

Decoding the Ubiquitin-Mediated Pathway of Arthropod Disease Vectors

Anthony Choy¹, Maiara S. Severo¹, Ruobai Sun², Thomas Girke², Joseph J. Gillespie³, Joao H. F. Pedra^{1,3*}

1 Institute for Integrative Genome Biology, Center for Disease Vector Research and Department of Entomology, University of California Riverside, Riverside, California, United States of America, **2** Institute for Integrative Genome Biology, Center for Plant Cell Biology, Department of Botany and Plant Sciences, University of California Riverside, Riverside, California, United States of America, **3** Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, Maryland, United States of America

Abstract

Protein regulation by ubiquitin has been extensively described in model organisms. However, characterization of the ubiquitin machinery in disease vectors remains mostly unknown. This fundamental gap in knowledge presents a concern because new therapeutics are needed to control vector-borne diseases, and targeting the ubiquitin machinery as a means for disease intervention has been already adopted in the clinic. In this study, we employed a bioinformatics approach to uncover the ubiquitin-mediated pathway in the genomes of *Anopheles gambiae*, *Aedes aegypti*, *Culex quinquefasciatus*, *Ixodes scapularis*, *Pediculus humanus* and *Rhodnius prolixus*. We observed that (1) disease vectors encode a lower percentage of ubiquitin-related genes when compared to *Drosophila melanogaster*, *Mus musculus* and *Homo sapiens* but not *Saccharomyces cerevisiae*; (2) overall, there are more proteins categorized as E3 ubiquitin ligases when compared to E2-conjugating or E1-activating enzymes; (3) the ubiquitin machinery within the three mosquito genomes is highly similar; (4) ubiquitin genes are more than doubled in the Chagas disease vector (*R. prolixus*) when compared to other arthropod vectors; (5) the deer tick *I. scapularis* and the body louse (*P. humanus*) genomes carry low numbers of E1-activating enzymes and HECT-type E3 ubiquitin ligases; (6) *R. prolixus* have low numbers of RING-type E3 ubiquitin ligases; and (7) *C. quinquefasciatus* present elevated numbers of predicted F-box E3 ubiquitin ligases, JAB and UCH deubiquitinases. Taken together, these findings provide novel opportunities to study the interaction between a pathogen and an arthropod vector.

Citation: Choy A, Severo MS, Sun R, Girke T, Gillespie JJ, et al. (2013) Decoding the Ubiquitin-Mediated Pathway of Arthropod Disease Vectors. PLoS ONE 8(10): e78077. doi:10.1371/journal.pone.0078077

Editor: Roman Ganta, Kansas State University, United States of America

Received: June 25, 2013; **Accepted:** September 16, 2013; **Published:** October 21, 2013

Copyright: © 2013 Choy et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the University of California, Riverside and the National Institutes of Health NIH R01AI043006 and NIH AI093653. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: jpedra@som.umaryland.edu

Introduction

Vector-borne diseases are some of the most prevalent infectious illnesses worldwide. According to the World Health Organization, there are approximately 216 million cases of malaria alone, with over 1.2 million estimated deaths [1]. Dengue fever, another prominent vector-borne disease is responsible for 50-100 million cases each year, and it is making one of the fastest growing infectious maladies. An estimated 120 million individuals are affected each year by lymphatic filariasis [2]. Lyme disease, the most common tick-borne illness in the northern hemisphere, is responsible for 30,000 clinical cases in the United States alone with the actual number possibly being much higher [3]. Overall, vector-borne diseases are responsible for 16% of disability-adjusted life years [4], and these statistics are compounded by the social costs that also affect heavily-infected communities.

Ubiquitin consists of a 76 amino acid protein that carries seven lysine (K) residues (K6, K11, K27, K29, K33, K48 and K63) [5]. Ubiquitin contributes to several fundamental biological processes within cells including, but not limited to: protein turnover, endocytosis, DNA repair, transcription and immunity [6,7]. Ubiquitination involves a ubiquitin-activating enzyme (E1), a ubiquitin-conjugating enzyme (E2) and a ubiquitin-protein ligase (E3) [5]. Ubiquitination is counteracted by de-ubiquitination, and de-ubiquitinases remove ubiquitin chains from molecules by cleaving precursors enzymatically, eliminating ubiquitin from proteins, or by altering the ubiquitin linkage type [5].

In *Drosophila*, it is established that developmental regulation by the evolutionarily conserved Notch receptor depends on ubiquitin [8], and some proteins associated with neuronal control and neurodegenerative disorders undergo ubiquitination [9]. Additionally, there is abundant evidence showing that ubiquitination licenses the Toll, the janus kinase (JAK)/signal

transducer and activator of transcription (STAT) and the immunodeficiency (IMD) pathways during immune challenge against bacterial, viral and fungal infection [5]. This immunological circuitry is not unique to *Drosophila* because ubiquitination also regulates these pathways in disease vectors, such as: *Anopheles gambiae*, *Aedes aegypti*, *Culex quinquefasciatus* and *Ixodes scapularis* [5].

The lack of information regarding the role of ubiquitin in disease vectors is problematic since molecular interactions depend on this posttranslational modification. These biochemical signatures may lead to novel factors to control pathogen transmission and/or acquisition [5]. For instance, drugs targeting the ubiquitin machinery for therapeutic development have entered clinical trials (<http://www.clinicaltrials.gov>) and have been approved by the Food and Drug Administration for use (e.g., Bortezomib - Millennium Pharmaceuticals and Nutlin - Roche) [5]. In this study, we used computational biology to identify the ubiquitin machinery of six clinically-relevant arthropod vectors: *An. gambiae*, *Ae. aegypti*, *C. quinquefasciatus*, *I. scapularis*, *Pediculus humanus* and *Rhodnius prolixus*. We compared our findings to datasets available for *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Mus musculus* and *Homo sapiens* [10-13]. While computational biology has been extensively used as a methodology in disease-causing arthropods for comparative genomics [14,15], transcriptomics [14,16-19] and quantitative proteomics [14,20,21], it has not been fully utilized to uncover the machinery associated with posttranslational modifications.

Materials and Methods

Protein and domain datasets

Complete protein datasets encoded within the genomes of *I. scapularis*, *An. gambiae*, *Ae. aegypti*, *C. quinquefasciatus*, *P. humanus* and *R. prolixus* were downloaded from VectorBase [22]. Protein datasets from *S. cerevisiae*, *D. melanogaster*, *H. sapiens* and *M. musculus* were downloaded from <http://www.yeastgenome.org> [23], <http://www.flybase.net> [24], and <http://ncbi.nih.gov> [25], respectively (Table S1). Based on a previously published study twenty-five Pfam domains for the ubiquitination machinery were selected [13], and downloaded from the Hidden Markov Model (HMM) library version 26.0 [26]. This included the following Pfam domains: APG12, PF04110; Atg8, PF02991; ubiquitin, PF00240; Ufm1, PF03671; Urm1, PF09138; ThiF, PF00899; UBACT, PF02134; UQ-con, PF00179; zf-C3HC4, PF00097; zf-Apc11, PF12861; RINGv, PF12906; Rtf2, PF04641; HECT, PF00632; Cullin, PF00888; U-box, PF04564; F-box, PF00646; OTU, PF02338; Josephin, PF02099; JAB, PF01398; DUF862, PF05903; WLM, PF08325; UCH, PF00443; Peptidase_C12, PF01088; Peptidase_C48, PF02902; and Peptidase_C54, PF03416.

Protein identification

Using the selected Pfam domains, the hmsearch program of the HMMER v3.0 package [27] was employed to identify analogous proteins within the chosen species (Tables S2, S3, S4, S5, S6, S7, S8, S9, S10). Proteins were assigned to the corresponding Pfam domains if searches returned an E-value

≤ 0.5 . Importantly, observed discrepancies between our results and findings previously published for *S. cerevisiae*, *D. melanogaster*, *H. sapiens* and *M. musculus* [10-13] were determined to result from one or more of the following: (1) changes made in the databases; (2) different programs and search parameters employed across studies; and (3) the high sensitivity of the relatively new HMMER 3.0 algorithm in detecting remotely related sequence similarities.

Computation of multiple alignments

The resulting protein sequences from *I. scapularis*, *An. gambiae*, *Ae. aegypti*, and *C. quinquefasciatus* were aligned using the Multiple Sequence Alignment tool MUSCLE [23] available on the Phylogeny.fr site [24]. The full mode option was utilized for the MUSCLE program and contained 3 stages: 1) draft progressive alignment, 2) improved progressive alignment, and 3) alignment refinement. Maximum number of iterations was 16, the default option. We did not remove any poorly aligned regions prior to phylogeny estimation because we were primarily interested in the relationship of proteins.

Generation of protein phylogenies

For each aligned dataset, protein phylogenies were estimated with the maximum likelihood method, employing the PhyML program [28] available through Phylogeny.fr. We presented cladograms, as opposed to phylograms to maintain readability of the large numbers of sequences considered in this study. Cladograms are also sufficient for illustrating the underlying function of related proteins, rather than inferring evolutionary relationships.

Identification of subfamilies/clusters

Protein sequences belonging to specific subcategories of the five ubiquitination-components (ubiquitin/ubiquitin-like proteins, ubiquitin-activating enzymes (E1), ubiquitin-conjugating enzymes (E2), ubiquitin ligases (E3) and deubiquitinases) were identified based on the Pfam domain matches obtained from hmsearch. Whenever possible, proteins were cataloged according to their major branching patterns in their respective phylogenies: ubiquitin/ubiquitin-like proteins (UB S30, small ubiquitin-like modifier (SUMO), neural precursor cell expressed developmentally downregulated gene 8 (NEDD8), autophagy-related protein 8 (ATG8), ubiquitin-related modifier-1 (URM1), homologous to ubiquitin 1 (HUB1), E1s (ubiquitin-activating enzyme 1 (UBA1), UBA2, UBA3, UBA1-like, autophagy-related E1-like enzyme (ATG7)), really interesting new gene (RING) (zf-C3HC4, zf-Apc11, RINGv, Rtf2), other E3s (homology to E6AP C-terminus (HECT), Cullin, U-box), and DUBs (ovarian tumor (OTU), Josephin, jun activation domain-binding protein (JAB), domain of unknown function (DUF) 862, Wss1p-like metalloproteases (WLM), ubiquitin carboxyl-terminal hydrolase (UCH), Peptidase_C12, Peptidase_C48, and Peptidase_C54). For direct classification of ubiquitin-like proteins, we manually analyzed and categorized each protein encoded in the genomes of *An. gambiae*, *Ae. aegypti*, *C. quinquefasciatus*, and *I. scapularis*. Ubiquitin-like proteins analyzed were SUMO, URM1, NEDD8, HUB1 and ATG8. Proteins containing two or more Pfam domains were observed and recorded in Table

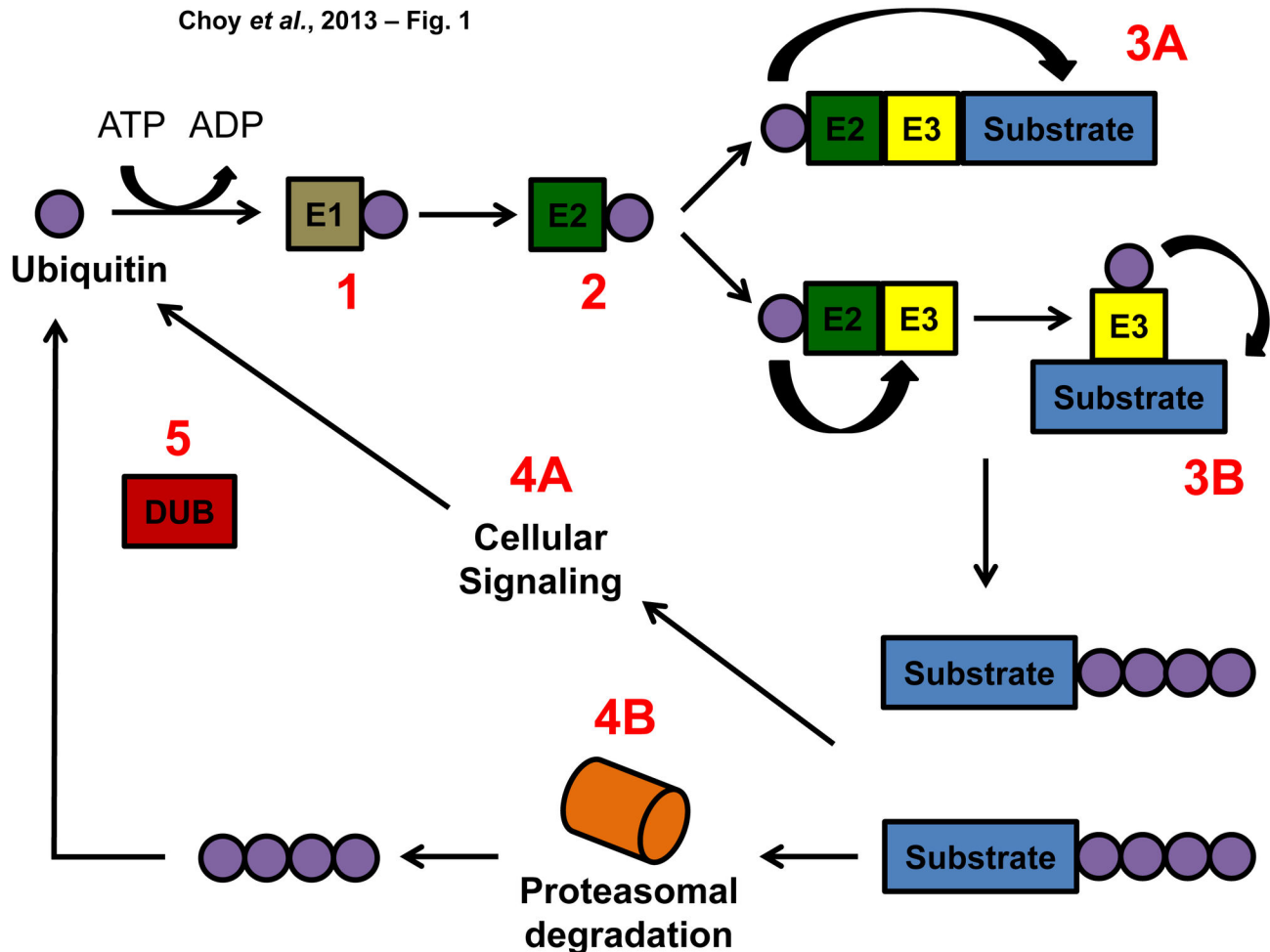


Figure 1. Schematic of the ubiquitination pathway. (1) A thiol ester bond between the E1-activating enzyme and ubiquitin is formed during ubiquitin activation. (2) The activated ubiquitin is then transferred to the E2-conjugating enzyme. (3) How ubiquitin is added onto a substrate varies according to the ubiquitin ligase (E3) category. (3A) RING-type E3 ligases function as a molecular scaffold that position the E2-conjugating enzyme with the substrate. (3B) HECT-type E3 ligases form an intermediary complex with ubiquitin leading to its transfer from the E2-conjugating enzyme to the substrate. (4A) Polyubiquitinated substrates may be targeted for cellular signaling or (4B) proteasomal degradation depending on the ubiquitin linkages (5). De-ubiquitinases (DUBs) recycle ubiquitin completing the cycle.

doi: 10.1371/journal.pone.0078077.g001

S11. Bootstrap values were used to evaluate the consistency of the cluster subcategories.

Results and Discussion

Composition of the ubiquitin machinery in model organisms versus disease vectors

The ubiquitin machinery regulates fundamental biological processes within eukaryotic cells (Figure 1). Thus, we used a bioinformatics approach to characterize this pathway in arthropod disease vectors. Our analysis demonstrated that the genomes of three clinically-relevant mosquitoes carried a similar ubiquitin pathway to total gene ratio (%U), ranging from 2.80 to 2.86 (Figure 2). *I. scapularis* contained the smallest %U

(1.79), while *P. humanus* had the highest ratio at 3.13. *R. prolixus* had a %U of 2.25. When compared to four model species, all disease vectors possessed a much lower %U, with the model species ranging from 4.38 to 5.02. The exception was *S. cerevisiae* (%U = 2.24). Interestingly, *D. melanogaster* carried a much higher %U at 4.50.

Ubiquitin

Using the ubiquitin Pfam domain, we analyzed the number of ubiquitin molecules reported for each of the selected species: *I. scapularis* (31), *An.gambiae* (40), *Ae. aegypti* (48), *C. quinquefasciatus* (40), *P. humanus* (35) and *R. prolixus* (129) (Table 1; Figure 3). We also analyzed the ubiquitin number in *S. cerevisiae* (7), *D. melanogaster* (30), *H. sapiens* (45) and *M.*

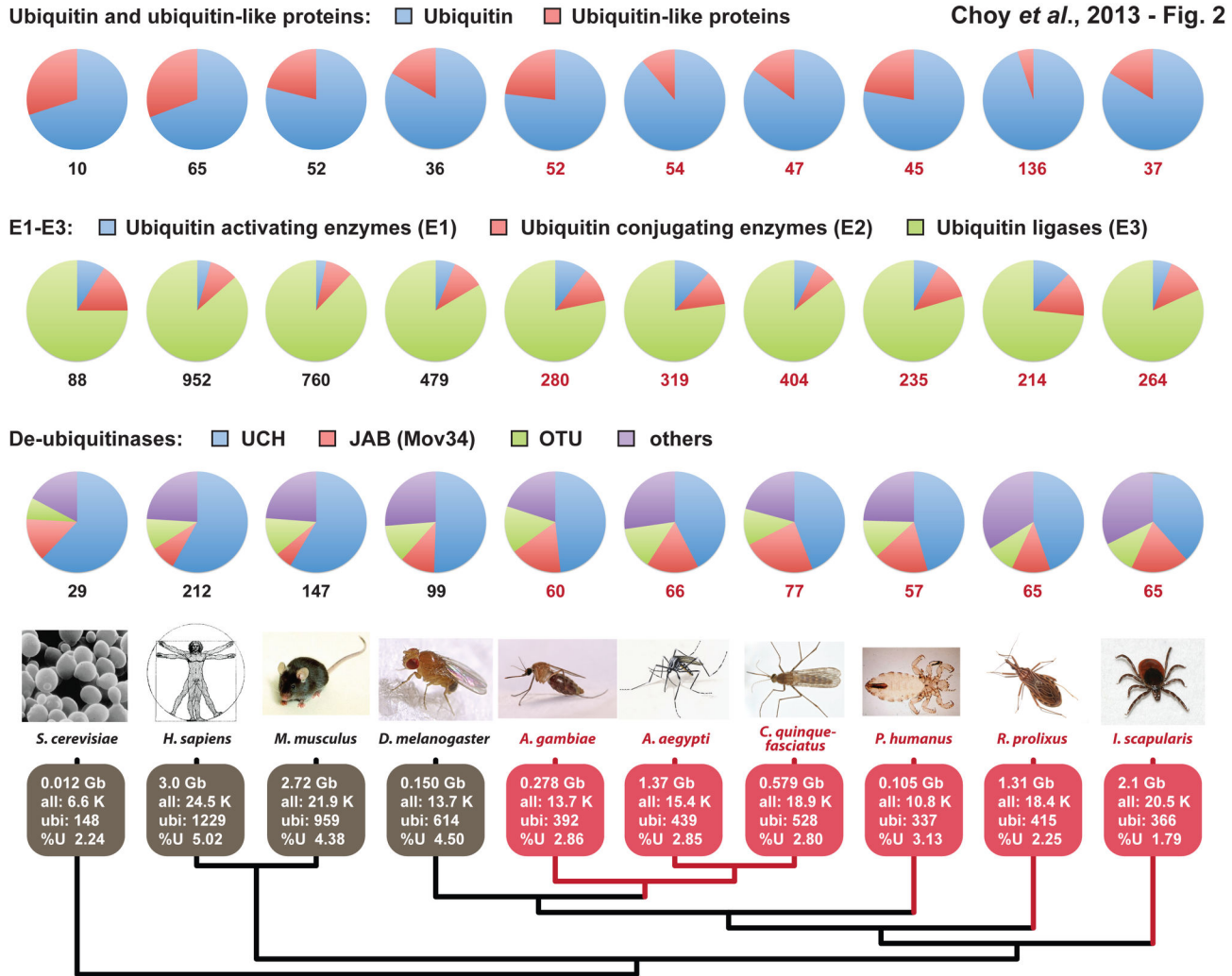


Figure 2. Proteins encoded for the ubiquitin machinery within the genomes of four model eukaryotes and six medically important arthropod vectors. Ten species are depicted at bottom, with model organisms colored gray (*S. cerevisiae*, *H. sapiens*, *M. musculus* and *D. melanogaster*) and arthropod vectors colored red (*A. gambiae*, *A. aegypti*, *C. quinquefasciatus*, *P. humanus*, *R. prolixus* and *I. scapularis*). The statistics shown are genome size in giga base pairs (Gb) and “all” referring to total number of predicted genes. “Ubi” refers to the total number of characterized (model organisms) or predicted (vectors) ubiquitin machinery proteins encoded within each genome, with “%U” the proportion of ubiquitin-related genes. Proteins were categorized into three major groups: Ubiquitin and ubiquitin-like proteins, E1-E3 enzymes, and de-ubiquitinases. For each genome, a breakdown within each major group depicts the composition of protein families (sub-groups).

doi: 10.1371/journal.pone.0078077.g002

musculus (41). Apart from *S. cerevisiae* (7) and *R. prolixus* (129), our analysis revealed a similar number of ubiquitin proteins across species. We further analyzed sequence similarities among ubiquitin family members (e.g. ubiquitin and ubiquitin-like proteins) for *An. gambiae*, *Ae. aegypti*, *C. quinquefasciatus*, and *I. scapularis* from their multiple alignments (Figure S1).

Ubiquitin-like proteins

Ubiquitin-like proteins were first discovered in 1979 and named after the interferon-stimulated gene 15 (ISG15) [29].

Since its discovery, several other ubiquitin-like proteins have been identified. In our study, we analyzed the paralogs for several ubiquitin-like proteins (Table 1; Figure 3). Specifically, we used Pfam hmmsearch domains for APG12, ATG8, UFM1 and URM1. For direct classification of ubiquitin-like proteins, we manually analyzed and categorized each protein encoded in the genomes of *An. gambiae*, *Ae. aegypti*, *C. quinquefasciatus*, and *I. scapularis* (Figure S1). Ubiquitin-like proteins analyzed were SUMO, NEDD8, URM1, HUB1 and ATG8.

Table 1. In silico ubiquitin machinery of medically relevant arthropod vectors when compared to known protein datasets* .

Domains/ Genomics	<i>I. scapularis</i>	<i>An. gambiae</i>	<i>Ae. aegypti</i>	<i>C. quinquefasciatus</i>	<i>P. humanus</i>	<i>R. prolixus</i>	<i>S. cerevisiae</i>	<i>D. melanogaster</i>	<i>H. sapiens</i>	<i>M. musculus</i>
Ubiquitin and ubiquitin-like proteins										
APG12	2	3	2	2	4	2	1	3	1	1
Atg8 (MAP1_LC3)	3	6	2	3	3	2	1	2	15	8
Ubiquitin	31	40	48	40	35	129	7	30	45	41
Ufm1 (UPF0185)	1	1	1	1	1	2	0	0	1	1
Urm1	0	2	1	1	2	1	1	1	3	1
Ubiquitin activating enzymes (E1)										
ThiF/UBACT	16	29	37	30	19	26	8	30	42	25
Ubiquitin conjugating enzymes (E2)										
UQ_con	32	32	36	28	29	31	14	49	87	67
Ubiquitin ligases (E3)										
RING	151	147	161	156	127	86	43	309	631	501
HECT	9	17	17	23	14	16	5	20	45	34
Cullin	5	6	9	4	7	8	4	14	16	11
U-Box	8	10	10	12	7	8	2	11	20	18
F-Box	43	39	49	151	32	39	12	46	111	104
De-ubiquitinases										
OTU	7	9	9	9	7	6	2	12	21	18
Josephin	2	1	1	2	2	2	0	1	13	10
JAB (Mov34)	12	10	11	18	10	8	4	11	17	8
DUF862	3	1	2	3	2	11	0	5	8	2
WLM	2	2	0	1	0	0	1	2	0	0
UCH	25	29	28	34	26	29	18	50	123	86
Peptidase_C12	4	2	7	3	4	3	1	6	7	6
Peptidase_C48	8	3	4	3	5	4	2	10	16	12
Peptidase_C54	2	3	4	4	1	2	1	2	7	5
Total	366	392	439	528	337	415	127	614	1229	959

Protein datasets were downloaded from online databases and matched with selected Pfam domains using the search program of the HMMER 3.0 package. Search results are listed individually.

doi: 10.1371/journal.pone.0078077.t001

One of the most extensively studied ubiquitin-like proteins is SUMO, a polypeptide with around 100 residues and 18% identity to ubiquitin [30]. SUMO modification has been known to regulate a multitude of biological functions. Among the most prominent are protein localization, gene expression, and DNA repair [31]. We observed a total of six proteins under the description of SUMO: *An. gambiae* (2), *Ae. aegypti* (2), *C. quinquefasciatus* (1), and *I. scapularis* (1) (Figure S1). NEDD8 is an ubiquitin-like protein with 81 amino acid residues [32]. It is around 60% identical to ubiquitin. Previous reports have identified the ligation of NEDD8 to members of the cullin family

protein, suggesting a critical role in cullin-RING E3 ubiquitin ligase (CRL) activation [33]. We observed four proteins categorized under the NEDD8 hmsearch description for *An. gambiae*, *Ae. aegypti*, *C. quinquefasciatus* and *I. scapularis* - one from each vector (Figure S1).

URM1 is a ubiquitin-like protein that shares very little homology with ubiquitin [34]. Four proteins were categorized under the URM1 description for *An. gambiae*, *Ae. aegypti*, *C. quinquefasciatus*, and *I. scapularis* (Table 1). HUB1 differs from ubiquitin in that instead of a C-terminal tail with a GG motif, it carries a double tyrosine (YY) motif [35]. In our

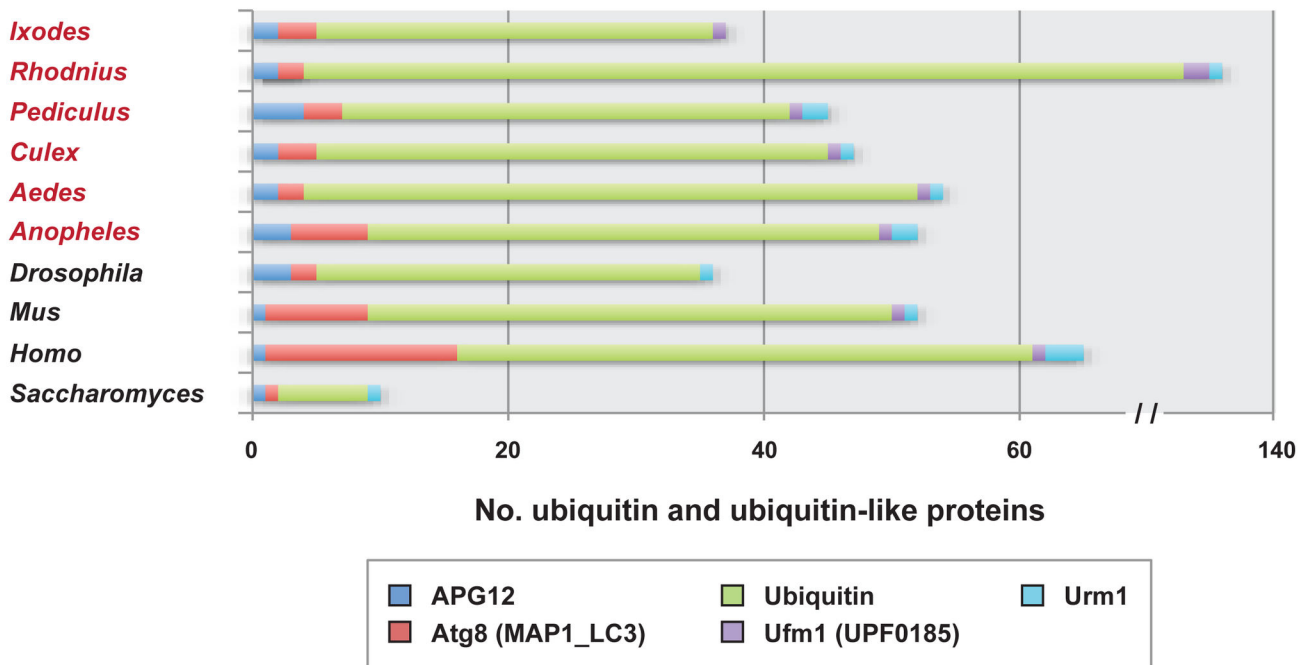
Choy *et al.*, 2013 - Fig. 3

Figure 3. Composition of ubiquitin and ubiquitin-like proteins within the genomes of four model eukaryotes and six medically important arthropod vectors. Using HMMER v.3.0, complete sets of proteins for all ten genomes were scanned for the presence of five Pfam domain models for ubiquitin and ubiquitin-like proteins: APG12 (PF04110), Atg8 (PF02991), ubiquitin (PF00240), Ufm1 (PF03671), and Urm1 (PF09138). Only protein matches to Pfam models with E-values <0.05 were selected for compilations.

doi: 10.1371/journal.pone.0078077.g003

analysis, we reported four HUB1 proteins within *An. gambiae*, *Ae. aegypti*, *C. quinquefasciatus* and *I. scapularis*, with one being identified in each vector (Figure S1). ATG8 is one of the most extensively studied ubiquitin-like proteins and has been identified as being a key component in the regulation of autophagy [36]. Of the ubiquitin-like proteins examined, the number of proteins similar to ATG8 was ranked the highest: *An. gambiae* (6), *Ae. aegypti* (2), *C. quinquefasciatus* (3), and *I. scapularis* (3) (Table 1).

Ubiquitin-activating enzymes (E1)

There are several classes of ubiquitin-activating enzymes, distinguishable by their ubiquitin-activating (UBA) domain. The E1 associated to ubiquitin is known as UBA1. UBA2 and UBA3 serve as activators for SUMO and NEDD8 respectively [13]. The E1 for URM1 is UBA4 [37]. ATG7, the E1 associated to ATG8, is recognized by an N-terminal motif [38]. UBA1-like, is slightly larger than UBA1 [13]. Using its Pfam domain, we analyzed the number of E1 molecules reported for each of the selected species: *I. scapularis* (16), *An. gambiae* (29), *Ae. aegypti* (37), *C. quinquefasciatus* (30), *P. humanus* (19) and *R. prolixus* (26) (Table 1; Figure 4). The three mosquitoes (*An. gambiae*, *Ae. aegypti*, and *C. quinquefasciatus*) have a very similar number of E1 proteins, while *I. scapularis* and *P.*

humanus have noticeably less E1s than the other disease vectors. We also analyzed the E1 number in *S. cerevisiae* (8), *D. melanogaster* (30), *H. sapiens* (42), and *M. musculus* (25) (Table 1). We further analyzed the proteins identified as E1 according to their subcategories for *An. gambiae*, *Ae. aegypti*, *C. quinquefasciatus* and *I. scapularis* through phylogeny estimation (Figure S2).

Ubiquitin-conjugating enzymes (E2)

E2s consists of a conserved domain with approximately 150 amino acids [39]. We analyzed the number of E2 molecules reported for *I. scapularis* (32), *An. gambiae* (32), *Ae. aegypti* (36), *C. quinquefasciatus* (28), *P. humanus* (29) and *R. prolixus* (31) (Table 1; Figure 4). The number of E2 proteins in each vector is very similar, ranging from 29-36. Interestingly, there are slightly more E1s than E2s for *A. aegypti* and *C. quinquefasciatus*. We further compared the E2 proteins in *An. gambiae*, *Ae. aegypti*, *C. quinquefasciatus*, and *I. scapularis* (Figure S3).

Ubiquitin ligases (E3)

RING domain E3 ligase. HMM searches were performed on selected vectors and the total number of RING E3s was calculated: *I. scapularis* (151), *An.gambiae* (147), *Ae. aegypti*

Choy et al., 2013 - Fig. 4

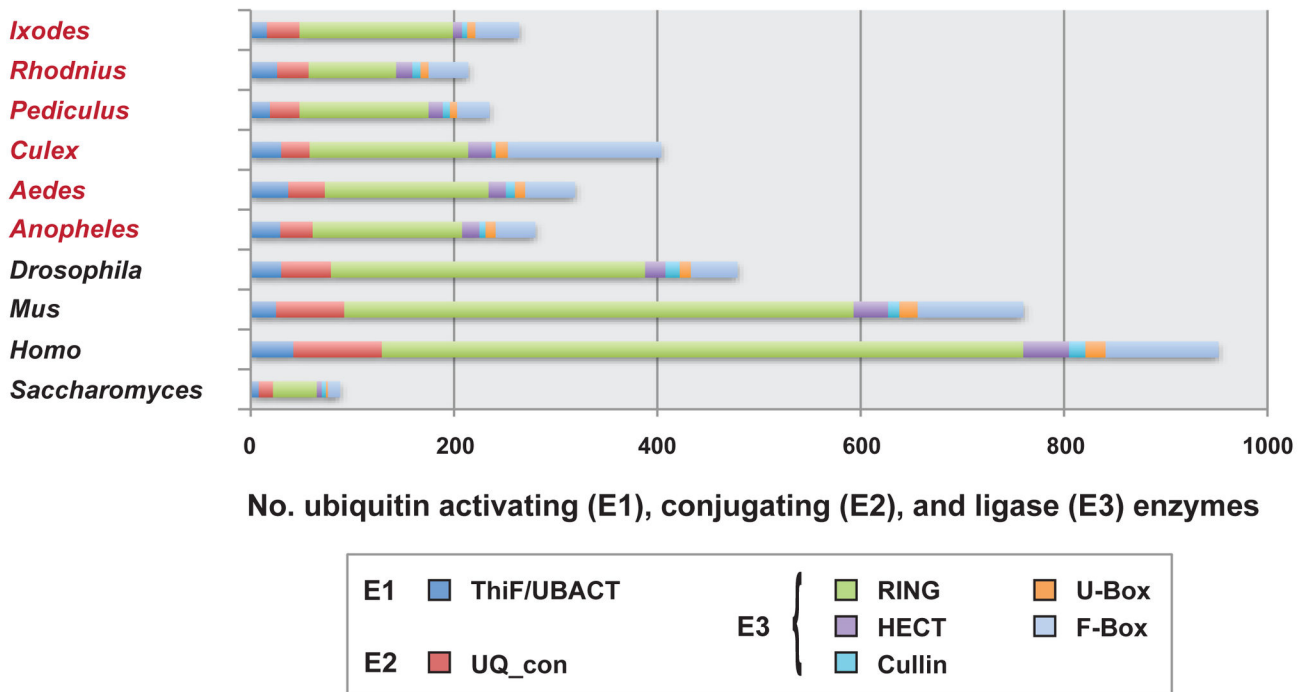


Figure 4. Composition of ubiquitin-activating enzymes (E1), ubiquitin conjugating enzymes (E2), and ubiquitin ligases (E3) within the genomes of four model eukaryotes and six medically important arthropod vectors. Using HMMER v.3.0, complete sets of proteins for all ten genomes were scanned for the presence of two Pfam domain models for ubiquitin activating enzymes (E1): ThiF (PF00899) and UBACT (PF02134); one Pfam domain model for ubiquitin conjugating enzymes (E2): UQ-con (PF00179); and eight Pfam domain models for ubiquitin ligases (E3): zf-C3HC4 (PF00097), zf-Apc11 (PF12861), RINGv (PF12906), Rtf2 (PF04641), HECT (PF00632), Cullin (PF00888), U-box (PF04564), and F-box (PF00646). Only protein matches to Pfam models with E-values <0.05 were selected for compilations.

doi: 10.1371/journal.pone.0078077.g004

(161), *C. quinquefasciatus* (156), *P. humanus* (127) and *R. prolixus* (86) (Table 1; Figure 4). The number of RING E3s found in each protein dataset is significantly higher than any other component in the ubiquitin machinery. We annotated four Pfam domains associated with RING E3 ligases: zf-C3HC4, zf-Apc11, RINGv, and Rtf2. The zf-C3HC4 domain contains the amino acid motif pattern Cys₃HisCys₄. The zf-C3HC4 subcategory was the most common RING-type ligase within selected vectors: *I. scapularis* (100), *An. gambiae* (82), *Ae. aegypti* (84), *C. quinquefasciatus* (87), *P. humanus* (80) and *R. prolixus* (56) (Table S2-7).

We report the number of zf-Apc11 proteins in *I. scapularis* (28), *An. gambiae* (42), *Ae. aegypti* (43), *C. quinquefasciatus* (47), *P. humanus* (31) and *R. prolixus* (19) (Table S2-7). RINGv proteins were observed in all six vectors: *I. scapularis* (17), *An. gambiae* (12), *Ae. aegypti* (22), *C. quinquefasciatus* (12), *P. humanus* (16) and *R. prolixus* (11) (Table S2-7). Rtf2 contains a C₂HC₂ motif and plays a role in DNA replication [40]. We analyzed the number of Rtf2 within each vector: *I. scapularis* (6), *An. gambiae* (11), *Ae. aegypti* (12), *C. quinquefasciatus* (10), *P. humanus* (6) and *R. prolixus* (8) (Table S2-7). It is important to note that some RING proteins

contained two or more RING domains and were annotated based on the subcategory with the lowest E-value (Table S11).

We further analyzed the RING E3s in *An. gambiae*, *Ae. aegypti*, *C. quinquefasciatus*, and *I. scapularis* by phylogeny estimation (Figure S4, S5). Because of computational limitations and the large number of RING proteins, we created a separate analysis for each of the four vectors. We also identified clusters related to each RING Pfam domain within the phylogeny (Figure S4, S5).

HECT domain E3 ligase. We analyzed the number of HECT E3s in *I. scapularis* (9), *An. gambiae* (17), *Ae. aegypti* (17), *C. quinquefasciatus* (23), *P. humanus* (14) and *R. prolixus* (16) (Table 1; Figure 4). The number of HECT E3s observed in each species is significantly lower than the number of RING ligases. We further examined the HECT E3s against other E3 ligases in *An. gambiae*, *Ae. aegypti*, *C. quinquefasciatus*, and *I. scapularis* by phylogeny estimation (Figure S6).

Cullin. Cullin proteins form a RING E3 in order to create a scaffold that facilitates the ligation of ubiquitin to the target substrate [41]. We calculated the number of proteins associated with Cullin in *I. scapularis* (5), *An. gambiae* (6), *Ae. aegypti* (9), *C. quinquefasciatus* (4), *P. humanus* (7) and *R.*

prolixus (8) (Table 1; Figure 4). The number of Cullin proteins is the smallest amongst the subcategories of ubiquitin ligases within each vector. This pattern is consistent with model species: *S. cerevisiae* (4), *D. melanogaster* (14), *H. sapiens* (16), and *M. musculus* (11) (Table 1). We then compared the Cullin proteins within *An. gambiae*, *Ae. aegypti*, *C. quinquefasciatus*, and *I. scapularis* against other E3 ligases (Figure S6). Similar to HECT ligases, most of the Cullin protein clusters were separated in the estimated tree.

U-box Ubiquitin Ligase. In addition to RING, HECT, and Cullin, recent studies have identified proteins that share many similarities with the RING finger, but lack the metal-chelating residue and many signature Cys residues in the original RING domain [42]. This protein family has been classified as the U-box, although it has previously been identified as E4 [43]. Here, we report the number of U-box ubiquitin ligases in *I. scapularis* (8), *An.gambiae* (10), *Ae. aegypti* (10), *C. quinquefasciatus* (12), *P. humanus* (7) and *R. prolixus* (8) (Table 1). Much like Cullin, U-box proteins account for only a small ratio of the total E3 ligases in each vector. When comparing U-boxes with other E3 ligases, we observed only two clusters of U-box proteins within the estimated tree (Figure S6).

F-box. The F-box domain is categorized by a structural motif that contains approximately 50 amino acids [44]. We used hmsearch to reveal the following numbers of F-box proteins in the various disease vectors: *I. scapularis* (43), *An. gambiae* (39), *Ae. aegypti* (49), *C. quinquefasciatus* (151), *P. humanus* (32) and *R. prolixus* (39) (Table 2; Figure 4). The number of proteins associated with F-box is the second largest component for five out of six arthropod vectors. In model species, we observed a similar increase in the number of F-box and most other ubiquitination components: *S. cerevisiae* (12), *D. melanogaster* (46), *H. sapiens* (111), and *M. musculus* (104) (Table 1). We then extracted the sequences of F-box proteins in *An. gambiae*, *Ae. aegypti*, *C. quinquefasciatus*, and *I. scapularis* and estimated phylogenies for each species (Figure S7, S8).

Deubiquitinases (DUBs)

Ovarian tumor domain (OTU). OTU DUBs have been known to hydrolyze K48, K63, and polyubiquitin chains [45]. Our study uncovered the number of OTU-associated proteins in the six selected vectors and four model organisms as: *I. scapularis* (7), *An. gambiae* (9), *Ae. aegypti* (9), *C. quinquefasciatus* (9), *P. humanus* (7), *R. prolixus* (6), *S. cerevisiae* (2), *D. melanogaster* (12), *H. sapiens* (21), and *M. musculus* (18) (Table 1; Figure 5). In most species, OTU-associated proteins were the third most common DUBs, accounting for approximately 12.5% of the total number of DUBs. Interestingly, the lowest ratio of OTU to DUBs was found in *H. sapiens*, where OTU accounted for only 9.9%. We further analyzed relatedness between OTU and other DUBs through estimated phylogenies across proteins from *An. gambiae*, *Ae. aegypti*, *C. quinquefasciatus*, and *I. scapularis* (Figure S9, S10).

Josephin. The Josephin domain is a highly conserved catalytic motif containing approximately 180 residues and was named after Machado-Joseph disease (MJD), a

neurodegenerative condition [46]. We reported the number of Josephin-associated proteins in *I. scapularis* (2), *An. gambiae* (1), *Ae. aegypti* (1), *C. quinquefasciatus* (2), *P. humanus* (2), *R. prolixus* (2), *S. cerevisiae* (0), *D. melanogaster* (1), *H. sapiens* (13), and *M. musculus* (10) (Table 1; Figure 5). All selected vectors carried a very limited number of proteins containing the Josephin domain, ranging from 1-2. Interestingly, *H. sapiens* and *M. musculus* contained a significantly larger number of Josephin-related proteins. We then analyzed Josephin-associated DUBs against other DUBs for *An. gambiae*, *Ae. aegypti*, *C. quinquefasciatus*, and *I. scapularis* (Figure S9, S10).

Jun activation domain-binding protein (JAB). The JAB domain is a component in a group of metalloproteases known as Jab1/MPN metalloenzyme (JAMM) DUBs [47]. This specific type of DUBs targets K63 polyubiquitin chains [48]. We analyzed the number of JAB-associated DUBs found within selected disease vectors: *I. scapularis* (12), *An. gambiae* (10), *Ae. aegypti* (11), *C. quinquefasciatus* (18), *P. humanus* (10) and *R. prolixus* (8) (Table 1; Figure 5). Proteins containing the JAB domain accounted for the second most abundant DUBs. Interestingly, the ratio of JAB-associated DUBs to total DUBs was much lower in *H. sapiens* and *M. musculus* than in the six vectors. Furthermore, the total number of proteins containing JAB found in *M. musculus* was smaller than any of those found in the analyzed vectors. We then compared JAB-containing DUBs with other DUBs in *An. gambiae*, *Ae. aegypti*, *C. quinquefasciatus*, and *I. scapularis* (Figure S9, S10).

Domain of Unknown Function 862 (DUF862). We identified the number of DUF862-containing DUBs within *I. scapularis* (3), *An.gambiae* (1), *Ae. aegypti* (2), *C. quinquefasciatus* (3), *P. humanus* (2) and *R. prolixus* (11) (Table 1; Figure 5). The number of proteins displaying the DUF862 domain was relatively small. The lone exception was *R. prolixus*, where DUF862 DUBs account for the second highest number of DUBs. We then compared the DUF862-containing proteins with other DUBs in *An. gambiae*, *Ae. aegypti*, *C. quinquefasciatus*, and *I. scapularis* by constructing estimated phylogenies (Figure S9, S10).

Wss1p-like metalloproteases (WLM). The WLM domain is the defining characteristic of a family within Zn-dependent peptidases, termed the WLM DUBs. Although its functions are not fully elucidated, WLM DUBs have been associated with deSUMOylation [49]. WLM was the rarest DUB found within all six vectors: *I. scapularis* (2), *An. gambiae* (2), *Ae. aegypti* (0), *C. quinquefasciatus* (1), *P. humanus* (0), and *R. prolixus* (0) (Table 1; Figure 5).

Ubiquitin carboxyl-terminal hydrolases (UCH). DUBs containing the UCH domain have been known to release ubiquitin in polyubiquitin chains by hydrolyzing its C-terminus [50]. We performed an hmsearch for the total numbers of UCH DUBs within *I. scapularis* (25), *An. gambiae* (29), *Ae. aegypti* (28), *C. quinquefasciatus* (34), *P. humanus* (26) and *R. prolixus* (29) (Table 1; Figure 5). UCH-associated proteins account for the largest number of DUBs within every vector. The average ratio between UCH-containing proteins to total DUBs for all six vectors was approximately 41%, with the lowest being *I. scapularis* at 38%. In the estimated trees

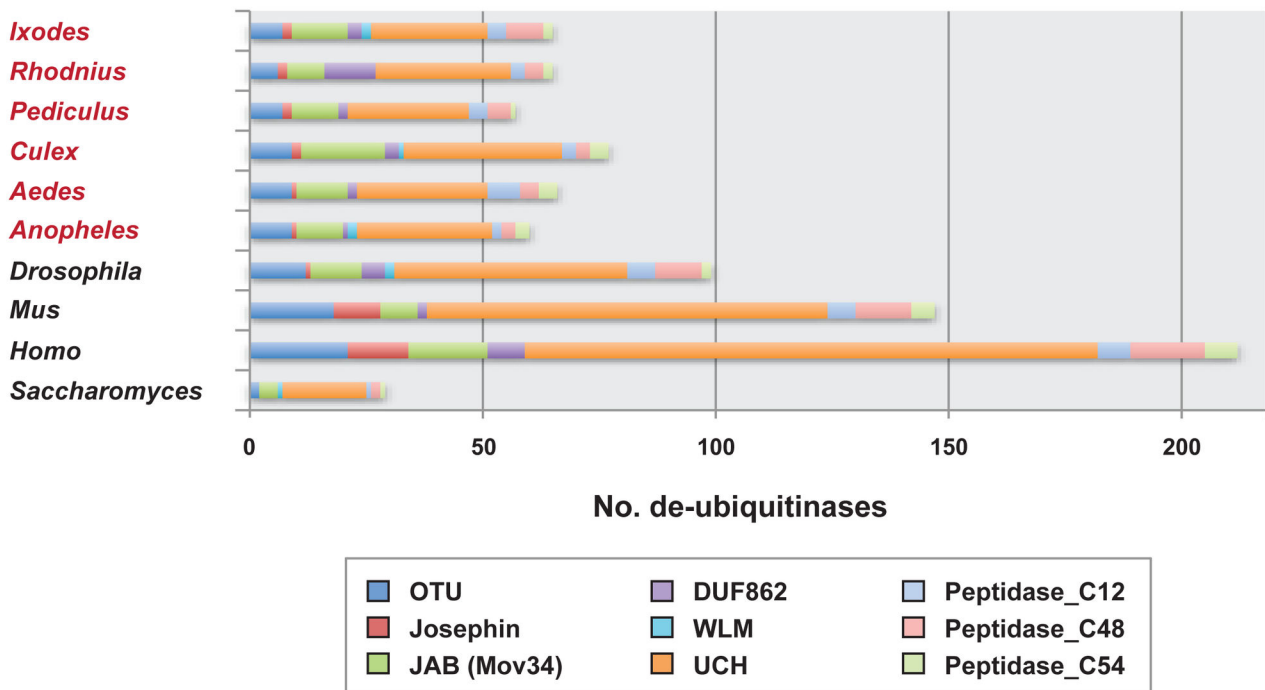
Choy *et al.*, 2013 - Fig. 5

Figure 5. Composition of de-ubiquitinases within the genomes of four model eukaryotes and six medically important arthropod vectors. Using HMMER v.3.0, complete sets of proteins for all ten genomes were scanned for the presence of nine Pfam domain models for de-ubiquitinases: OTU (PF02338), Josephin (PF02099), JAB (PF01398), DUF862 (PF05903), WLM (PF08325), UCH (PF00443), Peptidase_C12 (PF01088), Peptidase_C48 (PF02902), and Peptidase_C54 (PF03416). Only protein matches to Pfam models with E-values <0.05 were selected for compilations.

doi: 10.1371/journal.pone.0078077.g005

comparing UCH DUBs with other DUBs, we observed several UCH clusters for each vector: *An. gambiae* (6), *Ae. aegypti* (6), *C. quinquefasciatus* (8), and *I. scapularis* (5). (Figure S9, S10).

Peptidase_C12, Peptidase_C48, Peptidase_C54 domains. Peptidase_C12 has been linked to the UCH family of DUBs, while the Peptidase_C48 domain is a C-terminal catalytic domain associated with the Ubl-specific protease 1 (ULP1) protease family [51,52]. Peptidase_C54 has not been previously studied, although our analysis has identified it as a domain similar to a DUF related to URM1 proteases [53]. We analyzed the number to DUBs containing each of the three peptidase domains: *I. scapularis* (Peptidase_C12: 4, Peptidase_C48: 8, Peptidase_C54: 2), *An. gambiae* (Peptidase_C12: 2, Peptidase_C48: 3, Peptidase_C54: 3), *Ae. aegypti* (Peptidase_C12: 7, Peptidase_C48: 4, Peptidase_C54: 4), *C. quinquefasciatus* (Peptidase_C12: 3, Peptidase_C48: 3, Peptidase_C54: 4), *P. humanus* (Peptidase_C12: 4, Peptidase_C48: 5, Peptidase_C54: 1), *R. prolixus* (Peptidase_C12: 3, Peptidase_C48: 4, Peptidase_C54: 2), *S. cerevisiae* (Peptidase_C12: 1, Peptidase_C48: 2, Peptidase_C54: 1), *D. melanogaster* (Peptidase_C12: 6, Peptidase_C48: 10, Peptidase_C54: 2), *H. sapiens* (Peptidase_C12: 7, Peptidase_C48: 16, Peptidase_C54: 7), and *M. musculus* (Peptidase_C12: 6, Peptidase_C48: 12,

Peptidase_C54: 5) (Table 1; Figure 5). In most analyzed species, the number of Peptidase_C48 DUBs were the highest among the peptidase domain DUBs. The two exceptions were *Ae. aegypti*, where the number of Peptidase_C12-associated proteins was the highest, and *C. quinquefasciatus*, which contained one more deubiquitinase with a Peptidase_C54 domain. Each species possessed at least one DUB containing each of the peptidase domains. When analyzing the DUBs, we observed differences in the number of clusters formed for each vector (Figure S9, S10).

Conclusion

Vector-borne illnesses threaten public health in the tropics. Environmental changes due to globalization and the lack of effective vaccines are also contributing to the spread of these maladies to temperate climates. Thus, novel treatments are necessary in the clinic. One effective therapeutic strategy recently used to treat cancer, neurodegenerative disorders and some infectious diseases is pharmacological intervention targeting (de)ubiquitination. However, the use of (de)ubiquitination as a therapeutic target to combat arthropod-borne diseases has not been established. A possible reason for this gap in translational research is the inherent complexity

of vector-borne diseases. Another possibility is the lack of communication among scientists. Pharmacologists and chemists do not typically interact with vector biologists, and vector biologists have little incentive to interact with the drug development community.

Here we took advantage of publicly available arthropod genomes and described the ubiquitin machinery of *An. gambiae*, *Ae. aegypti*, *C. quinquefasciatus*, *I. scapularis*, *P. humanus* and *R. prolixus*. Although independent work will be necessary to validate our bioinformatics analysis, empirical evidence suggests that the ubiquitin machinery is present in disease vectors. Recently, Severo et al., 2013 characterized a RING-type E3 ubiquitin ligase named x-linked inhibitor of apoptosis protein (XIAP) and showed the importance of ubiquitination for microbial pathogenesis in ticks [54]. Similarly, targeting of genes by RNAi from the ubiquitin-mediated pathway affected bacterial infection in arthropod vectors [55]. Corroborating with these findings, Huang and colleagues demonstrated that monoubiquitinated proteins decorate pathogen-occupied vacuolar membranes during infection of embryonic tick cells [56]. In summary, this report should: (1) provide a framework for studying ubiquitination in disease vectors; (2) generate a basis for empirical experimentation correlating arthropod physiology and disease; and (3) potentially unveil novel pharmacological targets to interfere with vector-borne diseases.

Supporting Information

Table S1. Source and release date of protein datasets.
(PDF)

Table S2. List of proteins in *An. gambiae* identified by the HMM search.
(XLSX)

Table S3. List of proteins in *Ae. aegypti* identified by the HMM search.
(XLSX)

Table S4. List of proteins in *C. quinquefasciatus* identified by the HMM search.
(XLSX)

Table S5. List of proteins in *I. scapularis* identified by the HMM search.
(XLSX)

Table S6. List of proteins in *P. humanus* identified by the HMM search.
(XLSX)

Table S7. List of proteins in *R. prolixus* identified by the HMM search.
(XLSX)

Table S8. List of proteins in *D. melanogaster* identified by the HMM search.
(XLSX)

Table S9. List of proteins in *H. sapiens* identified by the HMM search.
(XLSX)

Table S10. List of proteins in *M. musculus* identified by the HMM search.
(XLSX)

Table S11. List of protein identifications containing sequences matching two or more Pfam domains.
(XLSX)

Figure S1. Ubiquitin and ubiquitin-like proteins in *I. scapularis*, *An. gambiae*, *Ae. aegypti*, and *C. quinquefasciatus*. Protein sequences matched with ubiquitin and ubiquitin-like proteins were aligned using MUSCLE and a phylogeny was estimated with the PhyML software. Sequences with a bootstrap value lower than 50 were manually removed. Bootstrap values ranged from 0.64 - 0.86 for highlighted clustered categories (SUMO, ATG8, UB S30, NEDD8, HUB1, URM1).
(TIF)

Figure S2. Ubiquitin and ubiquitin-like activating enzymes in *I. scapularis*, *An. gambiae*, *Ae. aegypti*, and *C. quinquefasciatus*. Phylogeny of ubiquitin and ubiquitin-like activating enzymes. Matched sequences for ubiquitin and ubiquitin-like activating enzymes were aligned using MUSCLE and a phylogeny was estimated with the PhyML software. Bootstrap values ranged from 0.43 - 0.83 for highlighted clustered categories (UBA1, UBA2, UBA3, ATG7, and UBA1-like).
(TIF)

Figure S3. Ubiquitin and ubiquitin-like conjugating enzymes in *I. scapularis*, *An. gambiae*, *Ae. aegypti*, and *C. quinquefasciatus*. Protein sequences matched with ubiquitin and ubiquitin-like conjugating enzymes were aligned using MUSCLE and a phylogeny was estimated with the PhyML software.
(TIF)

Figure S4. Phylogeny of RING and RING-like ubiquitin ligases in *An. gambiae* and *Ae. aegypti*. Protein sequences matched with ubiquitin and ubiquitin-like ligases were aligned using MUSCLE and a phylogeny was estimated with the PhyML software. Proteins not listed under a specific subset were categorized as zf-C3HC4. Phylogeny of RING and RING-like ubiquitin ligases in (A) *An. gambiae* and (B) *Ae. aegypti*. Bootstrap values ranged from (A) 0.41 - 0.80 and (B) 0.41 - 0.83 for highlighted clustered categories (RINGv, Rtf2, zf-Apc11).
(TIF)

Figure S5. Phylogeny of RING and RING-like ubiquitin ligases in *C. quinquefasciatus* and *I. scapularis*. Protein sequences matched with ubiquitin and ubiquitin-like ligases were aligned using MUSCLE and a phylogeny was estimated with the PhyML software. Proteins not listed under a specific subset were categorized into the zf-C3HC4 classification. Phylogeny of RING and RING-like ubiquitin ligases in (A) *C. quinquefasciatus* and (B) *I. scapularis*. Bootstrap values ranged from (A) 0.43 - 0.83 and (B) 0.46 - 0.83 for highlighted clustered categories (RINGv, Rtf2, zf-Apc11). (TIF)

Figure S6. HECT, Cullin and U-box ligases in *I. scapularis*, *An. gambiae*, *Ae. aegypti*, and *C. quinquefasciatus*. Protein sequences matched with ubiquitin and ubiquitin-like ligases were aligned using MUSCLE and the phylogeny was estimated with the PhyML software. Sequences with bootstrap values lower than 50 were manually removed. Bootstrap values ranged from 0.35 - 0.88 for highlighted clustered categories (HECT, Cullin, U-Box). (TIF)

Figure S7. Phylogenetic trees of F-box ubiquitin ligases in *An. gambiae* and *Ae. aegypti*. Protein sequences matched with F-box ligases were aligned using MUSCLE and the phylogeny was estimated with the PhyML software. Sequences with a bootstrap values lower than 50 were manually removed. Phylogenetic trees of F-box ubiquitin ligases in (A) *An. gambiae* and (B) *Ae. aegypti*. (TIF)

Figure S8. Phylogenetic trees of F-box ubiquitin ligases in *C. quinquefasciatus* and *I. scapularis*. Protein sequences matched with F-box ligases were aligned using MUSCLE and the phylogeny was estimated with the PhyML software. Sequences with a bootstrap values lower than 50 were manually removed. Phylogenetic trees of F-box ubiquitin ligases in (A) *C. quinquefasciatus* and (B) *I. scapularis*.

References

- (2004) The World. Health Reports: 2004 - Changing History. World Health Organization.
- (2010) Lymphatic Filariasis. Centers for Disease Control and Prevention.
- (2010) Lyme Disease. Centers for Disease Control and Prevention.
- McGraw EA, O'Neill SL (2013) Beyond insecticides: new thinking on an ancient problem. *Nat Rev Microbiol* 11: 181-193. doi:10.1038/nrmicro2968. PubMed: 23411863.
- Severo MS, Sakhon OS, Choy A, Stephens KD, Pedra JHF (2013) The 'ubiquitous' reality of vector immunology. *Cell Microbiol* 15: 1070-1078. doi:10.1111/cmi.12128. PubMed: 23433059.
- Low TY, Magliozzi R, Guardavaccaro D, Heck AJ (2013) Unraveling the ubiquitin-regulated signaling networks by mass spectrometry-based proteomics. *Proteomics* 13: 526-537. doi:10.1002/pmic.201200244. PubMed: 23019148.
- Collins CA, Brown EJ (2010) Cytosol as battleground: ubiquitin as a weapon for both host and pathogen. *Trends Cell Biol* 20: 205-213. doi: 10.1016/j.tcb.2010.01.002. PubMed: 20129784.
- Baron M (2012) Endocytic routes to Notch activation. *Semin Cell Dev Biol* 23: 437-442. doi:10.1016/j.semcdb.2012.01.008. PubMed: 22285298.
- Jaiswal M, Sandoval H, Zhang K, Bayat V, Bellen HJ (2012) Probing mechanisms that underlie human neurodegenerative diseases in *Drosophila*. *Annu Rev Genet* 46: 371-396. doi:10.1146/annurev-genet-110711-155456. PubMed: 22974305.
- Matsumoto M, Hatakeyama S, Oyama K, Oda Y, Nishimura T et al. (2005) Large-scale analysis of the human ubiquitin-related proteome. *Proteomics* 5: 4145-4151. doi:10.1002/pmic.200401280. PubMed: 16196087.
- Peng J, Cheng D (2005) Proteomic analysis of ubiquitin conjugates in yeast. 399: 367-381. doi:10.1016/S0076-6879(05)99025-3. PubMed: 16338369.
- Sample CA, Group RG, Members GSL (2003) The comparative proteomics of ubiquitination in mouse. *Genome Res* 13: 1389-1394
- Ponts N, Yang J, Chung D-WD, Prudhomme J, Girke T et al. (2008) Deciphering the ubiquitin-mediated pathway in apicomplexan parasites: a potential strategy to interfere with parasite virulence. *PLOS ONE* 3: e2386. doi:10.1371/journal.pone.0002386. PubMed: 18545708.
- Severson DW, Behura SK (2012) Mosquito genomics: progress and challenges. *Annu Rev Entomol* 57: 143-166. doi:10.1146/annurev-ento-120710-100651. PubMed: 21942845.

(TIF)

Figure S9. Phylogenetic trees of deubiquitinases in *An. gambiae* and *Ae. aegypti*. Protein sequences were aligned using MUSCLE and a phylogeny was estimated using the maximum likelihood method. Phylogeny of deubiquitinases in (A) *An. gambiae* and (B) *Ae. aegypti*. Bootstrap values ranged from (A) 0.43 - 0.87 and (B) 0.37 - 0.84 for highlighted clustered categories (UCH, OTU, WLM, JAB, Peptidase_C12, Peptidase_C54, Peptidase_C48, Josephin and DUF862). (TIF)

Figure S10. Phylogenetic trees of deubiquitinases in *C. quinquefasciatus* and *I. scapularis*. Protein sequences were aligned using MUSCLE and a phylogeny was estimated using the maximum likelihood method. Phylogeny of deubiquitinases in (A) *C. quinquefasciatus* and (B) *I. scapularis*. Bootstrap values ranged from (A) 0.39 - 0.83 and (B) 0.37 - 0.78 for highlighted clustered categories (UCH, OTU, WLM, JAB, Peptidase_C12, Peptidase_C54, Peptidase_C48, Josephin and DUF862). (TIF)

Acknowledgements

We thank Abdu Azad and Lindsey Brown (University of Maryland School of Medicine) for scientific discussions, Carmen Wong (University of California, Riverside) for preparation of supplemental figures and Olivia Sakhon (University of Maryland School of Medicine) for illustrations. We also acknowledge the Institute for Integrative Genome Biology Bioinformatics Facility at the University of California, Riverside.

Author Contributions

Conceived and designed the experiments: AC TG JJJ JHFP. Performed the experiments: AC MSS RS JJJ. Analyzed the data: AC RS TG JJJ JHFP. Contributed reagents/materials/analysis tools: RS TG. Wrote the manuscript: AC JHFP.

15. Neafsey DE, Christophides GK, Collins FH, Emrich SJ, Fontaine MC, et al. (2013) The evolution of the *Anopheles* 16 genomes project. *G3* 3:1191-1194 PubMed: 23708298
16. David JP, Coissac E, Melodelima C, Poupardin R, Riaz MA et al. (2010) Transcriptome response to pollutants and insecticides in the dengue vector *Aedes aegypti* using next-generation sequencing technology. *BMC Genomics* 11: 216. doi:10.1186/1471-2164-11-216. PubMed: 20356352.
17. Francischetti IM, Sa-Nunes A, Mans BJ, Santos IM, Ribeiro JM (2009) The role of saliva in tick feeding. *PLoS One* 14: 2051-2088. PubMed: 19273185.
18. Mans BJ, Andersen JF, Francischetti IM, Valenzuela JG, Schwan TG et al. (2008) Comparative sialomics between hard and soft ticks: implications for the evolution of blood-feeding behavior. *Insect Biochem Mol Biol* 38: 42-58. doi:10.1016/j.ibmb.2007.09.003. PubMed: 18070664.
19. Bartholomay LC, Waterhouse RM, Mayhew GF, Campbell CL, Michel K et al. (2010) Pathogenomics of *Culex quinquefasciatus* and meta-analysis of infection responses to diverse pathogens. *Science* 330: 88-90. doi:10.1126/science.1193162. PubMed: 20929811.
20. Dinglasan RR, Jacobs-Lorena M (2008) Flipping the paradigm on malaria transmission-blocking vaccines. *Trends Parasitol* 24: 364-370. doi:10.1016/j.pt.2008.05.002. PubMed: 18599352.
21. Patramool S, Choumet V, Surasombatpattana P, Sabatier L, Thomas F et al. (2012) Update on the proteomics of major arthropod vectors of human and animal pathogens. *Proteomics* 12: 3510-3523. doi:10.1002/pmic.201200300. PubMed: 23077092.
22. Megy K, Emrich SJ, Lawson D, Campbell D, Dialynas E et al. (2012) VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics. *BMC Genomics* 40: D729-D734. doi:10.1093/nar/gkr1089. PubMed: 22135296.
23. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G et al. (2012) Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* 40: D700-D705. doi:10.1093/nar/gkr1029. PubMed: 22110037.
24. Marygold SJ, Leyland PC, Seal RL, Goodman JL, Thurmond J et al. (2013) FlyBase: improvements to the bibliography. *Nucleic Acids Res* 41: D751-D757. doi: 10.1093/nar/gks1024. PubMed: 23125371.
25. Cooper PS, Lipshultz D, Matten WT, McGinnis SD, Pechous S et al. (2010) Education resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 38: 563-569. doi:10.1093/nar/bbq022. PubMed: 20570844.
26. Finn RD, Mistry J, Schuster-Böckler B, Griffiths-Jones S, Hollich V et al. (2006) Pfam: clans, web tools and services. *Nucleic Acids Res* 34: D247-D251. doi: 10.1093/nar/gkj149. PubMed: 16381856.
27. Bateman A, Birney E, Durbin R, Eddy SR, Finn RD, et al. (1999) Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. *Nucleic Acids Res* 27: 260-262
28. Guindon S, Lethiec F, Duroux P, Gascuel O (2005) PHYML Online - a web server for fast maximum likelihood-based phylogenetic inference. *Bioinformatics* 21: W557-W559. doi:10.1093/nar/gki352. PubMed: 15980534.
29. Bade VN, Nickels J, Keusekotten K, Praefcke GJK (2012) Covalent protein modification with ISG15 via a conserved cysteine in the hinge region. *PLOS ONE* 7: e38294. doi:10.1371/journal.pone.0038294. PubMed: 22693631.
30. Bayer P, Arndt A, Metzger S, Mahajan R, Melchior F et al. (1998) Structure determination of the small ubiquitin-related modifier SUMO-1. *J Mol Biol* 280: 275-286. doi:10.1006/jmbi.1998.1839. PubMed: 9654451.
31. Gill G (2004) SUMO and ubiquitin in the nucleus: different functions, similar mechanisms? *Genes* 18: 2046-2059 PubMed: 15342487
32. Kumar S, Yoshida Y, Noda M (1993) Cloning of a cDNA which encodes a novel ubiquitin-like protein. *Biochem Biophys Res Commun* 195: 393-399. doi:10.1006/bbrc.1993.2056. PubMed: 8395831.
33. Hjerpe R, Thomas Y, Kurz T (2012) NEDD8 overexpression results in neddylation of ubiquitin substrates by the ubiquitin pathway. *J Mol Biol* 421: 27-29. doi:10.1016/j.jmb.2012.05.013. PubMed: 22608973.
34. Furukawa K, Mizushima N, Noda T, Ohsumi Y (2000) A protein conjugation system in yeast with homology to biosynthetic enzyme reaction of prokaryotes. *J Biol Chem* 275: 7462-7465. doi:10.1074/jbc.275.11.7462. PubMed: 10713047.
35. Mishra SK, Ammon T, Popowicz GM, Krajewski M, Nagel RJ et al. (2011) Role of the ubiquitin-like protein Hub1 in splice-site usage and alternative splicing. *Nature* 474: 173-178. doi:10.1038/nature10143. PubMed: 21614000.
36. Kraft C, Kijanska M, Kalie E, Siergiejuk E, Lee SS et al. (2012) Binding of the Atg1/ULK1 kinase to the ubiquitin-like protein Atg8 regulates autophagy. *EMBO J* 31: 3691-3703. doi:10.1038/emboj.2012.225. PubMed: 22885598.
37. Hofmann K, Bucher P, Kajava AV (1998) A model of Cdc25 phosphatase catalytic domain and Cdk-interaction surface based on the presence of a rhodanese homology domain. *J Mol Biol* 282: 195-208. doi:10.1006/jmbi.1998.1998. PubMed: 9733650.
38. Komatsu M, Tanida I, Ueno T, Ohsumi M, Ohsumi Y et al. (2001) The C-terminal region of an Apg7p/Cvt2p is required for homodimerization and is essential for its E1 activity and E1-E2 complex formation. *J Biol Chem* 276: 9846-9854. doi:10.1074/jbc.M007737200. PubMed: 11139573.
39. Pickart CM (2001) Mechanisms underlying ubiquitination. *Annu Rev Biochem* 70: 503-533. doi:10.1146/annurev.biochem.70.1.503. PubMed: 11395416.
40. Inagawa T, Yamada-Inagawa T, Eydmann T, Mian IS, Wang TS et al. (2009) *Schizosaccharomyces pombe* Rtf2 mediates site-specific replication termination by inhibiting replication restart. *Mol Cell* 106: 7927-7932. doi:10.1073/pnas.0812323106. PubMed: 19416828.
41. Jin J, Li J-M (2012) CRL ubiquitin ligases and DNA damage response. *Front Oncol* 2: 1-10. doi:10.3389/fonc.2012.00001. PubMed: 22979629.
42. Marin I (2010) Ancient origin of animal U-box ubiquitin ligases. *BMC Evol Biol* 10: 331. doi:10.1186/1471-2148-10-331. PubMed: 20979629.
43. Chatterjee A, Upadhyay S, Chang X, Nagpal JK, Trink B, et al. (2008) U-box-type ubiquitin E4 ligase, UFD2a attenuates cisplatin mediated degradation of Δ Np63 α . *Cell Cycle* 7: 1231-1237.
44. Correa RL, Bruckner FP, de Souza Cascardo R, Alfenas-Zerbini P (2013) The role of F-Box proteins during viral infection. *Int J Mol Sci* 14: 4030-4049. doi:10.3390/ijms14024030. PubMed: 23429191.
45. Evans PC, Smith TS, Lai M-J, Williams MG, Burke DF et al. (2003) A novel type of deubiquitinating enzyme. *J Biol Chem* 278: 23180-23186. doi:10.1074/jbc.M301863200. PubMed: 12682062.
46. Mauri PL, Riva M, Ambu D, De Palma A, Secundo F et al. (2006) Ataxin-3 is subject to autolytic cleavage. *FEBS J* 273: 4277-4286. doi: 10.1111/j.1742-4658.2006.05419.x. PubMed: 16939621.
47. Patterson-Fortin J, Shao G, Bretscher H, Messing TE, Greenberg RA (2010) Differential regulation of JAMM domain deubiquitinating enzyme activity within the RAP80 complex. *J Biol Chem* 285: 30971-30981. doi: 10.1074/jbc.M110.135319. PubMed: 20656689.
48. Cooper EM, Cutcliffe C, Kristiansen TZ, Pandey A, Pickart CM et al. (2009) K63-specific deubiquitination by two JAMM/MPN+ complexes: BRISC-associated Brcc36 and proteasomal Poh1. *EMBO J* 28: 621-631. doi:10.1038/emboj.2009.27. PubMed: 19214193.
49. Su D, Hochstrasser M (2010) A WLM Protein with SUMO-directed protease activity. *Mol Cell Biol* 30: 3734-3736. doi:10.1128/MCB.00673-10. PubMed: 20566698.
50. Larsen CN, Krantz BA, Wilkinson KD (1998) Substrate specificity of deubiquitinating enzymes: ubiquitin C-terminal hydrolases. *Biochimie* 37: 3358-3368. doi:10.1021/bi972274d.
51. Johnston SC, Larsen CN, Cook WJ, Wilkinson KD, Hill CP (1997) Crystal structure of a deubiquitinating enzyme (human UCH-L3) at 1.8 Å resolution. *EMBO J* 16: 3787-3796. doi:10.1093/emboj/16.13.3787. PubMed: 9233788.
52. Mossessova E, Lima CD (2000) Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol Cell* 5: 865-876. doi: 10.1016/S1097-2765(00)80326-3. PubMed: 10882122.
53. Kang SH, Kim GR, Seong M, Baek SH, Seol JH et al. (2007) Two novel ubiquitin-fold modifier 1 (Ufm1)-specific proteases, UfSP1 and UfSP2. *J Biol Chem* 282: 5256-5262. PubMed: 17182609.
54. Severo MS, Choy A, Stephens KD, Sakhon OS, Chen G et al. (2013) The E3 ubiquitin ligase XIAP restricts *Anaplasma phagocytophilum* colonization of *Ixodes scapularis* Ticks. *J Infect Dis*. doi:10.1093/infdis/jit298. PubMed: 23066989.
55. de la Fuente J, Blouin EF, Manzano-Roman R, Naranjo V, Almazán C et al. (2007) Functional genomic studies of tick cells in response to infection with the cattle pathogen, *Anaplasma marginale*. *Genomics* 90: 712-722. doi:10.1016/j.ygeno.2007.08.009. PubMed: 17964755.
56. Huang B, Ojogun N, Ragland SA, Carlyon JA (2012) Monoubiquitinated proteins decorate the *Anaplasma phagocytophilum*-occupied vacuolar membrane. *FEMS Immunol Med Microbiol* 64: 32-41. doi:10.1111/j.1574-695X.2011.00873.x. PubMed: 22066989.