

## RESEARCH ARTICLE

Prediction pipeline for discovery of regulatory motifs associated with *Brugia malayi* moltingAlexandra Grote<sup>1‡</sup>, Yichao Li<sup>2‡</sup>, Canhui Liu<sup>3</sup>, Denis Voronin<sup>4</sup>, Adam Geber<sup>1</sup>, Sara Lustigman<sup>4</sup>, Thomas R. Unnasch<sup>3</sup>, Lonnie Welch<sup>2\*</sup>, Elodie Ghedin<sup>1,5\*</sup>

**1** Department of Biology, Center for Genomics and Systems Biology, New York University, New York, New York, United States of America, **2** School of Computer Science and Electrical Engineering, Ohio University, Athens, Ohio, United States of America, **3** Center for Global Infectious Disease Research, University of South Florida, Tampa, FL, Florida, United States of America, **4** Laboratory of Molecular Parasitology, Lindsley F. Kimball Research Institute, New York Blood Center, New York, New York, United States of America, **5** Department of Epidemiology, School of Global Public Health, New York University, New York, New York, United States of America

‡ These authors share first authorship on this work

\* [welch@ohio.edu](mailto:welch@ohio.edu) (LW); [elodie.ghedin@nyu.edu](mailto:elodie.ghedin@nyu.edu) (EG)

## OPEN ACCESS

**Citation:** Grote A, Li Y, Liu C, Voronin D, Geber A, Lustigman S, et al. (2020) Prediction pipeline for discovery of regulatory motifs associated with *Brugia malayi* molting. PLoS Negl Trop Dis 14(6): e0008275. <https://doi.org/10.1371/journal.pntd.0008275>

**Editor:** James B. Lok, University of Pennsylvania, UNITED STATES

**Received:** November 14, 2019

**Accepted:** April 7, 2020

**Published:** June 23, 2020

**Copyright:** © 2020 Grote et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All sequencing data have been deposited to NCBI's SRA (NCBI PRJNA557263) and the prediction pipeline is available on GitHub ([https://github.com/YichaoOU/Emotif\\_Alpha](https://github.com/YichaoOU/Emotif_Alpha)).

**Funding:** This work was supported by US National Institute of Allergy and Infectious Diseases (NIAID)/National Institutes of Health (NIH) grant number R21 AI135172-01A1 to TRU; R56 AI101372 to EG, TRU, and SL; and R21 AI126466 to EG. LW received funding from The Ohio

## Abstract

Filarial nematodes can cause debilitating diseases in humans. They have complicated life cycles involving an insect vector and mammalian hosts, and they go through a number of developmental molts. While whole genome sequences of parasitic worms are now available, very little is known about transcription factor (TF) binding sites and their cognate transcription factors that play a role in regulating development. To address this gap, we developed a novel motif prediction pipeline, Emotif Alpha, that integrates ten different motif discovery algorithms, multiple statistical tests, and a comparative analysis of conserved elements between the filarial worms *Brugia malayi* and *Onchocerca volvulus*, and the free-living nematode *Caenorhabditis elegans*. We identified stage-specific TF binding motifs in *B. malayi*, with a particular focus on those potentially involved in the L3-L4 molt, a stage important for the establishment of infection in the mammalian host. Using an *in vitro* molting system, we tested and validated three of these motifs demonstrating the accuracy of the motif prediction pipeline.

## Author summary

Diseases caused by parasitic worms such as the filariae are among the leading causes of morbidity in the developing world. Very little is known about how development is regulated in these vector-transmitted parasites. We have developed a computational method to identify motifs that correspond to transcription factor binding sites in the genome of the parasitic worm, *Brugia malayi*, one of the causative agents of lymphatic filariasis. Using this approach, we were able to predict stage-specific transcription factor binding sites involved in a stage of the molting process important for the establishment of the infection. We validated the role of these motifs using an *in vitro* molting system.

University GERB Program. AG received funding from the T32 Ruth L. Kirschstein Institutional National Research Service Award (T32AI007180) and the F31 Ruth L. Kirschstein Pre-doctoral Individual NRSA (F31AI131527). 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.

## Introduction

*Brugia malayi* is a mosquito-borne filarial nematode and one of the causative agents of lymphatic filariasis, commonly known as elephantiasis. Currently, 856 million people in 52 countries require preventative chemotherapy to stop the spread of the disease [1]. Transmission occurs when the mosquito vector introduces infective third-stage larvae (L3) during their blood meal. The larvae then migrate to the lymphatic vessels where they molt twice and develop into adults. Over their lifespan adult females produce millions of microfilariae (immature larvae) that circulate in the blood, allowing for continued transmission. Chronic lymphatic filariasis can cause permanent and disfiguring damage, characterized by lymphoedema (tissue swelling) and elephantiasis (tissue thickening) of the lower limbs, and hydrocele (scrotal swelling).

Over the past decade, a few parasitic nematode genomes have been sequenced, including *B. malayi* [2], *Loa loa* [3] and *Onchocerca volvulus* [4]. Transcriptomic experiments have helped quantify differentially expressed genes and their biological implications [5,6,7,8,9]. However, little is known about how these genes are regulated through cis-regulatory motifs. Motifs that have been characterized in *B. malayi* and that are available in the CIS-BP database [10] are purely bioinformatic predictions based on transcription factor binding site (TFBS) homology. *De novo* DNA motif discovery is an effective bioinformatic method for studying transcriptional gene regulation [11] and a number of motif discovery methods and tools currently exist. These include expectation-maximization methods, such as MEME [12] and Improbizer [13]; Gibbs sampling methods, such as BioProspector [14] and MotifSampler [15]; k-mer enumeration methods such as Weeder [16], DME [17], and DECOD [18]; ensemble methods such as W-ChIPMotifs [19], and GimmeMotifs [20]; and deep learning methods such as DanQ [21] and DeepFinder [22]. Based on the input types, motif discovery approaches can also be classified as either generative or discriminative. Generative motif discovery models use pre-defined background models (e.g., the Hidden Markov Model), while discriminative motif discovery models need to explicitly specify a set of background sequences. In this study, we developed Emotif Alpha that integrates a number of the current methods based on the aforementioned models, and filters the motifs using a Z-test, random forest feature importance, and sequence homology.

Gene promoter regions play a crucial role in gene regulation yet remain largely uncharacterized in *B. malayi*. Among the very few promoters that have been previously described and validated in *B. malayi* is that of Heat Shock Protein 70 (HSP70) [23]. A previous study showed that while the regulatory domains of the HSP70 promoter were similar to other eukaryotes, the core promoter domains appeared to be distinct [24]. Furthermore, nothing is known about motifs regulating developmentally expressed genes in *B. malayi*. There is thus a need for systematic identification, annotation, and experimental validation of *B. malayi* promoter motifs associated with gene regulation to better characterize filaria gene expression patterns. To better understand how promoter elements regulate stage-specific gene expression, we did differential gene expression analysis of the L3 to L4 molt, the first developmental step important for the establishment of infection in the mammalian host, and motif discovery using the Emotif Alpha pipeline. Several promoter motifs appeared to be associated with the regulation of the L3 to L4 molt. Our results provide an initial overview of the putative regulatory mechanisms in the filariae that could be targeted using novel intervention strategies for control.

## Results

### Stage-specific expression of serpins, peptidases, cysteine protease inhibitors, and structural constituents of the cuticle during the L3 to L4 molt

Since the L3 to L4 molt is of particular interest because it corresponds to the life cycle stage when infective larvae establish infection, we focused in this study on identifying genes that are

Table 1. RNA-Seq summary.

Library <sup>a</sup>	Total reads (millions) <sup>b</sup>	Stage Total (millions)	Total Mapped Reads (millions)	Stage Total Mapped Reads (millions)	% <i>B. malayi</i> Genes Expressed
L3a	294	763	266	639	84
L3b	223		204		
L3c	246		169		
L3D6a	508	971	131	477	82.3
L3D6b	243		152		
L3D6c	220		194		
L3D9a	300	440	170	266	81.2
L3D9b	140		96		
L4a	104	195	47	96	72.4
L4b	90		49		

<sup>a</sup>Lower case a, b, and c refer to separate biological replicates.

<sup>b</sup>Total reads sequenced and mapped in each biological replicate at each developmental stage, L3 to L4.

<https://doi.org/10.1371/journal.pntd.0008275.t001>

differentially expressed during this unique process. We used RNA-seq to profile transcription at different time points during the molt, collecting samples from the infective L3 (from mosquitoes), L3 Day 6, and L3 Day 9 worms recovered from infected gerbils (NCBI PRJNA557263). We combined this transcriptome data with previously published L4 data [7] that corresponds to Day 14 post infection of gerbils (Table 1). In total, 2.36 billion reads were generated, with 1.38 billion reads mapping to the *B. malayi* genome. Each biological replicate received an average of 272 million reads, with an average of 173 million reads that were successfully mapped (Table 1). Of the 11,841 *B. malayi* gene models, 87.6% were expressed in at least one stage of the L3 to L4 molt (Fig 1). The molting expression data shows unique stage-specific profiles for each stage of the molt with significant differential expression between days. Principle component analysis based on gene expression shows close clustering of biological replicates (S1 Fig).

We determined differentially expressed genes during the iL3 to L4 molt using both DESeq [25] and EdgeR [26] to perform pairwise comparisons between the four samples. To get a high confidence list of differentially expressed genes, we used the consensus of the two algorithms (FDR  $\leq 0.01$ , log fold change  $\geq 2.5$ ) (S1 Table). All differentially expressed genes had a minimum coverage of 7 read counts in at least one stage. For the purposes of this study, we focused on the genes that were up-regulated at each stage of molting, as compared to the other stages, and did a gene annotation enrichment analysis for each stage comparison (S2 Table). We found that up-regulated genes in iL3 larvae—as compared to L3D6, L3D9, and L4—were enriched for annotations involving cysteine-type peptidase activity as well as serpin domains and serpin family proteins. Cysteine-type peptidases are essential for molting in *B. malayi* [27,28], and serpins are serine protease inhibitors that have previously been shown to be involved in immunomodulation and host immune evasion during infection [29]. We identified five different cysteine-type peptidases and two cysteine-type endopeptidase inhibitors that were upregulated in iL3 larvae. By day 6 of molting, structural constituents of the cuticle, including collagen (the main component of the cuticle) were enriched in the up-regulated gene sets. We also see the up-regulation of several metalloproteases. At day 9, genes involved in signaling were enriched among the up-regulated genes, as were several different metalloproteases. At day 14 (L4 larvae), we again see an enrichment of structural constituents of the cuticle. As for those enriched in L3 day 6 larvae, they are all mostly orthologs of *C. elegans* col (COLLAGEN) genes, which are themselves orthologs of human MARCO genes (macrophage

receptor with collagenous structure). The structural constituents enriched at day 14 are, however, a completely unique set of collagen genes as compared to the genes observed at day 6. These stage-specific enrichments reflect the order of peptidases and structural constituents necessary for the building of a new L4 cuticle, the separation of the old L3 cuticle from the developing L4 cuticle, and the shedding of the old L3 cuticle.

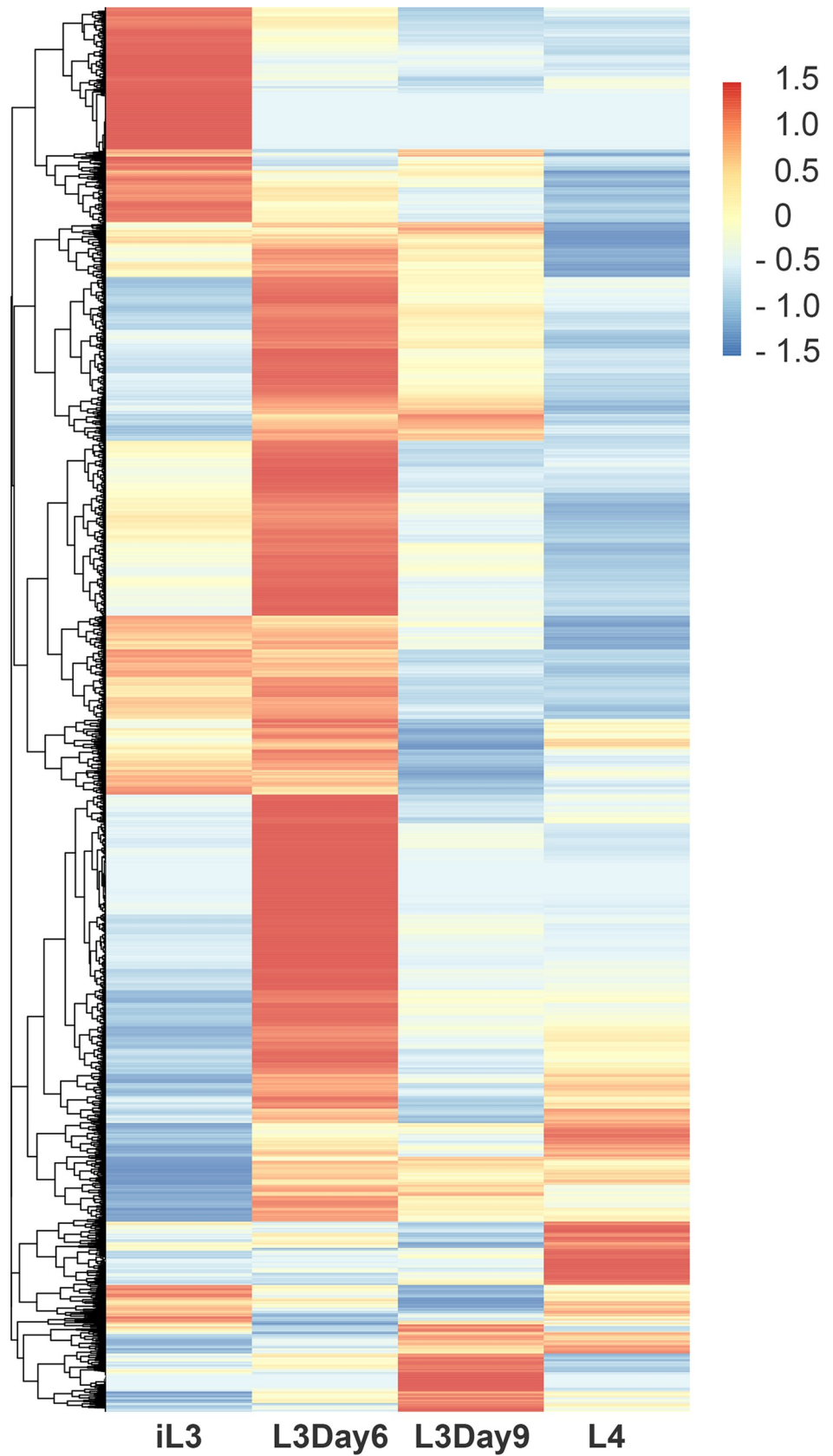
### Identification of 12 motifs associated with transcription factor binding that are enriched in the L3 to L4 molt

To better understand the regulatory program of *B. malayi* during the L3 to L4 molt, we analyzed statistically over-represented DNA motifs in regions upstream of genes that were up-regulated during molting. To do so, we developed a motif identification pipeline called Emotif Alpha (Fig 2). First, we used the transcriptome data from the different stages of the L3 to L4 molt to generate using pair-wise comparisons lists of genes up-regulated at each stage of the molt. We then did a motif discovery analysis on each gene set using a combination of three motif discovery tools: GimmeMotifs, DME, and DECOD. GimmeMotifs is an ensemble of generative motif discovery tools—including Homer [30], AMD [31], BioProspector, MDmodule [32], MEME, Weeder, GADEM [33], and Improbizer—while DME and DECOD are discriminative motif discovery tools. We did a discriminative motif discovery analysis by randomly selecting background promoter region sets from all *B. malayi* genes, excluding the differentially expressed genes. These background sets are three times larger than the foreground sets. We selected motif lengths between 6- and 15-mer. In total, we identified 20,025 motifs.

To select statistically significant motifs, we first assessed the motifs by a random forest classifier using scikit-learn [34]. The random forest algorithm uses bootstrap sampling and constructs a decision tree for each sub-sample. To evaluate the motifs, we used both Gini impurity [35] and information gain [36] criteria, and retained the union of the resulting top 40 motifs. We further filtered the motifs by foreground coverage (i.e. UG%), removing motifs occurring in less than 30% of the genes up-regulated at that stage of molting. We then used a Z-test to compare the frequency of a motif in the up-regulated genes with the expected frequency in the background promoters. Using a significance level (p-value) cutoff of  $10^{-3}$ , we selected 395 motifs (S3 Table).

We retrieved a collection of 163 known nematode transcription factor binding sites (TFBSs) from the MEME suite (<http://meme-suite.org/>), searching the motif databases JASPAR CORE 2016 nematodes [37], CIS-BP *Brugia malayi* [14], and Uniprobe worm [38]. We matched the remaining motifs to known TFBSs with TOMTOM [39]. If two motifs were matched to the same binding site and they were discovered from the same gene list, we considered them to be redundant and kept the one with the lowest over-representation p-value. This step narrowed our list down to 27 motifs that had matches to known binding sites.

We next performed a conservation analysis amongst nematodes using an adaptation of a published method [40]. We retrieved orthologous gene information among *B. malayi*, *C. elegans*, and *O. volvulus* from Wormbase ParaSite Biomart [41]. We extracted up to 1kb upstream from the translation start sites for *B. malayi* genes, assuming these regions would contain the promoter. We performed multiple sequence alignments using CLUSTALW2 [42] and defined a motif as conserved if it occurred at the same position in the orthologous promoter region alignment of either *C. elegans* or *O. volvulus*. This step resulted in 12 remaining motifs (Fig 3, Table 2) that were (1) enriched (p-value  $< 10^{-3}$ ) and (2) conserved in either *C. elegans* or *O. volvulus*. The frequency of motif occurrence in the putative promoter regions of up-regulated genes ranges from 33% to 94%. The fold enrichment, representing the ratio between motif





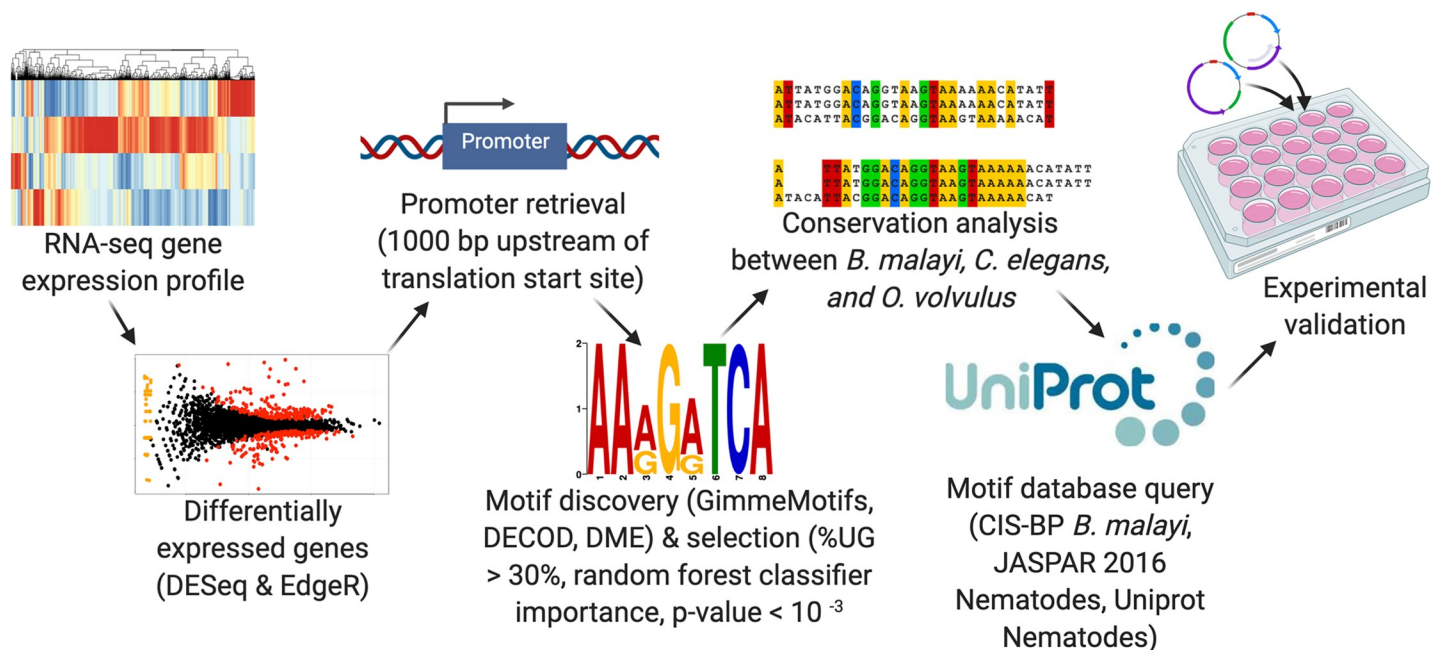
**Fig 1. Expression of *Brugia malayi* genes during the L3 to L4 molt.** Expression is in FPKMs and is Z-scale normalized by row prior to clustering. High expression is indicated by red and low expression by blue. Time-points included infective L3 larvae (iL3), L3 larvae at Day 6 of molting (L3D6), L3 larvae at Day 9 of molting (L3D9), and L4 larvae. Biological replicates have been combined.

<https://doi.org/10.1371/journal.pntd.0008275.g001>

frequencies in the up-regulated gene promoters and background promoters, ranges from 1.28 to 2.19.

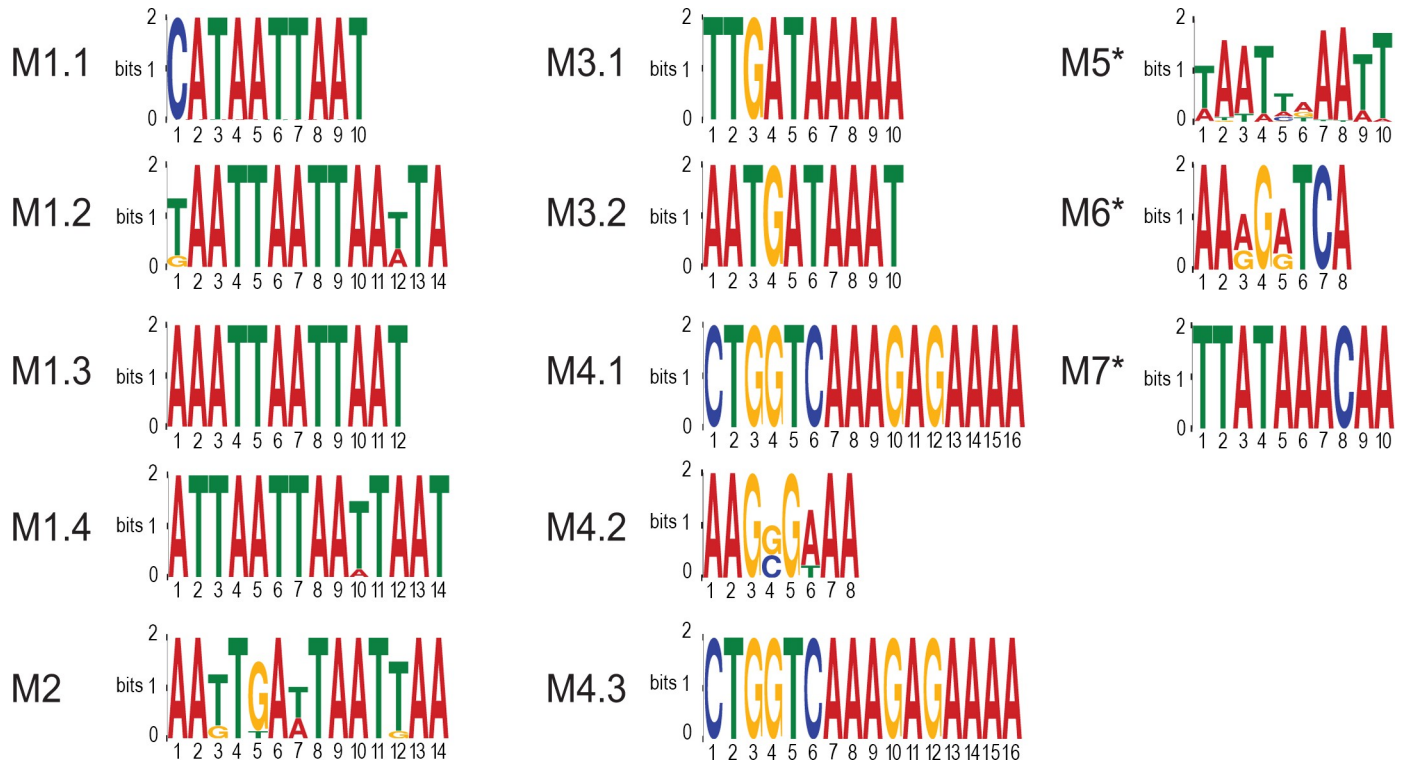
The 12 selected motifs matched known binding sites for 6 transcription factors in *C. elegans* (Table 2), all of which are involved in development, aging, and/ or movement. Motifs M1.1, M1.2, M1.3 and M1.4 matched a zinc-finger protein, *zfh-2*, which is involved in hermaphrodite genitalia development, locomotion, nematode larval development and receptor-mediated endocytosis. Motif M2 matched *vab-7*, which is associated with DB motor neuron identity and posterior DB axonal polarity. Motifs M3.1 and M3.2 matched *elt-3*, which is related to aging [43]. Motifs M4.1, M4.2, and M4.3 matched *blmp-1*. M5 matched a homeobox protein, Bm8528, and M6 matched a nuclear receptor, Bm4429. The 12 motifs are conserved in either *C. elegans* or *O. volvulus* (Table 2, last column; S4 Table). Moreover, the occurrence of M3.2 in the Bm10655 promoter region is conserved in orthologs in both *C. elegans* (promoter of C28A5.3) and *O. volvulus* (promoter of OVOC9600). We find that a number of the *B. malayi* orthologs to the *C. elegans* transcription factors that we predict could bind the predicted TFBS are differentially expressed as well (S5 Table).

The motif analysis reveals how some of the differential expression of different proteases may be orchestrated during the L3 to L4 molt. Motif M1.4 is found in the promoter region of Bm2270, a predicted metalloprotease significantly up-regulated in L3D9 worms (S4 Table).



**Fig 2. Workflow for promoter motif identification.** The six main steps for motif discovery were: (1) generation of an RNA-seq profile; (2) determination of up-regulated genes for every pairwise comparison using DESeq2 (FDR<0.01) and EdgeR (adjusted P-value<0.01); (3) promoter retrieval: 1000 bp upstream of the translation start site; (4) ensemble motif discovery using GimmeMotifs (Homer, AMD, BioProspector, MDmodule, MEME, Weeder, GADEM, and Improbizer), DECOD, DME, and selection of enriched motifs: %UG > 30% and random forest classifier feature importance and over-representation p-value < 10<sup>-3</sup>; (5) TFBS conservation analysis between *B. malayi*, *C. elegans*, and *O. volvulus*; and (6) motif database query: CIS-BP *B. malayi*, JASPAR 2016 Nematodes, Uniprot Nematodes. Finally, a subset of those identified motifs were experimentally validated.

<https://doi.org/10.1371/journal.pntd.0008275.g002>



**Fig 3. Logos of enriched promoter motifs over the L3 to L4 molt.** Motifs found to be enriched ( $p$ -value  $< 10^{-3}$ ) in the upstream elements of up-regulated genes between different stages of molting and to be conserved in either *C. elegans* or *O. volvulus*. The \* denotes the three motifs that have been validated experimentally. Note that M7 was included in the experimental validation because it passed 4 out of 5 filters, including the foreground coverage filter, the random forest filter, the known motif filter, and the conservation filter. However, it was not included in the 12 reported motifs due to its non-significant  $p$ -value.

<https://doi.org/10.1371/journal.pntd.0008275.g003>

Bm2270 is an ortholog of *nas-37* in *C. elegans* and has been shown to be involved in collagen and cuticulin-based cuticle development and ecdysis. Motif M5 is found in the promoter region of Bm1938 and is predicted to encode a serpin (S4 Table). Bm1938 is one of the serpins that was found to be significantly up-regulated in the iL3 larvae.

### L3 stage-specific transcription factor binding motifs can be validated *in vitro*

Three of the motifs (M5, M6, and M7) were chosen for validation based on their enrichment in the promoters of genes up-regulated in the mid to late stages of the L3 to L4 molt. Three separate genes, each containing one chosen motif, were tested. The 1 kb upstream region of each gene was amplified from *B. malayi* genomic DNA and cloned upstream of the firefly luciferase reporter gene in the expression vector pGL3 Basic [23]. *B. malayi* L3 were then transfected with the constructs in a co-culture system, as previously described [44]. The parasites were induced to molt *in vitro* and then assayed for luciferase activity. The number of relative light units (RLUs) observed were normalized to those obtained from parasites transfected in parallel with a construct consisting of the *B. malayi* HSP70 promoter driving the expression of the firefly luciferase reporter [44]. The experiment was performed with both the native promoter and a mutant promoter where the nucleotides of the motif had been randomly shuffled. All of the native promoters produced significant amounts of reporter luciferase activity in the molting parasites (ranging from 40–70% of the activity produced by the HSP70 construct transfected

**Table 2. List of enriched promoter motifs over the L3 to L4 molt.**

Name <sup>a</sup>	% UG <sup>b</sup>	Ratio <sup>c</sup>	P-value	Matched known motif	transcription factor (TF)	Function	Gene pair of conserved sites
M1.1 <L4,L3>	64	1.63	1.4*10 <sup>-4</sup>	MA0928.1 ( <i>C. elegans</i> )	zfh-2	Involved in hermaphrodite genitalia development, locomotion, nematode larval development and receptor-mediated endocytosis	[Bm4560, OVOC6906], [Bm856, OVOC3639]
M1.2 <L3D9, L3>	84.1	1.31	7.1*10 <sup>-5</sup>				[Bm856, OVOC3639], [Bm17348, OVOC8896], [Bm17988, OVOC3292]
M1.3 <L3D6, L3>	94.4	1.28	7.6*10 <sup>-6</sup>				[Bm2559, C34C6.3], [Bm2821, OVOC10446], [Bm17348, OVOC8896], [Bm3341, OVOC10396], [Bm4257, OVOC2553]
M1.4 <L3D9, L3D6>	84.7	1.55	9.6*10 <sup>-8</sup>				[Bm7179, OVOC394], [Bm2270, OVOC2391]
M2 <L3D6, L4>	93.3	1.31	2.9*10 <sup>-5</sup>	MA0927.1 ( <i>C. elegans</i> )	vab-7	Required for DB motorneuron identity and posterior DB axonal polarity	[Bm4904, OVOC2123]
M3.1 <L3D9, L3D6>	80.6	1.33	3.0*10 <sup>-4</sup>	MA0542.1 ( <i>C. elegans</i> )	elt-3	Controlling hypodermal cell differentiation	[Bm2802, OVOC9504]
M3.2 <L3D9, L3>	81.8	1.32	8.3*10 <sup>-5</sup>				[Bm10655, C28A5.3, OVOC9600]
M4.1 <L4, L3D6>	34.8	2.19	4.9*10 <sup>-4</sup>	MA0537.1 ( <i>C. elegans</i> )	blimp-1	Loss of blmp-1 activity via deletion mutation has been reported to result in small, dumpy animals with abnormal fat content	[Bm2802, OVOC9504], [Bm6190, OVOC827]
M4.2 <L4,L3>	56	1.71	1.6*10 <sup>-4</sup>				[Bm7019, OVOC7405], [Bm6190, OVOC827]
M4.3 <L3D9, L3D6>	38.9	2.01	1.0*10 <sup>-5</sup>				[Bm2802, OVOC9504], [Bm7179, OVOC394]
M5* <L3, L3D6>	93.3	1.23	2.9*10 <sup>-4</sup>	M5348_1.02 (CIS-BP <i>Brugia</i> inferred)	Bm8528	Retinal homeobox protein Rx3	[Bm1938, OVOC2080], [Bm8228, C27D6.4], [Bm1559, OVOC7210], [Bm4184, OVOC3386]
M6* <L3, L3D6>	33.3	2.03	3.0*10 <sup>-5</sup>	M5221_1.02 (CIS-BP <i>Brugia</i> inferred)	Bm4429	Involved in regulation of transcription, DNA-templated and steroid hormone mediated signaling pathway	[Bm1938, OVOC2080], [Bm6642, Y48B6A.12]
M7* <L3, L3D9>	54.5	1.19	6.1*10 <sup>-2</sup>	M0739_1.02 (CIS-BP inferred)	Bm3608	Involved in cell growth, proliferation, differentiation, and longevity.	[Bm4184, OVOC3386]

<sup>a</sup>The text listed in <> brackets correspond to comparisons between stages of molting where upstream elements of upregulated genes were found to be enriched with specific motifs.

<sup>b</sup>Frequency of a motif in up-regulated gene promoters.

<sup>c</sup>Relative frequency of a motif in up-regulated gene promoters vs. background promoters. \*Denotes the three motifs validated experimentally.

<https://doi.org/10.1371/journal.pntd.0008275.t002>

positive controls; Fig 4). However, when the putative motifs were mutated, the activity in all the promoters tested decreased by 80–90% (Fig 4).

## Discussion

Parasitic nematodes such as *B. malayi* maintain a complicated lifecycle involving both an insect and a mammalian host, and undergo a number of molts. The L3 to L4 molt that occurs immediately upon infection of the mammalian host is of particular interest as it marks the establishment of infection and thus represents an attractive point for drug intervention. Little



is known, however, about how *B. malayi* regulates the transitions between these stages. Prior to our study, nothing was known about promoter motifs that regulate developmentally expressed genes in *B. malayi*. Because stage transitions rely on precise transcriptional control through the interaction of transcription factors and their binding sites, we set out to characterize the potential transcription factor binding motifs of these parasites and to identify motifs that contribute to stage-specific expression of genes involved in early worm development in the mammalian host.

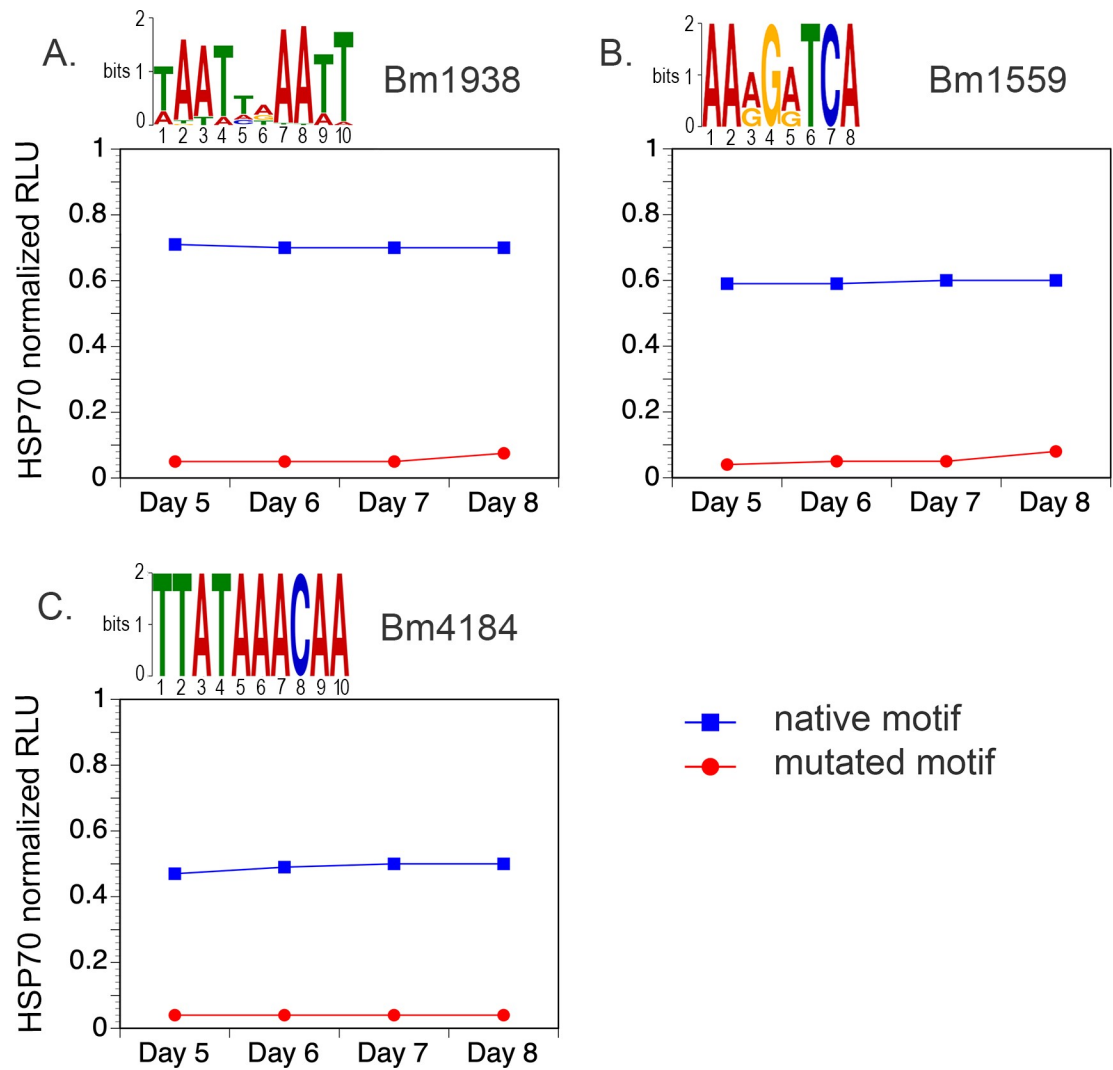
We found a number of enriched motifs and were able to define both conserved motifs across molting as well as stage-specific motifs. While some of the motifs we identified are conserved in other nematodes, such as *C. elegans*, a number of motifs represent novel binding sites potentially reflecting the differences in development and the parasitic lifestyle. It is known that molting is regulated by an ecdysone-like response system [45,46,47,48]. Two of the identified motifs and their cognate TFs appear to be related to the ecdysone response. For example, *zfh-2*, the transcription factor predicted to bind four of our identified motifs, is a common cofactor implicated in ecdysone signaling in *D. melanogaster* [49]. *Blimp-1*, the transcription factor predicted to bind three of our identified motifs, is an ecdysone-inducible repressor that is essential for the prepupal development in *Drosophila* [50]. Validation results suggest that our pipeline is able to identify biologically-relevant motifs involved in molting. This analysis provided biological insight into the development of the parasite as well as the identification of novel drug targets.

Future studies should expand this analysis across the lifecycle of the nematode at different stages of development, in its different hosts (i.e. human vs. mosquito). While the validation results from our *in vitro* system are promising, we must assume there are biological differences to molting *in vivo*. Transcriptomic data from other life stages already exist and can be used to predict motifs; however, validation at other stages *in vivo* will prove more difficult. This is due, in large part, to the fact that the human filariae are obligate parasites without a free-living stage. This has limited the ability to genetically manipulate these stages in ways necessary to validate predictions made by transcriptomic analyses. Until recently, studies of promoter structure and function were limited by the fact that the only way to transfect *B. malayi* involved biolistic transfection of isolated embryos [23] which were developmentally incompetent. While this system has been used to identify the conserved motifs necessary for promoter function in constitutive promoters [24,51,52,53] and cis acting regulatory regions in some regulated promoters [54,55], these studies were limited to genes expressed in embryos. However, with recent innovations in filarial transgenics, it is now possible to efficiently produce stably transfected developmentally competent infective larvae, which can in turn be used to produce transgenic parasites in which transgenes are stably integrated into the parasite genome [44]. This study represents the first application of this new technology to the study of promoter function in a lifecycle stage other than isolated developmentally incompetent embryos. The ability to produce stable transgenic parasite lines will permit *in vivo* functional testing of promoter motifs for stage-specific expressed genes in all lifecycle stages of the parasite.

## Materials and Methods

### Transcriptomic study design

All parasites were obtained from FR3 (Filariasis Research Reagent Resource Center; BEI Resources, Manassas, VA, USA) where they were isolated from infected gerbils (*Meriones unguiculatus*) or mosquitoes (*Aedes aegypti*). Worms were flash-frozen and shipped to the New York Blood Center for processing. For transcriptomic sequencing, infective third-stage larvae (iL3) were recovered from mosquitoes and mammalian stage larvae were recovered



**Fig 4. Promoter motif validation.** A) Promoter motif validation in L3 worms that were molting *in vitro* using the native promoter of Bm1559 and a mutated motif M5 version of the same promoter. B) Promoter motif validation using the native promoter of Bm1938 and a mutated motif M6 version of the promoter. C) Promoter motif validation using the native promoter of Bm4184 with a mutated motif M7 version of the promoter. A total of four wells, each containing 100 L3, were transfected with each construct, as previously described [44]. Parasites in each well were harvested on days 5–8 and firefly luciferase activity determined as previously described [23]. In each panel luciferase activity obtained from the constructs is normalized against parasites transfected with a construct containing the Bm HSP70 promoter driving the expression of the firefly luciferase reporter.

<https://doi.org/10.1371/journal.pntd.0008275.g004>

from gerbils at 6 and 9 days post infection (dpi). At 6 dpi, larvae are typically undergoing the molt from L3 to L4, while by 9 dpi some worms may be finished molting [56] while others may not finish until 10 dpi [57]. Data was combined with previously published stages 14 dpi (L4) [7].

### RNA isolation, library preparation and sequencing

Total RNA was prepared from *B. malayi* worms as previously described [7]. RNA was prepared from 3 biological replicates of infective L3 (iL3; 2000 larvae each), 3 replicates of 6 dpi larvae (1500 each) and 2 replicates of 9 dpi larvae (1300 each). *B. malayi* worms were homogenized in Trizol (ThermoFisher) using a hand-held pestle in 1.5mL tubes containing the

worms. Total RNA was extracted by organic extraction using Trizol and the PureLink RNA mini kit (ThermoFisher) and after being treated with DNaseI (New England Biolabs). Ribosomal RNA (rRNA) depletion was performed using Terminator (Epicentre), a 5'-phosphate-dependent exonuclease that degrades transcripts with a 5' monophosphate. Libraries were prepared using the NEBNext Ultra II RNA Library Prep Kit for Illumina (New England Biolabs) according to manufacturer instructions. Library quality was assessed using a D1000 Screen-Tape Assay (Agilent) prior to sequencing. Library concentrations were assessed using the qPCR library quantification protocol (KAPA biosystems). Libraries were sequenced on the Illumina NextSeq500 platform with 150bp paired-end reads. To minimize the confounding effects of lane-to-lane variation, libraries were multiplexed and sequenced with technical replicates on multiple lanes. Each biological replicate received an average of 135 million mapped reads (PRJNA557263).

### Sequencing alignment and expression analysis

Read quality was assessed using FastQC (Babraham Bioinformatics). Sequence reads from each sample were analyzed with the Tuxedo suite of tools [58,59,60]. Reads were mapped with Tophat2's Bowtie2-very-sensitive algorithm to the annotated *B. malayi* genome assembly (WormBase.org). The resulting BAM files were then used with HtSeq to obtain raw read counts. Differential gene expression analysis was performed using both DESeq and EdgeR, and the overlapping genes with FDR < 0.01 were retained. Up-regulated genes were characterized for each pairwise comparison between L3, L3 Day 6 (L3D6), L3 day 9 (L3D9), and L4 worms. For example, <L3,L4> refers to the up-regulated genes in L3 compared to L4. Two pairs of comparisons, <L3D9,L4> and <L4,L3D9>, were dropped due to the limited number of up-regulated genes (<= 3) passing our stringent filters, possibly due to the L3D9 being actually younger L4. The up-regulated gene lists were filtered using log<sub>2</sub> fold change (logFC) with the following thresholds: |logFC| = 7 for <L3D6,L3>, <L3D6,L3D9>, <L3D6,L4>; |logFC| = 4 for <L3D9,L3>, <L3,L3D6>, <L3,L3D9>, <L3,L4>, <L4,L3>; |logFC| = 2.5 for <L3D9,L3D6>, <L4,L3D6>. The reason for varying the threshold was that the number of up-regulated genes in each list varied significantly. For motif discovery tools to search efficiently, the number of sequences were limited to less than one hundred. In total, 10 up-regulated gene lists were used for motif discovery (S1 Table). Functional annotation enrichment analysis of the upregulated gene lists was done using DAVID 6.8 [61].

Potential promoter sequences were retrieved from WormBase ParaSite Biomart [41] web interface, capturing the 1000bp upstream of the translation start site for each gene.

### The Emotif Alpha pipeline for regulatory motif identification

The Emotif Alpha pipeline (freely available at: [https://github.com/YichaoOU/Emotif\\_Alpha](https://github.com/YichaoOU/Emotif_Alpha)) was developed to automate motif discovery analysis for the 10 up-regulated gene lists. This pipeline was written in python and was applied to perform all aforementioned motif analyses. The motif discovery step used multiple tools and was run in parallel at the Ohio SuperComputer Center. Motif length search was from 6 to 14. Motif scanning was done using FIMO [62] with a default p-value threshold of  $10^{-4}$ . We implemented 5 different motif filters: (1) Foreground coverage (i.e., UG%) was defined as the proportion of up-regulated gene promoters containing the given motif. We set a minimal foreground coverage at 30%; (2) Motifs were then filtered by a random forest classifier. The union of the top 40 motifs that resulted from either Gini impurity or information gain criterion was retained; (3) Motif enrichment p-value was calculated using Z-test and the cutoff was  $10^{-3}$ ; (4) Known motif filter was performed using TOMTOM and a collection of 163 known nematode TFBSs. The motif similarity p-

value threshold was  $10^{-4}$ ; (5) Conservation analysis was performed using a method described in [40]. Only conserved motifs were kept.

### ***In vitro* validation of promoter transcription motifs**

The putative TF motifs M5, M6, and M7 were chosen for validation based on their enrichment in the promoters of genes upregulated in the mid to late stages of the L3 to L4 molt. Three different genes, each containing one of the chosen motifs, were used for the validation assay. As previously described [23], we amplified the 1 kbp region upstream of each gene from *B. malayi* genomic DNA and cloned it upstream of the firefly luciferase reporter gene in the expression vector pGL3 Basic. We then transfected *B. malayi* L3 larvae with the constructs in a co-culture system, as previously described [44]. The parasites were induced to molt by the addition of ascorbic acid on day 5, and parasites were assayed for luciferase activity on days 5–8, as by day 10 the molting was complete. We normalized the number of RLU observed to those obtained from parasites transfected in parallel with a construct consisting of the *B. malayi* HSP70 promoter driving the expression of the firefly luciferase reporter [23] to control for accumulation of the firefly luciferase over time during the duration of the experiment. We did the experiment with both the native promoter and a mutant promoter where the nucleotides of the motif had been randomly shuffled (S6 Table). Only one replicate was carried out for each motif validation.

### **Supporting information**

**S1 Fig. Principle component analysis of biological replicates based on stage-specific expression.**

(PDF)

**S1 Table. Stage comparisons of upregulated genes.**

(XLSX)

**S2 Table. Functional annotation enrichment of gene lists upregulated during molting.**

(XLSX)

**S3 Table. List of statistically significant *de novo* motifs identified.**

(PDF)

**S4 Table. Wormbase annotations for the gene pairs of conserved sites from Fig 3.**

(XLSX)

**S5 Table. Expression of *B. malayi* transcription factors predicted to interact with the identified motifs in Fig 3.**

(XLSX)

**S6 Table. Primers used for validation mutagenesis.**

(XLSX)

### **Author Contributions**

**Conceptualization:** Alexandra Grote, Yichao Li, Sara Lustigman, Thomas R. Unnasch, Lonnie Welch, Elodie Ghedin.

**Data curation:** Alexandra Grote, Yichao Li, Lonnie Welch, Elodie Ghedin.

**Formal analysis:** Alexandra Grote, Yichao Li, Canhui Liu, Denis Voronin, Adam Geber, Thomas R. Unnasch, Lonnie Welch, Elodie Ghedin.

**Funding acquisition:** Sara Lustigman, Thomas R. Unnasch, Lonnie Welch, Elodie Ghedin.

**Investigation:** Alexandra Grote, Yichao Li, Canhui Liu, Denis Voronin, Sara Lustigman, Thomas R. Unnasch, Lonnie Welch, Elodie Ghedin.

**Methodology:** Alexandra Grote, Yichao Li, Canhui Liu, Denis Voronin, Adam Geber, Sara Lustigman, Thomas R. Unnasch, Lonnie Welch, Elodie Ghedin.

**Project administration:** Sara Lustigman, Lonnie Welch, Elodie Ghedin.

**Resources:** Sara Lustigman, Lonnie Welch.

**Software:** Yichao Li, Lonnie Welch.

**Supervision:** Sara Lustigman, Thomas R. Unnasch, Lonnie Welch, Elodie Ghedin.

**Validation:** Alexandra Grote, Yichao Li, Canhui Liu, Denis Voronin, Thomas R. Unnasch, Lonnie Welch.

**Visualization:** Canhui Liu, Lonnie Welch.

**Writing – original draft:** Alexandra Grote, Yichao Li, Denis Voronin, Thomas R. Unnasch, Lonnie Welch, Elodie Ghedin.

**Writing – review & editing:** Alexandra Grote, Yichao Li, Denis Voronin, Sara Lustigman, Thomas R. Unnasch, Lonnie Welch, Elodie Ghedin.

## References

1. Gordon CA, Jones MK, McManus DP. The history of Bancroftian Lymphatic Filariasis in Australasia and Oceania: Is there a Threat of Re-occurrence in Mainland Australia? *Tropical Medicine and Infectious Disease*. 2018; 3(2), 58.
2. Ghedin E, Wang S, Spiro D, Caler E, Zhao Q, Crabtree J, et al. Draft Genome of the Filarial Nematode Parasite *Brugia malayi*. *Science*. 2007; 317(5845), 1756–1760. <https://doi.org/10.1126/science.1145406> PMID: 17885136
3. Desjardins CA, et al. Genomics of *Loa loa*, a *Wolbachia*-free Filarial Parasite of Humans. *Nature Genetics*. 2013; 45(5): 495–500. <https://doi.org/10.1038/ng.2585> PMID: 23525074
4. Cotton JA, Bennuru S, Grote A, Harsha B, Tracey A, Beech R, Doyle S, et al. The Genome of *Onchocerca volvulus*, Agent of River Blindness. *Nature Microbiology*. 2016; 2 (2).
5. Bennuru S, Cotton JA, Ribeiro JM, Grote A, Harsha B, Holroyd N, Mhashilkar A, Molina DM, Randall AZ, Shandling AD, Unnasch TR, Ghedin E, Berriman M, Lustigman S, Nutman TB. Stage-Specific Transcriptome and Proteome Analyses of the Filarial Parasite *Onchocerca volvulus* and Its *Wolbachia* Endosymbiont. *mBio*; 2016 7(6).
6. Choi YJ, Ghedin E, Berriman M, McQuillan J, Holroyd N, Mayhew GF, et al. A Deep Sequencing Approach to Comparatively Analyze the Transcriptome of Lifecycle Stages of the Filarial Worm, *Brugia malayi*. *PLOS Neglected Tropical Diseases*. 2011; 5(12), 1–10.
7. Grote A, Voronin D, Ding T, Twaddle A, Unnasch T, Lustigman S, Ghedin E. Defining *Brugia Malayi* and *Wolbachia* Symbiosis by Stage-Specific Dual RNA-Seq. *PLoS Neglected Tropical Diseases*. 2017; 11 (3): e0005357. <https://doi.org/10.1371/journal.pntd.0005357> PMID: 28358880
8. Kariuki MM, Hearne LB, Beerntsen BT. Differential Transcript Expression between the Microfilariae of the Filarial Nematodes, *Brugia malayi* and *B. pahangi*. *BMC Genomics*. 2010; 11(1), 225.
9. Li BW, Wang Z, Rush AC, Mitreva M, Weil GJ. Transcription Profiling Reveals Stage- and Function-Dependent Expression Patterns in the Filarial Nematode *Brugia malayi*. *BMC Genomics*. 2012; 13(1), 184.
10. Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, et al. Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity. *Cell*. 2014; 158(6): 1431–43. <https://doi.org/10.1016/j.cell.2014.08.009> PMID: 25215497
11. Dieterich C, & Sommer RJ. A *Caenorhabditis* Motif Compendium for Studying Transcriptional Gene Regulation. *BMC Genomics*. 2008; 9(1), 30.



12. Bailey TL, Williams N, Misleh C, Li WW. MEME: Discovering and Analyzing DNA and Protein Sequence Motifs. *Nucleic Acids Research*. 2006; 34:369–373.
13. Ao W, Gaudet J, Kent WJ, Muttumu S, Mango SE. Environmentally Induced Foregut Remodeling by PHA-4/FoxA and DAF-12/NHR. *Science*. 2004; 305(5691), 1743–1746. <https://doi.org/10.1126/science.1102216> PMID: 15375261
14. Liu X, Brutlag DL, Liu JS. BioProspector: Discovering Conserved DNA Motifs in Upstream Regulatory Regions of Co-expressed Genes. *Pacific Symposium on Biocomputing*. 2001; 6: 127–138.
15. Thijs G, Marchal K, Lescot M, Rombauts S, Moor BD, Rouze P, Moreau Y. A Gibbs Sampling Method to Detect Overrepresented Motifs in the Upstream Regions of Co-expressed Genes. *Journal of Computational Biology*. 2002; 9(2): 447–464. <https://doi.org/10.1089/10665270252935566> PMID: 12015892
16. Pavesi G, Mereghetti P, Mauri G, Pesole G. Weeder Web: Discovery of Transcription Tactor Binding Sites in a Set of Sequences from Co-regulated Genes. *Nucleic Acids Research*. 2004; 32: 199–203.
17. Smith AD, Sumazin P, Zhang MQ. Identifying tissue-selective transcription factor binding sites in vertebrate promoters. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102(5): 1560–1565. <https://doi.org/10.1073/pnas.0406123102> PMID: 15668401
18. Huggins P, Zhong S, Shiff I, Beckerman R, Laptenko O, Prives C, et al. DECOD: Fast and Accurate Discriminative DNA Motif Finding. *Bioinformatics*. 2011; 27(17): 2361–2367. <https://doi.org/10.1093/bioinformatics/btr412> PMID: 21752801
19. Jin VX, Apostolos J, Nagisetty NS, Farnham PJ. W-ChIPMotifs: a Web Application Tool for De Novo Motif Discovery from ChIP-Based High-Throughput Data. *Bioinformatics*. 2009; 25(23): 3191–3193. <https://doi.org/10.1093/bioinformatics/btp570> PMID: 19797408
20. Heeringen SJ van Veenstra GJC. GimmeMotifs: a De Novo Motif Prediction Pipeline for ChIP-sequencing Experiments. *Bioinformatics*. 2011; 27(2), 270–271. <https://doi.org/10.1093/bioinformatics/btq636> PMID: 21081511
21. Quang D, Xie X. DanQ: a Hybrid Convolutional and Recurrent Deep Neural Network for Quantifying the Function of DNA Dequences. *Nucleic Acids Research*. 2016; 44 (11), e107. <https://doi.org/10.1093/nar/gkw226> PMID: 27084946
22. Lee NK, Azizan FL, Wong YS, Omar N. DeepFinder: An Integration of Feature-Based and Deep Learning Approach for DNA Motif Discovery. *Biotechnology & Biotechnological Equipment*. 2018; 32(3): 759–768.
23. Shu L, Katholi CR, Higazi T, Unnasch TR. Analysis of the *Brugia Malayi* HSP70 Promoter Using a Homologous Transient Transfection System. *Molecular and Biochemical Parasitology*. 2003; 128 (1): 67–75. [https://doi.org/10.1016/s0166-6851\(03\)00052-5](https://doi.org/10.1016/s0166-6851(03)00052-5) PMID: 12706798
24. Higazi TB, DeOliveira A, Katholi CR, Shu L, Barchue J, Lisanby M, Unnasch TR. Identification of Elements Essential for Transcription in *Brugia malayi* Promoters. *Journal of Molecular Biology*. 2005; 353 (1), 1–13. <https://doi.org/10.1016/j.jmb.2005.08.014> PMID: 16154590
25. Anders S, Huber W. Differential Expression Analysis for Sequence Count Data. *Genome Biology*. 2010; 11(R106).
26. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics*. 2010; 26(1): 139–140. <https://doi.org/10.1093/bioinformatics/btp616> PMID: 19910308
27. Guiliano DB, Hong X, McKerrow JH, Blaxter ML, Oksov Y, Liu J, Ghedin E, Lustigman S. A Gene Family of Cathepsin L-like Proteases of Filarial Nematodes are Associated with Larval Molting and Cuticle and Eggshell Remodeling. *Mol Biochem Parasitol*. 2004; 136: 227–42. <https://doi.org/10.1016/j.molbiopara.2004.03.015> PMID: 15478801
28. Lustigman S, Zhang J, Liu J, Oksov Y, Hashmi S. RNA Interference Targeting Cathepsin L and Z-like Cysteine Proteases of *Onchocerca volvulus* Confirmed their Essential Function during L3 Molting. *Mol Biochem Parasitol*. 2004; 138: 165–70. <https://doi.org/10.1016/j.molbiopara.2004.08.003> PMID: 15555728
29. Zang X, Maizels RM. Serine Proteinase Inhibitors from Nematodes and the Arms Race between Host and Pathogen. *Trends Biochemical Science*. 2001; 26: 191–197.
30. Heinz S, Benner C, Spann N, Bertolino E, Lin Y, Laslo P, Cheng J, Murre C, Singh H, Glass C. Simple Combinations of Lineage-Determining Transcription Factors Prime Cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell*. 2010; 38(4): 576–589.
31. Shi J, Yang W, Chen M, Du Y, Zhang J, Wang K. AMD, an Automated Motif Discovery Tool Using Step-wise Refinement of Gapped Consensuses. *PLoS One*. 2011; 6 (9): e24576. <https://doi.org/10.1371/journal.pone.0024576> PMID: 21931761

32. Conlon E.M., Liu X.S., Lieb J., & Liu J. (2003). Integrating Regulatory Motif Discovery and Genome-Wide Expression Analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 100 (6): 3339–44. <https://doi.org/10.1073/pnas.0630591100> PMID: 12626739
33. Leping Li. GADEM: A Genetic Algorithm Guided Formation of Spaced Dyads Coupled with an EM Algorithm for Motif Discovery. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*. 2009; 16(2): 317–29.
34. Varoquaux G, Buitinck L, Louppe G, Grisel L, Pedregosa F, Mueller A. Scikit-Learn. *GetMobile: Mobile Computing and Communications*. 2015; 19(1).
35. Gordon AD, Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. *Biometrics*. 1984; 40(30), 874.
36. Quinlan JR. Constructing Decision Trees. *C4.5*. San Francisco: Morgan Kaufmann Publishers Inc; 1993.
37. Mathelier A, Fornes O, Arenillas D, Chen C, Denay G, Lee J, Shi W, et al. JASPAR 2016: A Major Expansion and Update of the Open-Access Database of Transcription Factor Binding Profiles. *Nucleic Acids Research*. 2016; 44 (1): 110–15.
38. Newburger DE, Bulyk ML. UniPROBE: An Online Database of Protein Binding Microarray Data on Protein-DNA Interactions. *Nucleic Acids Research*. 2009; 37: 77–82.
39. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble W. Quantifying Similarity between Motifs. *Genome Biology*. 2007; 8 (2): R24. <https://doi.org/10.1186/gb-2007-8-2-r24> PMID: 17324271
40. Roy S, Wapinski I, Pfiffner J, French C, Socha A, Konieczka J, Habib N, Kellis M, Thompson D, Regev A. Arboretum: Reconstruction and Analysis of the Evolutionary History of Condition-Specific Transcriptional Modules. *Genome Research*. 2013; 23 (6): 1039–50. <https://doi.org/10.1101/gr.146233.112> PMID: 23640720
41. Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. WormBase ParaSite – a Comprehensive Resource for Helminth Genomics. *Molecular and Biochemical Parasitology*. 2017; 215: 2–10. <https://doi.org/10.1016/j.molbiopara.2016.11.005> PMID: 27899279
42. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, et al. Clustal W and Clustal X Version 2.0. *Bioinformatics*. 2007; 23(21): 2947–48. <https://doi.org/10.1093/bioinformatics/btm404> PMID: 17846036
43. Budovskaya YV, Wu K, Southworth L, Jiang M, Tedesco P, Johnson T, Kim S. An E1t-3/elt-5/elt-6 GATA Transcription Circuit Guides Aging in *C. Elegans*. *Cell*. 2008; 134 (2): 291–303. <https://doi.org/10.1016/j.cell.2008.05.044> PMID: 18662544
44. Liu C, Mhashilkar AS, Chabanon J, Xu S, Lustigman S, Adams JH, et al. Development of a Toolkit for piggyBac-Mediated Integrative Transfection of the Human Filarial Parasite *Brugia malayi*. *PLoS Negl Trop Dis*. 2018; 12(5): e0006509. <https://doi.org/10.1371/journal.pntd.0006509> PMID: 29782496
45. Barker GC, Mercer JC, Rees HH, Howells RE. The Effect of Ecdysteroids on the Microfilarial Production of *Brugia pahangi* and the Control of Meiotic Reinitiation in the Oocytes of *Dirofilaria immitis*. *Parasitology Research*. 1991; 77 (1): 65–71. <https://doi.org/10.1007/BF00934388> PMID: 1994372
46. Mhashilkar AS, Adapa SR, Jiang RH, Williams SA, Zaky W, Slatko BE, Luck AN, Moorhead AR, Unnasch TR. Phenotypic and Molecular Analysis of the Effect of 20-Hydroxyecdysone on the Human Filarial Parasite *Brugia malayi*. *Int J Parasitol*. 2016; 46(5–6): 333–41. <https://doi.org/10.1016/j.ijpara.2016.01.005> PMID: 26896576
47. Mhashilkar AS, Vankayala SL, Lui C, Kearns F, Mehrotra P, Tzertzinis G, Palli SR, Woodcock HL, Unnasch TR. Identification of Ecdysone Hormone Receptor Agonists as a Therapeutic Approach for Treating Filarial Infections. *PLoS Neglected Tropical Disease*. 2016; 10(6): e0004772.
48. Warbrick EV, Barker GC, Rees HH, Howells RE. The Effect of Invertebrate Hormones and Potential Hormone Inhibitors on the Third Larval Molt of the Filarial Nematode, *Dirofilaria immitis*, in vitro. *Parasitology*. 1993; 107: 459–463. <https://doi.org/10.1017/s0031182000067822> PMID: 8278225
49. Davis MB, SanGil I, Berry G, Olayokun R, Neves L. Identification of Common and Cell Type Specific LXXLL Motif EcR Cofactors Using a Bioinformatics Refined Candidate RNAi Screen in *Drosophila melanogaster* Cell Lines. *BMC Developmental Biology*. 2011; 11: 66. <https://doi.org/10.1186/1471-213X-11-66> PMID: 22050674
50. Akagi K, Ueda H. Regulatory Mechanisms of Ecdysone-Inducible Blimp-1 Encoding a Transcriptional Repressor that is Important for the Prepupal Development in *Drosophila*. *Development, Growth & Differentiation*. 2011; 53 (5): 697–703.
51. de Oliveira AD, Katholi CR, Unnasch TR. Characterization of the promoter of the *Brugia malayi* 12kDa small subunit ribosomal protein (RPS12) gene. *International Journal of Parasitology*. 2008; 38:1111–9. <https://doi.org/10.1016/j.ijpara.2008.02.002> PMID: 18364245

52. Liu C, Chauhan C, Katholi CR, Unnasch TR. The splice leader addition domain represents an essential conserved motif for heterologous gene expression in *B. malayi*. *Mol Biochem Parasitol*. 2009; 166(1): 15–21. <https://doi.org/10.1016/j.molbiopara.2009.02.004> PMID: 19428668
53. Bailey M, Chauhan C, Liu C, Unnasch TR. The Role of Polymorphisms in the Spliced Leader Addition Domain in Determining Promoter Activity in *Brugia malayi*. *Molecular and Biochemical Parasitology*. 2011; 176:37–41. <https://doi.org/10.1016/j.molbiopara.2010.11.012> PMID: 21111761
54. Liu C, Enright T, Tzertzinis G, Unnasch TR. Identification of Genes Containing Ecdysone Response Elements in the Genome of *Brugia malayi*. *Mol Biochem Parasitol*. 2012; 186: 38–43. <https://doi.org/10.1016/j.molbiopara.2012.09.005> PMID: 23017214
55. Liu C, Vander Kelen P, Ghedin E, Lustigman S, Unnasch TR. Analysis of Transcriptional Regulation of Tetracycline Responsive Genes in *Brugia malayi*. *Molecular and Biochemical Parasitology*. 2011; 180:106–11. <https://doi.org/10.1016/j.molbiopara.2011.09.004> PMID: 21944995
56. Mutafchiev Y, Bain O, Williams Z, McCall JW, Michalski ML. Intra-peritoneal Development of the Filarial Nematode *Brugia malayi* in the Mongolian Jird (*Meriones Unguiculatus*). *Parasitology Research*. 2014; 113(5): 1827–35. <https://doi.org/10.1007/s00436-014-3829-5> PMID: 24664084
57. Babu S, Porte P, Klei TR, Shultz L, Rajan TV. Host NK Cells are Required for the Growth of the Human Filarial Parasite *Brugia malayi* in Mice. *Journal of Immunology*. 1998; 161(3): 1428–1432.
58. Kim D, Pertea D, Trapnell C, Pimentel H, Kelley R, Salzberg S. TopHat2: Accurate Alignment of Transcriptomes in the Presence of Insertions, Deletions and Gene Fusions. *Genome Biology*. 2013; 14 (4): R36. <https://doi.org/10.1186/gb-2013-14-4-r36> PMID: 23618408
59. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential Analysis of Gene Regulation at Transcript Resolution with RNA-Seq. *Nature Biotechnology*. 2013; 31(1): 46–53. <https://doi.org/10.1038/nbt.2450> PMID: 23222703
60. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript Assembly and Quantification by RNA-Seq Reveals Unannotated Transcripts and Isoform Switching during Cell Differentiation. *Nature Biotechnology*. 2010; 28(5): 511–15. <https://doi.org/10.1038/nbt.1621> PMID: 20436464
61. Huang D, Sherman B, Lempicki R. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protoc*. 2009; 4, 44–57.
62. Grant CE, Bailey TL, Noble WS. FIMO: Scanning for Occurrences of a Given Motif. *Bioinformatics*. 2011; 27(7):1017–1018. <https://doi.org/10.1093/bioinformatics/btr064> PMID: 21330290