad (7)$$

where S_{GDV-P} and S_{GDV-H} denote the graphlet degree vector similarity [57] derived from the host-pathogen interaction network and the host-PPI network, respectively; S_{EB} represents the BLAST expected value (E-value) similarity [58], defined as $1 - E$ -value for E -values ≤ 1 and 0 otherwise; and S_{GO} denotes the GO term annotation similarity calculated using the Jaccard similarity measure [59]. If a specific type of annotation is not provided, the HPIA algorithm assigns the similarity value 0 to the corresponding similarity parameters, e.g., if BLAST E -values are not provided, the value of S_{EB} for all pairs of nodes is set to 0. We did not add the graphlet degree vector similarity for the pathogen networks because only a few pathogen-PPI networks are available, none of which are for *B. mallei*.

Data. For the topological node annotation, we used graphlet degree vectors [57] of all host and pathogen proteins from the host-pathogen PPI networks. Host proteins found in the host PPI network [25] were additionally annotated with another set of graphlet degree vectors calculated based on the host PPI network topology. We used GO annotation [23] downloaded from UNIPROT [60] [the lowest (leaf) level] as the biological node annotation. Additionally, we used BLAST E -values of ≤ 0.01 to define similarities between proteins [58]. Protein sequences were downloaded from UNIPROT and aligned using BLAST pairwise sequence alignment.

Alignment quality. To assess the topological quality of the alignment, we used *edge correctness* (EC), defined as the percentage of edges in G_1 that were aligned to edges in G_2 [51]. To assess the biological quality of the alignment, we evaluated whether the number of aligned protein pairs that share one or more GO term(s) was statistically significant compared to the number we could expect at random using the standard model of sampling without replacement, as described in previous studies [51–53]. We used the same approach to assess the statistical significance of the alignment of two bipartite networks, $G_1(U_1, V_1, E_1)$ and $G_2(U_2, V_2, E_2)$, with the EC of $x\%$ (similar to the implementation described above [51–53]). We aligned each pair of host-pathogen interaction networks 30 times and reported the average and standard deviations of the alignment scores over all runs, as well as the best score (S8 Table). We ascertained the robustness of the alignments with respect to the E -value cutoff and observed no significant differences in the results when lowering the cutoff value from 10^{-2} to 10^{-3} . To assess the biological quality of the alignment, we evaluated whether the number of aligned protein pairs that share one or more GO term(s) was statistically significant compared to the number we could expect from a random alignment. Given that all obtained alignments were of similar biological quality, we further refined our prediction by using the alignments that had the highest EC score, i.e., we used the alignment with the highest EC score to infer the role of *B. mallei* proteins.

Implementation and availability

All statistical analyses were performed in R. All networks were plotted using Cytoscape [61]. The cross-species network alignment algorithm was developed in C++. Executable files and examples for the HPIA algorithm are provided at <http://www.bhsai.org/downloads/hpia/>.

Supporting Information

S1 Data. This file contains a set of human-*B. mallei* PPIs and a set of human PPIs used in this study.

(XLSX)

S1 Text. This file contains the HPIA algorithm pseudo code and supplementary tables.

(DOCX)

S1 Table. A list of interaction modules statistically significantly enriched in Gene Ontology (GO) biological processes for human proteins interacting with known *B. mallei* virulence factors.

(DOCX)

S2 Table. Host proteins associated with distinct biological processes that interacted with known *B. mallei* virulence factors.

(DOCX)

S3 Table. Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways statistically significantly enriched in human proteins interacting with known and/or putative *B. mallei* virulence factors.

(DOCX)

S4 Table. Subnetworks statistically significantly enriched in Gene Ontology (GO) biological processes for human proteins interacting with known and/or putative *B. mallei* virulence factors.

(DOCX)

S5 Table. Functional annotation of *B. mallei* proteins inferred from the host-pathogen interaction network alignment.

(DOCX)

S6 Table. A summary of host-*B. mallei* interactions used in the study.

(DOCX)

S7 Table. Characteristics of host-pathogen networks used in network alignment.

(DOCX)

S8 Table. Functional annotation of *B. mallei* proteins inferred from the host-pathogen network alignment.

(DOCX)

Acknowledgments

The opinions and assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the United States (U.S.) Army or of the U.S. Department of Defense. This paper has been approved for public release with unlimited distribution.

Author Contributions

Conceived and designed the experiments: NZ RP DD JR AW VM. Generated the interaction data: SVR KK RP. Developed the algorithms and performed the calculations: VM. Analyzed the data: VM NZ AW. Wrote the paper: VM AW JR.

References

- Estes DM, Dow SW, Schweizer HP, Torres AG (2010) Present and future therapeutic strategies for melioidosis and glanders. *Expert Rev Anti Infect Ther* 8: 325–338. doi: [10.1586/eri.10.4](https://doi.org/10.1586/eri.10.4) PMID: [20192686](https://pubmed.ncbi.nlm.nih.gov/20192686/)
- Allwood EM, Devenish RJ, Prescott M, Adler B, Boyce JD (2011) Strategies for intracellular survival of *Burkholderia pseudomallei*. *Front Microbiol* 2: 170. doi: [10.3389/fmicb.2011.00170](https://doi.org/10.3389/fmicb.2011.00170) PMID: [22007185](https://pubmed.ncbi.nlm.nih.gov/22007185/)
- Galyov EE, Brett PJ, DeShazer D (2010) Molecular insights into *Burkholderia pseudomallei* and *Burkholderia mallei* pathogenesis. *Annu Rev Microbiol* 64: 495–517. doi: [10.1146/annurev.micro.112408.134030](https://doi.org/10.1146/annurev.micro.112408.134030) PMID: [20528691](https://pubmed.ncbi.nlm.nih.gov/20528691/)
- Lazar Adler NR, Stevens JM, Stevens MP, Galyov EE (2011) Autotransporters and their role in the virulence of *Burkholderia pseudomallei* and *Burkholderia mallei*. *Front Microbiol* 2: 151. doi: [10.3389/fmicb.2011.00151](https://doi.org/10.3389/fmicb.2011.00151) PMID: [21811486](https://pubmed.ncbi.nlm.nih.gov/21811486/)
- Losada L, Ronning CM, DeShazer D, Woods D, Fedorova N, et al. (2010) Continuing evolution of *Burkholderia mallei* through genome reduction and large-scale rearrangements. *Genome Biol Evol* 2: 102–116. doi: [10.1093/gbe/evq003](https://doi.org/10.1093/gbe/evq003) PMID: [20333227](https://pubmed.ncbi.nlm.nih.gov/20333227/)
- Nierman WC, DeShazer D, Kim HS, Tettelin H, Nelson KE, et al. (2004) Structural flexibility in the *Burkholderia mallei* genome. *Proc Natl Acad Sci U S A* 101: 14246–14251. doi: [10.1073/pnas.0403306101](https://doi.org/10.1073/pnas.0403306101) PMID: [15377793](https://pubmed.ncbi.nlm.nih.gov/15377793/)
- Schell MA, Ulrich RL, Ribot WJ, Brueggemann EE, Hines HB, et al. (2007) Type VI secretion is a major virulence determinant in *Burkholderia mallei*. *Mol Microbiol* 64: 1466–1485. doi: [10.1111/j.1365-2958.2007.05734.x](https://doi.org/10.1111/j.1365-2958.2007.05734.x) PMID: [17555434](https://pubmed.ncbi.nlm.nih.gov/17555434/)
- Memišević V, Zavaljevski N, Pieper R, Rajagopala SV, Kwon K, et al. (2013) Novel *Burkholderia mallei* virulence factors linked to specific host-pathogen protein interactions. *Mol Cell Proteomics* 12: 3036–3051. doi: [10.1074/mcp.M113.029041](https://doi.org/10.1074/mcp.M113.029041) PMID: [23800426](https://pubmed.ncbi.nlm.nih.gov/23800426/)
- Durmus Tekir S, Cakir T, Ulgen KO (2012) Infection strategies of bacterial and viral pathogens through pathogen-human protein-protein interactions. *Front Microbiol* 3: 46. doi: [10.3389/fmicb.2012.00046](https://doi.org/10.3389/fmicb.2012.00046) PMID: [22347880](https://pubmed.ncbi.nlm.nih.gov/22347880/)
- Dyer MD, Murali TM, Sobral BW (2008) The landscape of human proteins interacting with viruses and other pathogens. *PLOS Pathog* 4: e32. doi: [10.1371/journal.ppat.0040032](https://doi.org/10.1371/journal.ppat.0040032) PMID: [18282095](https://pubmed.ncbi.nlm.nih.gov/18282095/)
- Dyer MD, Neff C, Dufford M, Rivera CG, Shattuck D, et al. (2010) The human-bacterial pathogen protein interaction networks of *Bacillus anthracis*, *Francisella tularensis*, and *Yersinia pestis*. *PLOS ONE* 5: e12089. doi: [10.1371/journal.pone.0012089](https://doi.org/10.1371/journal.pone.0012089) PMID: [20711500](https://pubmed.ncbi.nlm.nih.gov/20711500/)
- Schleker S, Sun J, Raghavan B, Srnec M, Muller N, et al. (2012) The current *Salmonella*-host interactome. *Proteomics Clin Appl* 6: 117–133. doi: [10.1002/prca.201100083](https://doi.org/10.1002/prca.201100083) PMID: [22213674](https://pubmed.ncbi.nlm.nih.gov/22213674/)
- Yang H, Ke Y, Wang J, Tan Y, Myeni SK, et al. (2011) Insight into bacterial virulence mechanisms against host immune response via the *Yersinia pestis*-human protein-protein interaction network. *Infect Immun* 79: 4413–4424. doi: [10.1128/IAI.05622-11](https://doi.org/10.1128/IAI.05622-11) PMID: [21911467](https://pubmed.ncbi.nlm.nih.gov/21911467/)
- Arnold R, Boonen K, Sun MG, Kim PM (2012) Computational analysis of interactomes: Current and future perspectives for bioinformatics approaches to model the host-pathogen interaction space. *Methods* 57: 508–518. doi: [10.1016/j.ymeth.2012.06.011](https://doi.org/10.1016/j.ymeth.2012.06.011) PMID: [22750305](https://pubmed.ncbi.nlm.nih.gov/22750305/)
- Solava RW, Michaels RP, Milenković T (2012) Graphlet-based edge clustering reveals pathogen-interacting proteins. *Bioinformatics* 28: i480–i486. doi: [10.1093/bioinformatics/bts376](https://doi.org/10.1093/bioinformatics/bts376) PMID: [22962470](https://pubmed.ncbi.nlm.nih.gov/22962470/)
- Kshirsagar M, Carbonell J, Klein-Seetharaman J (2013) Multitask learning for host-pathogen protein interactions. *Bioinformatics* 29: i217–i226. doi: [10.1093/bioinformatics/btt245](https://doi.org/10.1093/bioinformatics/btt245) PMID: [23812987](https://pubmed.ncbi.nlm.nih.gov/23812987/)
- Anderson DM, Frank DW (2012) Five mechanisms of manipulation by bacterial effectors: A ubiquitous theme. *PLOS Pathog* 8: e1002823. doi: [10.1371/journal.ppat.1002823](https://doi.org/10.1371/journal.ppat.1002823) PMID: [22927812](https://pubmed.ncbi.nlm.nih.gov/22927812/)
- Bhavsar AP, Guttman JA, Finlay BB (2007) Manipulation of host-cell pathways by bacterial pathogens. *Nature* 449: 827–834. doi: [10.1038/nature06247](https://doi.org/10.1038/nature06247) PMID: [17943119](https://pubmed.ncbi.nlm.nih.gov/17943119/)
- Burnick MN, DeShazer D, Nair V, Gherardini FC, Brett PJ (2010) *Burkholderia mallei* cluster 1 type VI secretion mutants exhibit growth and actin polymerization defects in RAW 264.7 murine macrophages. *Infect Immun* 78: 88–99. doi: [10.1128/IAI.00985-09](https://doi.org/10.1128/IAI.00985-09) PMID: [19884331](https://pubmed.ncbi.nlm.nih.gov/19884331/)

20. Stevens JM, Ulrich RL, Taylor LA, Wood MW, Deshazer D, et al. (2005) Actin-binding proteins from *Burkholderia mallei* and *Burkholderia thailandensis* can functionally compensate for the actin-based motility defect of a *Burkholderia pseudomallei* *bimA* mutant. *J Bacteriol* 187: 7857–7862. doi: [10.1128/JB.187.22.7857-7862.2005](https://doi.org/10.1128/JB.187.22.7857-7862.2005) PMID: [16267310](https://pubmed.ncbi.nlm.nih.gov/16267310/)
21. Ulrich RL, DeShazer D (2004) Type III secretion: A virulence factor delivery system essential for the pathogenicity of *Burkholderia mallei*. *Infect Immun* 72: 1150–1154. doi: [10.1128/IAI.72.2.1150-1154.2004](https://doi.org/10.1128/IAI.72.2.1150-1154.2004) PMID: [14742564](https://pubmed.ncbi.nlm.nih.gov/14742564/)
22. Memišević V, Kumar K, Cheng L, Zavaljevski N, DeShazer D, et al. (2014) DBSecSys: A database of *Burkholderia mallei* secretion systems. *BMC Bioinformatics* 15: 244. doi: [10.1186/1471-2105-15-244](https://doi.org/10.1186/1471-2105-15-244) PMID: [25030112](https://pubmed.ncbi.nlm.nih.gov/25030112/)
23. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene Ontology: Tool for the unification of biology. *Nat Genet* 25: 25–29. doi: [10.1038/75556](https://doi.org/10.1038/75556) PMID: [10802651](https://pubmed.ncbi.nlm.nih.gov/10802651/)
24. Shanks J, Burnick MN, Brett PJ, Waag DM, Spurgers KB, et al. (2009) *Burkholderia mallei* *tssM* encodes a putative deubiquitinase that is secreted and expressed inside infected RAW 264.7 murine macrophages. *Infect Immun* 77: 1636–1648. doi: [10.1128/IAI.01339-08](https://doi.org/10.1128/IAI.01339-08) PMID: [19168747](https://pubmed.ncbi.nlm.nih.gov/19168747/)
25. Yu X, Wallqvist A, Reifman J (2012) Inferring high-confidence human protein-protein interactions. *BMC Bioinformatics* 13: 79. doi: [10.1186/1471-2105-13-79](https://doi.org/10.1186/1471-2105-13-79) PMID: [22558947](https://pubmed.ncbi.nlm.nih.gov/22558947/)
26. Van Der Hoeven PC, Van Der Wal JC, Ruurs P, Van Dijk MC, Van Blitterswijk J (2000) 14–3–3 isotypes facilitate coupling of protein kinase C-zeta to Raf-1: Negative regulation by 14–3–3 phosphorylation. *Biochem J* 345 Pt 2: 297–306. doi: [10.1042/0264-6021.3450297](https://doi.org/10.1042/0264-6021.3450297)
27. Jeffery CJ (2009) Moonlighting proteins—an update. *Mol Biosyst* 5: 345–350. doi: [10.1039/b900658n](https://doi.org/10.1039/b900658n) PMID: [19396370](https://pubmed.ncbi.nlm.nih.gov/19396370/)
28. Ovadi J (2011) Moonlighting proteins in neurological disorders. *IUBMB Life* 63: 453–456. doi: [10.1002/iub.491](https://doi.org/10.1002/iub.491) PMID: [21698748](https://pubmed.ncbi.nlm.nih.gov/21698748/)
29. Henderson B, Martin A (2011) Bacterial virulence in the moonlight: Multitasking bacterial moonlighting proteins are virulence determinants in infectious disease. *Infect Immun* 79: 3476–3491. doi: [10.1128/IAI.00179-11](https://doi.org/10.1128/IAI.00179-11) PMID: [21646455](https://pubmed.ncbi.nlm.nih.gov/21646455/)
30. Sriram G, Martinez JA, McCabe ER, Liao JC, Dipple KM (2005) Single-gene disorders: What role could moonlighting enzymes play? *Am J Hum Genet* 76: 911–924. doi: [10.1086/430799](https://doi.org/10.1086/430799) PMID: [15877277](https://pubmed.ncbi.nlm.nih.gov/15877277/)
31. Wieland T, Mittmann C (2003) Regulators of G-protein signalling: Multifunctional proteins with impact on signalling in the cardiovascular system. *Pharmacol Ther* 97: 95–115. doi: [10.1016/S0163-7258\(02\)00326-1](https://doi.org/10.1016/S0163-7258(02)00326-1) PMID: [12559385](https://pubmed.ncbi.nlm.nih.gov/12559385/)
32. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, et al. (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 27: 29–34. doi: [10.1093/nar/27.1.29](https://doi.org/10.1093/nar/27.1.29) PMID: [9847135](https://pubmed.ncbi.nlm.nih.gov/9847135/)
33. Rottner K, Stradal TE, Wehland J (2005) Bacteria-host-cell interactions at the plasma membrane: Stories on actin cytoskeleton subversion. *Dev Cell* 9: 3–17. doi: [10.1016/j.devcel.2005.06.002](https://doi.org/10.1016/j.devcel.2005.06.002) PMID: [15992537](https://pubmed.ncbi.nlm.nih.gov/15992537/)
34. Yu X, Ivanic J, Memišević V, Wallqvist A, Reifman J (2011) Categorizing biases in high-confidence high-throughput protein-protein interaction data sets. *Mol Cell Proteomics* 10: M111.012500. doi: [10.1074/mcp.M111.012500](https://doi.org/10.1074/mcp.M111.012500) PMID: [21876202](https://pubmed.ncbi.nlm.nih.gov/21876202/)
35. Blaylock B, Riordan KE, Missiakas DM, Schneewind O (2006) Characterization of the *Yersinia enterocolitica* type III secretion ATPase YscN and its regulator, YscL. *J Bacteriol* 188: 3525–3534. doi: [10.1128/JB.188.10.3525-3534.2006](https://doi.org/10.1128/JB.188.10.3525-3534.2006) PMID: [16672607](https://pubmed.ncbi.nlm.nih.gov/16672607/)
36. Freeman JA, Rappal C, Kuhle V, Hensel M, Miller SI (2002) SpiC is required for translocation of *Salmonella pathogenicity* island 2 effectors and secretion of translocon proteins SseB and SseC. *J Bacteriol* 184: 4971–4980. doi: [10.1128/JB.184.18.4971-4980.2002](https://doi.org/10.1128/JB.184.18.4971-4980.2002) PMID: [12193612](https://pubmed.ncbi.nlm.nih.gov/12193612/)
37. Woestyn S, Allaoui A, Wattiau P, Cornelis GR (1994) YscN, the putative energizer of the *Yersinia* Yop secretion machinery. *J Bacteriol* 176: 1561–1569. PMID: [8132449](https://pubmed.ncbi.nlm.nih.gov/8132449/)
38. Yu XJ, Liu M, Holden DW (2004) SsaM and SpiC interact and regulate secretion of *Salmonella* pathogenicity island 2 type III secretion system effectors and translocators. *Mol Microbiol* 54: 604–619. doi: [10.1111/j.1365-2958.2004.04297.x](https://doi.org/10.1111/j.1365-2958.2004.04297.x) PMID: [15491354](https://pubmed.ncbi.nlm.nih.gov/15491354/)
39. Gray KA, Daugherty LC, Gordon SM, Seal RL, Wright MW, et al. (2013) Genenames.org: The HGNC resources in 2013. *Nucleic Acids Res* 41: D545–D552. doi: [10.1093/nar/gks1066](https://doi.org/10.1093/nar/gks1066) PMID: [23161694](https://pubmed.ncbi.nlm.nih.gov/23161694/)
40. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, et al. (2012) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 40: D13–D25. doi: [10.1093/nar/gkr1184](https://doi.org/10.1093/nar/gkr1184) PMID: [22140104](https://pubmed.ncbi.nlm.nih.gov/22140104/)
41. Csardi G, Nepusz T (2006) The igraph software package for complex network research. *InterJournal Complex Systems*: 1695.

42. Massey FJ (1951) The Kolmogorov-Smirnov test for goodness of fit. *J Am Stat Assoc* 46: 68–78. doi: [10.1080/01621459.1951.10500769](https://doi.org/10.1080/01621459.1951.10500769)
43. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, et al. (2005) BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics* 21: 3439–3440. doi: [10.1093/bioinformatics/bti525](https://doi.org/10.1093/bioinformatics/bti525) PMID: [16082012](https://pubmed.ncbi.nlm.nih.gov/16082012/)
44. Zhang JD, Wiemann S (2009) KEGGgraph: A graph approach to KEGG PATHWAY in R and bioconductor. *Bioinformatics* 25: 1470–1471. doi: [10.1093/bioinformatics/btp167](https://doi.org/10.1093/bioinformatics/btp167) PMID: [19307239](https://pubmed.ncbi.nlm.nih.gov/19307239/)
45. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Statist Soc* 57: 289–300.
46. Seah BS, Bhowmick SS, Dewey CF Jr., (2014) DualAligner: A dual alignment-based strategy to align protein interaction networks. *Bioinformatics* 30: 2619–2626. doi: [10.1093/bioinformatics/btu358](https://doi.org/10.1093/bioinformatics/btu358) PMID: [24872427](https://pubmed.ncbi.nlm.nih.gov/24872427/)
47. Hu J, Kehr B, Reinert K (2014) NetCoffee: A fast and accurate global alignment approach to identify functionally conserved proteins in multiple networks. *Bioinformatics* 30: 540–548. doi: [10.1093/bioinformatics/btt715](https://doi.org/10.1093/bioinformatics/btt715) PMID: [24336806](https://pubmed.ncbi.nlm.nih.gov/24336806/)
48. Alkan F, Erten C (2014) BEAMS: Backbone extraction and merge strategy for the global many-to-many alignment of multiple PPI networks. *Bioinformatics* 30: 531–539. doi: [10.1093/bioinformatics/btt713](https://doi.org/10.1093/bioinformatics/btt713) PMID: [24336414](https://pubmed.ncbi.nlm.nih.gov/24336414/)
49. Neyshabur B, Khadem A, Hashemifar S, Arab SS (2013) NETAL: A new graph-based method for global alignment of protein-protein interaction networks. *Bioinformatics* 29: 1654–1662. doi: [10.1093/bioinformatics/btt202](https://doi.org/10.1093/bioinformatics/btt202) PMID: [23696650](https://pubmed.ncbi.nlm.nih.gov/23696650/)
50. Aladag AE, Erten C (2013) SPINAL: Scalable protein interaction network alignment. *Bioinformatics* 29: 917–924. doi: [10.1093/bioinformatics/btt071](https://doi.org/10.1093/bioinformatics/btt071) PMID: [23413436](https://pubmed.ncbi.nlm.nih.gov/23413436/)
51. Kuchaiev O, Milenković T, Memišević V, Hayes W, Pržulj N (2010) Topological network alignment uncovers biological function and phylogeny. *J R Soc Interface* 7: 1341–1354. doi: [10.1098/rsif.2010.0063](https://doi.org/10.1098/rsif.2010.0063) PMID: [20236959](https://pubmed.ncbi.nlm.nih.gov/20236959/)
52. Kuchaiev O, Pržulj N (2011) Integrative network alignment reveals large regions of global network similarity in yeast and human. *Bioinformatics* 27: 1390–1396. doi: [10.1093/bioinformatics/btr127](https://doi.org/10.1093/bioinformatics/btr127) PMID: [21414992](https://pubmed.ncbi.nlm.nih.gov/21414992/)
53. Memišević V, Pržulj N (2012) C-GRAAL: Common-neighbors-based global GRAph ALignment of biological networks. *Integr Biol* 4: 734–743. doi: [10.1039/c2ib00140c](https://doi.org/10.1039/c2ib00140c)
54. Patro R, Kingsford C (2012) Global network alignment using multiscale spectral signatures. *Bioinformatics* 28: 3105–3114. doi: [10.1093/bioinformatics/bts592](https://doi.org/10.1093/bioinformatics/bts592) PMID: [23047556](https://pubmed.ncbi.nlm.nih.gov/23047556/)
55. Rivera C, Murali TM (2009) Identifying evolutionarily conserved protein interaction modules using GraphHopper. In: S. Rajasekaran, editor. *Bioinformatics and Computational Biology*. Springer Berlin Heidelberg. pp. 67–78.
56. Sharan R, Suthram S, Kelley RM, Kuhn T, McCuine S, et al. (2005) Conserved patterns of protein interaction in multiple species. *Proc Natl Acad Sci U S A* 102: 1974–1979. doi: [10.1073/pnas.0409522102](https://doi.org/10.1073/pnas.0409522102) PMID: [15687504](https://pubmed.ncbi.nlm.nih.gov/15687504/)
57. Milenković T, Pržulj N (2008) Uncovering biological network function via graphlet degree signatures. *Cancer Inform* 6: 257–273. PMID: [19259413](https://pubmed.ncbi.nlm.nih.gov/19259413/)
58. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410. doi: [10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: [2231712](https://pubmed.ncbi.nlm.nih.gov/2231712/)
59. Jaccard P (1901) Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bulletin de la Société Vaudoise des Sciences Naturelles* 37: 241–272.
60. Magrane M, Consortium U (2011) UniProt Knowledgebase: A hub of integrated protein data. *Database* 2011: bar009.
61. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498–2504. doi: [10.1101/gr.1239303](https://doi.org/10.1101/gr.1239303) PMID: [14597658](https://pubmed.ncbi.nlm.nih.gov/14597658/)
62. Essex-Lopresti AE, Boddey JA, Thomas R, Smith MP, Hartley MG, et al. (2005) A type IV pilin, PilA, contributes to adherence of *Burkholderia pseudomallei* and virulence *in vivo*. *Infect Immun* 73: 1260–1264. doi: [10.1128/IAI.73.2.1260-1264.2005](https://doi.org/10.1128/IAI.73.2.1260-1264.2005) PMID: [15664977](https://pubmed.ncbi.nlm.nih.gov/15664977/)
63. Whitlock GC, Valbuena GA, Popov VL, Judy BM, Estes DM, et al. (2009) *Burkholderia mallei* cellular interactions in a respiratory cell model. *J Med Microbiol* 58: 554–562. doi: [10.1099/jmm.0.007724-0](https://doi.org/10.1099/jmm.0.007724-0) PMID: [19369515](https://pubmed.ncbi.nlm.nih.gov/19369515/)

64. Stevens MP, Haque A, Atkins T, Hill J, Wood MW, et al. (2004) Attenuated virulence and protective efficacy of a *Burkholderia pseudomallei* bsa type III secretion mutant in murine models of melioidosis. *Microbiology* 150: 2669–2676. doi: [10.1099/mic.0.27146-0](https://doi.org/10.1099/mic.0.27146-0) PMID: [15289563](https://pubmed.ncbi.nlm.nih.gov/15289563/)
65. Suparak S, Kespichayawattana W, Haque A, Easton A, Damnin S, et al. (2005) Multinucleated giant cell formation and apoptosis in infected host cells is mediated by *Burkholderia pseudomallei* type III secretion protein BipB. *J Bacteriol* 187: 6556–6560. doi: [10.1128/JB.187.18.6556-6560.2005](https://doi.org/10.1128/JB.187.18.6556-6560.2005) PMID: [16159789](https://pubmed.ncbi.nlm.nih.gov/16159789/)
66. Pilatz S, Breitbach K, Hein N, Fehlhaber B, Schulze J, et al. (2006) Identification of *Burkholderia pseudomallei* genes required for the intracellular life cycle and *in vivo* virulence. *Infect Immun* 74: 3576–3586. doi: [10.1128/IAI.01262-05](https://doi.org/10.1128/IAI.01262-05) PMID: [16714590](https://pubmed.ncbi.nlm.nih.gov/16714590/)