Microarray-based analysis of differential gene expression between infective and noninfective larvae of Strongyloides stercoralis.

Roshan Ramanathan
Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland

Sudhir Varma
Bioinformatics and Computational Biosciences Branch, National Institutes of Health, Bethesda, Maryland

José M C Ribeiro
Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases National Institutes of Health, Bethesda, Maryland

Timothy G Myers
Microarray Research Facility, National Institutes of Health, Bethesda, Maryland

Recommend Citation
Ramanathan, Roshan; Varma, Sudhir; Ribeiro, José M C; Myers, Timothy G; Nolan, Thomas J; Abraham, David; Lok, James B; and Nutman, Thomas B, "Microarray-based analysis of differential gene expression between infective and noninfective larvae of Strongyloides stercoralis." (2011). Department of Microbiology and Immunology Faculty Papers. Paper 22.
https://jdc.jefferson.edu/mifp/22

This Article is brought to you for free and open access by the Jefferson Digital Commons. The Jefferson Digital Commons is a service of Thomas Jefferson University's Center for Teaching and Learning (CTL). The Commons is a showcase for Jefferson books and journals, peer-reviewed scholarly publications, unique historical collections from the University archives, and teaching tools. The Jefferson Digital Commons allows researchers and interested readers anywhere in the world to learn about and keep up to date with Jefferson scholarship. This article has been accepted for inclusion in Department of Microbiology and Immunology Faculty Papers by an authorized administrator of the Jefferson Digital Commons. For more information, please contact:
JeffersonDigitalCommons@jefferson.edu.
Microarray-Based Analysis of Differential Gene Expression between Infective and Noninfective Larvae of Strongyloides stercoralis

Roshan Ramanathan1, Sudhir Varma2, José M. C. Ribeiro3, Timothy G. Myers4, Thomas J. Nolan5, David Abraham6, James B. Lok5, Thomas B. Nutman1
1 Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, United States of America, 2 Bioinformatics and Computational Biosciences Branch, National Institutes of Health, Bethesda, Maryland, United States of America, 3 Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases National Institutes of Health, Bethesda, Maryland, United States of America, 4 Microarray Research Facility, National Institutes of Health, Bethesda, Maryland, United States of America, 5 Department of Pathobiology, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, 6 Department of Microbiology and Immunology, Thomas Jefferson University, Philadelphia, Pennsylvania, United States of America

Abstract

Background: Differences between noninfective first-stage (L1) and infective third-stage (L3i) larvae of parasitic nematode Strongyloides stercoralis at the molecular level are relatively uncharacterized. DNA microarrays were developed and utilized for this purpose.

Methods and Findings: Oligonucleotide hybridization probes for the array were designed to bind 3,571 putative mRNA transcripts predicted by analysis of 11,335 expressed sequence tags (ESTs) obtained as part of the Nematode EST project. RNA obtained from S. stercoralis L3i and L1 was co-hybridized to each array after labeling the individual samples with different fluorescent tags. Bioinformatic predictions of gene function were developed using a novel cDNA Annotation System software. We identified 935 differentially expressed genes (469 L3i-biased; 466 L1-biased) having two-fold expression differences or greater and microarray signals with a p value < 0.01. Based on a functional analysis, L1 larvae have a larger number of genes putatively involved in transcription (p = 0.004), and L3i larvae have biased expression of putative heat shock proteins (such as hsp-90). Genes with products known to be immunoreactive in S. stercoralis-infected humans (such as SsIR and NIE) had L3i biased expression. Abundantly expressed L3i contigs of interest included S. stercoralis orthologs of cytochrome oxidase ucr 2.1 and hsp-90, which may be potential chemotherapeutic targets. The S. stercoralis ortholog of fatty acid and retinol binding protein-1, successfully used in a vaccine against Ancylostoma ceylanicum, was identified among the 25 most highly expressed L3i genes. The sperm-containing glycoprotein domain, utilized in a vaccine against the nematode Cooperia punctata, was exclusively found in L3i biased genes and may be a valuable S. stercoralis target of interest.

Conclusions: A new DNA microarray tool for the examination of S. stercoralis biology has been developed and provides new and valuable insights regarding differences between infective and noninfective S. stercoralis larvae. Potential therapeutic and vaccine targets were identified for further study.


Editor: Elodie Ghedin, University of Pittsburgh, United States of America

Received July 14, 2010; Accepted March 16, 2011; Published May 3, 2011

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CCO public domain dedication.

Funding: This work was supported by the Intramural Research Program of the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health and by NIH grants AI-050688, AI-022662, AI-082548, and RR02512. 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: ramanathanr@niaid.nih.gov

Introduction

Strongyloides stercoralis is a parasitic nematode endemic to the tropics and subtropics that infects an estimated 30–100 million people worldwide. Chronically infected individuals have the potential to develop hyperinfection syndrome or disseminated disease, clinical entities that carry a very high (87–100%) mortality if unrecognized [1].

Free-living S. stercoralis infective third stage (L3i) larvae residing in the soil penetrate intact skin and blood vessels, ultimately developing to adults in the small intestine. Adult females, typically residing in the duodenum of the host, produce eggs by mitotic parthenogenesis that develop into first-stage (L1) larvae that are excreted into the stool. L1 larval progeny of parasitic females develop into free-living adults unless triggered by genetic, environmental, or host factors to develop directly into L3i larvae [2,3]. Despite sharing many characteristics, L1 and L3i larvae can be distinguished by their behavior and morphology. L1 larvae have a short, trilobed pharynx and expend much of their energy on feeding and growth [3]. L3i larvae, by contrast, can survive in...
Author Summary

*Strongyloides stercoralis* is a soil-transmitted helmint that affects an estimated 30–100 million people worldwide. Chronically infected persons who are exposed to corticosteroids can develop disseminated disease, which carries a high mortality (87–100%) if untreated. Despite this, little is known about the fundamental biology of this parasite, including the features that enable infection. We developed the first DNA microarray for this parasite and used it to compare infective third-stage larvae (L3i) with non-infective first stage larvae (L1). Using this method, we identified 935 differentially expressed genes. Functional characterization of these genes revealed L3i biased expression of heat shock proteins and genes with products that have previously been shown to be immunoreactive in infected humans. Genes putatively involved in transcription were found to have L1 biased expression. Potential chemotherapeutic and vaccine targets such as far-1, ucr 2.1 and hsp-90 were identified for further study.

harsh environmental conditions, enabled by a comparatively thickened cuticle, constricted gastrointestinal tract, and closed mouth. These larvae are developmentally arrested, non-feeding, stress resistant, and long lived [3–5].

A high degree of specificity between these stages has been suggested by expressed sequence tag (EST) based analysis of free living L1 and L3i larvae for *S. stercoralis* [6–8]. These comparisons, however, are based on short reads of cDNA libraries and assumptions about abundance. There remain many unanswered questions about the basic molecular features underlying the apparent morphologic and behavioral differences between these larval stages. An improved understanding of these differences can provide insights into what defines infectivity and may ultimately prove useful in defining targets for the development of vaccines and therapeutics against this parasite.

In order to answer these questions, a DNA microarray tool for *S. stercoralis* – the species causing the vast majority of human infection worldwide - is needed. Although a DNA microarray has recently been developed for *Strongyloides ratti*, the natural parasite of brown rats (Rattus norvegicus) [9], previous work has suggested little conservation of gene expression profiles between these two species [10], underscoring the need for a DNA microarray specific to this species.

The availability of a *S. stercoralis* DNA microarray enables comparative analyses across nematodes, which can be utilized to further our understanding of the biologic determinants of parasitism. The free-living, non-parasitic, nematode *C. elegans* has been used as a model species for comparison with *S. stercoralis*. *C. elegans* dauer stage larvae and *S. stercoralis* L3i larvae share many morphologic and physiologic characteristics. The 'dauer hypothesis' recognizes these similarities and suggests that the same molecular genetic mechanisms control the morphogenesis of these stages [11]. Comparative genomics of gene expression based on EST abundance data for *S. stercoralis* suggests a higher degree of similarity between *S. stercoralis* L1 and *C. elegans* non-dauer expressed genes [6]. By contrast, a robust 'dauer-L3i expression signature' has not been found [6]. A comparative analysis based on microarray expression data for these species could prove useful not only in identifying a 'dauer-L3i expression signature' should it exist, but also in uncovering potentially significant determinants of *S. stercoralis* L3i infectivity.

The purpose of this study was to: 1) develop and optimize a DNA microarray tool for *S. stercoralis*, 2) utilize this microarray to examine differences in gene expression between L3i and L1 larvae and 3) perform a comparative microarray analysis between parasitic *S. stercoralis* and non-parasitic *C. elegans* in order to develop further insights into the biologic determinants of parasitism.

Methods

Ethics statement

Animal handling and experimental procedures were undertaken in compliance with the University of Pennsylvania’s Institutional Animal Care and Use Committee (IACUC) guidelines. Ethical approval was obtained for the study (protocol number 702342) from IACUC (University of Pennsylvania, Philadelphia, PA).

Parasites

All larvae used in this analysis were obtained from laboratory dogs infected with *S. stercoralis*, UPD strain [12]. Fecal samples from dogs were processed using the charcoal coproculture followed by Baermann funnel technique, as outlined elsewhere [13]. Post parasitic L1 larvae were recovered from freshly deposited stool samples; L3i larvae were recovered after 7 days of stool incubation at 25°C. L3i larvae underwent surface decontamination by migration through low-melting-point agarose. L1 larvae were decontaminated by 3 washes with phosphate buffered saline (PBS) containing an antibiotic cocktail. Decontaminated parasites were subsequently stored in Trizol reagent (Invitrogen, San Diego, CA) at −80°C. Using this method, 30,700 post-parasitic L1 and 50,000 L3i larvae were collected.

Isolation of total RNA from larvae

Total RNA was extracted by thawing pooled samples of L1 and L3i larvae at 37°C in a warm water bath and centrifuging the samples at 4°C (805 × g) for 10 minutes to obtain a pellet. The pellet was frozen in liquid nitrogen, ground thoroughly with an autoclaved mortar and pestle and then purified using an RNAeasy mini kit (Qiagen, Valencia, CA) following the manufacturer’s protocol. A Nano Drop-1000 spectrophotometer (NanoDrop Technologies, Santa Clara, CA) was used to determine the RNA concentration in each sample. RNA was more precisely quantified and quality assessed using the 2100 Bioanalyzer (Agilent, Santa Clara, CA).

Amplification and labeling

RNA samples from L1 and L3i stage larvae were co-hybridized using Cy3 and Cy5 labels to discriminate the relative level of target bound to the microarray probe. Fluorescent-labeled cDNA targets were prepared from total RNA using the Ovation amino-allyl kit (NuGEN, San Carlos, CA) according to the manufacturer's protocol. A Nano Drop-1000 spectrophotometer (NanoDrop Products, Wilmington DE) was used to determine the RNA concentration in each sample. RNA was more precisely quantified and quality assessed using the 2100 Bioanalyzer (Agilent, Santa Clara, CA).

Hybridization procedure

Labeled samples were combined with blocking components poly(dA), yeast tRNA, and human Cot-1, in hybridization buffer composed of 25% formamide/5× saline-sodium citrate (SSC)/0.2% (w/v) sodium dodecyl sulfate (SDS) to a total volume of 60 μl. After heating the sample (95°C for 3 minutes), it was centrifuged (20,000 × g) for 3 minutes. Fifty eight μl of the sample (1.6 μg of labeled cDNA) was loaded onto the microarray chip. The microarray chips were hybridized overnight at 45°C using the MicroArray User Interface (MAUI) hybridization system (BioMicro Systems, Inc., Salt Lake City, UT). The following day, the chips were washed twice in 1× SSC/0.05% (w/v) SDS buffer
(3 minutes each wash) and twice in 0.1 × SSC buffer (5 minutes each wash).

For the present study, four technical replicate experiments using pooled L1 and L3i larvae were performed, including one dye swap. The microarray chips were imaged using a GenePix 4000 B scanner (Molecular Devices, Sunnyvale, CA). Agilent Feature Extraction software was used for image analysis, protocol GE2-v5.10 Apr08. The data discussed in this publication have been deposited in the National Center for Biotechnical Information (NCBI) Gene Expression Omnibus (GEO) and are accessible through GEO Series accession number GSE24735 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE24735).

Microarray design
ESTs (11,335) were identified from L1 and L3i cDNA libraries created as part of the nematode EST project [6,7]. ESTs were organized into 3,571 contigs by bioinformatics analysis [14]. Oligonucleotide probes designed to hybridize with these contigs were used to develop early versions (V1 and V2) of chips manufactured by Combimatrix (Irvine, CA) based on a variety of algorithms for oligonucleotide design. Versions 1 and 2 were assessed for performance using RNA from L1 and L3i larvae. After testing the performance of these two versions of the arrays, an optimized version (V3) was developed. The best probe for each target was selected based on the average signal intensity for all arrays and the number of arrays with detectable signal. The spot density was 22K spots per array. Of the six oligonucleotides designed per target, one was designed using the Array Designer program (Premier Biosoft International, Palo Alto, CA), two were designed using E-Array (Agilent, Santa Clara, CA) using the “base composition” method (replicated twice), two were designed using E-array “best Tm” method, and the last was a 40-mer designed using Array Designer. Probes were selected to avoid cross-hybridization to other sequences in the target (contig) dataset manufactured by Agilent SurePrint. The probes designed to make the V3 microarray are found in Table S1 in Supporting Information Text S1.

Functional annotation
All data were exported into the cDNA Annotation System (dCAS) [14,15]. This tool enabled annotation of each S. stercoralis contig based on Basic Local Alignment Search Tool (BLAST) alignments against multiple databases (NCBI nr protein database (NR), Gene Ontology (GO), eukaryotic Orthologous Groups (KOG), Pfam protein families database (PFAM), Simple Modular Architecture Research Tool (SMART), Wormbase (CELEG), and Saccharomyces genome database (YESAT) and provided the corresponding E-values. The database was also annotated manually with a composite categorization that summarized the findings across databases. The entire annotated database, with hyperlinks to the NIAID exon website, is accessible for download at: http://exon.niaid.nih.gov/transcriptome/S_stercoralis/SS-Supp-Web.zip. A stand-alone version can also be accessed and downloaded at: http://exon.niaid.nih.gov/transcriptome/S_stercoralis/SS-Supp-StandAlone.zip. Extract the excel file and the links directory to your own computer for browsing the hyperlinks locally.

Statistical analysis
Spot values were calculated using a linear lowess dye normalization. Further, the 50th percentile of a set containing all the ribosomal genes in the array was applied to all spot values. In cases of multiple spots for the same S. stercoralis contig, the average of the log2 signal was calculated for each array. The mean signal ratio (log2 L3i/L1) was calculated from the signals for all 4 arrays. No surrogate values were applied. A single group t-test analysis was calculated on the data set. Variance shrinkage was not used when calculating p-values for differential expression. Differentially expressed genes were identified using a ‘cutoff’ of 2 fold expression difference or greater for log2 L3i/L1 signal ratios, and p<0.01 for microarray signal data (false discovery rate (FDR) = 2.5%).

Functional analysis
A functional analysis was performed based on annotations provided by each database (Pfam, SMART, KOG, etc.). The number of genes per functional category (e.g., transcription, cytoskeleton, metabolism, etc.) was compared between L1 and L3i differentially expressed genes (as defined by the above cutoff). To ascertain whether genes belonging to certain functional classes were more likely to be highly expressed in one stage or another, we used a statistical test for one proportion using Normal approximation. Assuming a null proportion of 0.5 (i.e., that there is no difference in the number of genes of that category for the two classes), p values were calculated for deviation from 0.5 using Normal approximation. P values were adjusted for multiple comparisons using the Bonferroni criterion.

Gene-set enrichment analysis
Gene Set Enrichment Analysis (GSEA) is a robust method for analyzing molecular profiling data examines the clustering of a pre-defined group of genes (gene set) across the entire microarray database [12] using the Kolmogorov-Smirnov test. Gene sets were compiled by first downloading GO categories from Wormbase (www.wormbase.org) for C. elegans genes. Definitions for each GO category used can be found at http://www.wormbase.org/db/ontology/gene. S. stercoralis orthologs for C. elegans genes were determined by dCAS based on BLAST alignments to the C. elegans gene. BLAST matches with E values<0.05 were excluded. Gene sets with fewer than 5 S. stercoralis orthologs were excluded from GSEA analysis. Using these criteria, 18 S. stercoralis gene sets were created (see Figure 1A). Additional manually compiled gene sets included the group of S. stercoralis genes whose products have been shown to be immunoreactive in humans infected with S. stercoralis [17–19], and a group of putatively identified heat shock proteins.

Comparative microarray analysis of S. stercoralis and C. elegans
Microarray expression data for S. stercoralis L3i and C. elegans dauer larvae were compared using several methods as follows: 1) We defined three gene sets comprising the S. stercoralis orthologs of “dauer-enriched” C. elegans genes derived from either C. elegans microarray expression data alone, both serial analysis of gene expression (SAGE) and microarray expression data or from the Gene Ontology category dauer larval development (Figure 1A) [20,21]. We then used GSEA to determine whether these gene sets
showed significant L3i enrichment. 2) We examined whether a correlation exists between *C. elegans* dauer/L1 microarray expression data obtained by Wang and colleagues [20] with our *S. stercoralis* L3i/L1 microarray expression data. The previously obtained *C. elegans* microarray expression data can be found at http://cmgm.stanford.edu/kimlab/dauer/ExtraData.htm, Table S1 in Supporting Information Text S1, column “AdjD/L1_Ratio” which corresponds to the average log2 expression ratio for each gene set.

---

**Figure 1. Gene Set Enrichment Analysis and enrichment plot.** A. Gene sets were compiled by listing the *S. stercoralis* orthologs of *C. elegans* genes assigned to Gene Ontology (GO) categories (downloaded from www.wormbase.org). Some gene sets were manually compiled. Only gene sets with at least 5 *S. stercoralis* contigs were included in this analysis. The results of the GSEA are listed for each gene set. The enrichment score reflects the degree to which each gene set is represented at the top or bottom of the list of 3,571 contigs ranked by fold change (L3i enriched = more positive, L1 enriched = more negative). The normalized enrichment score can be used to compare results across gene sets. The nominal p value estimates the statistical significance of the enrichment for a single gene set and does not correct for gene set size and multiple hypothesis testing. *ID = Gene Ontology Identification.** The False Discovery Rate (FDR) is adjusted for gene set size and multiple hypotheses testing. B. This plot depicts the distribution of individual genes (vertical black lines) encoding immunoreactive antigens recognized by sera from patients infected with *S. stercoralis*. This gene set was analyzed against a list of 3,571 *S. stercoralis* contigs ranked by fold change of log2 L3i/L1 mean signal ratios. The clustering of individual genes towards the left side of the list (above the red bar) suggests L3i-biased enrichment of this gene set. These genes are individually listed in Table S8 in Supporting Information Text S1. doi:10.1371/journal.pntd.0001039.g001
values for *C. elegans* dauer larva at time 0 relative to L1 larva [20]. 3 Using these data, we calculated the absolute value of the difference between fold change values for *C. elegans* genes and their *S. stercoralis* orthologs (C. elegans dauer/L1 fold change - S. stercoralis L3i/L1 fold change). Only those genes with robust microarray expression data (p values <0.01) were included. In order to identify those genes that are expressed differently by *S. stercoralis* L3i and *C. elegans* dauer larva, a list was generated of all *S. stercoralis*- *C. elegans* orthologs with the greatest differences in fold change values (absolute value >2). The list was further narrowed to include only those *S. stercoralis*- *C. elegans* gene pairs where gene expression was regulated in opposite directions between the two nematodes (Table 1).

Microarray validation by quantitative real-time polymerase chain reaction (qPCR)

The sequences of L3i biased genes (contigs 24, 25, 65, 243, 2136) and L1 biased genes (contigs 55, 222, 307, 2328) were used to create primer-probe sets designed and manufactured by Applied Biosystems (Foster City, CA). The sequences for these primer probes are listed in Table S2 in Supporting Information Text S1. The *S. stercoralis* control genes for qPCR analysis was *S. stercoralis* glyceraldehyde 3 phosphate dehydrogenase (GAPDH; GenBank accession number BI77392; contig 90; log2 L3i/L1 = −0.28179). Post-parasitic L1 and L3i larvae (distinct from those hybridized onto the microarray) were collected and total RNA made as described above. Total RNA (1 μg) from L1 and L3i larvae was used to synthesize cDNA. qPCR was performed using all 9 primer probe sets in separate reactions with L1 cDNA and also with L3i cDNA. The reaction was performed using 10× RT buffer (10 μl), 25 mM MgCl2 (22 μl), dNTP (20 μl), random hexamers (5 μl), RNase inhibitor (2 μl), and multiscribe reverse transcriptase (50 U/μl; 6.25 μl) in a microamp 96-well reaction plate (Applied Biosystems). De-ionized, distilled water was added to total volume of 65.25 μl. Cycling conditions were: 25°C for 10 minutes, 37°C for 60 minutes, 95°C for 5 minutes, then 4.0°C. Each experiment was performed in triplicate. The mean negative delta threshold cycle (delta C_T) was calculated for each sample. The data generated by performing qPCR using primer probes for 9 contigs on L1 and L3i cDNA (n = 18) was plotted

<table>
<thead>
<tr>
<th>C. elegans match</th>
<th><em>S. stercoralis</em> contig</th>
<th>Putative identification</th>
<th>Fold change direction</th>
<th>C. elegans (dauer/L1)</th>
<th>S. stercoralis (L3i/L1)</th>
<th>Fold change</th>
<th>Absolute difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>C130.9.8</td>
<td>1269</td>
<td>pek-1 eukaryotic translation initiation factor 2 alpha kinase PEK</td>
<td>↑</td>
<td>3.59</td>
<td>0.39</td>
<td>3.20</td>
<td></td>
</tr>
<tr>
<td>T14D7.2</td>
<td>1652</td>
<td>trp-2 trehalose phosphate synthase</td>
<td>↑</td>
<td>21.75</td>
<td>0.40</td>
<td>21.36</td>
<td></td>
</tr>
<tr>
<td>C06B3.4</td>
<td>2873</td>
<td>stdh-1 estradiol 17 beta-dehydrogenase</td>
<td>↑</td>
<td>3.71</td>
<td>0.44</td>
<td>3.27</td>
<td></td>
</tr>
<tr>
<td>T02D1.5</td>
<td>1790</td>
<td>pmp-4 ABC transporters</td>
<td>↑</td>
<td>16.57</td>
<td>0.45</td>
<td>16.12</td>
<td></td>
</tr>
<tr>
<td>T19B10.2</td>
<td>504</td>
<td>protein_id:CAA98547</td>
<td>↑</td>
<td>2.56</td>
<td>0.55</td>
<td>2.01</td>
<td></td>
</tr>
<tr>
<td>T04B2.5</td>
<td>2003</td>
<td>protein_id:CAA92628</td>
<td>↑</td>
<td>3.97</td>
<td>0.56</td>
<td>3.41</td>
<td></td>
</tr>
<tr>
<td>F11B10.1</td>
<td>1269</td>
<td>pek-1 eukaryotic translation initiation factor 2 alpha kinase PEK</td>
<td>↑</td>
<td>3.61</td>
<td>0.57</td>
<td>3.04</td>
<td></td>
</tr>
<tr>
<td>F1086.5</td>
<td>1652</td>
<td>trp-2 trehalose phosphate synthase</td>
<td>↑</td>
<td>4.22</td>
<td>0.65</td>
<td>3.57</td>
<td></td>
</tr>
<tr>
<td>Y53G11C.15</td>
<td>2873</td>
<td>stdh-1 estradiol 17 beta-dehydrogenase</td>
<td>↑</td>
<td>4.60</td>
<td>0.66</td>
<td>3.94</td>
<td></td>
</tr>
<tr>
<td>F42D1.2</td>
<td>118</td>
<td>tyrosine aminotransferase</td>
<td>↑</td>
<td>2.92</td>
<td>0.76</td>
<td>2.16</td>
<td></td>
</tr>
<tr>
<td>F42D1.2</td>
<td>117</td>
<td>tyrosine aminotransferase</td>
<td>↑</td>
<td>0.17</td>
<td>2.22</td>
<td>2.05</td>
<td></td>
</tr>
<tr>
<td>C53B1.4</td>
<td>2200</td>
<td>col-119 collagen</td>
<td>↑</td>
<td>0.17</td>
<td>2.87</td>
<td>2.69</td>
<td></td>
</tr>
<tr>
<td>C28C2.1</td>
<td>836</td>
<td>egf-10 G-protein beta subunit GPB-2</td>
<td>↑</td>
<td>0.10</td>
<td>2.98</td>
<td>2.88</td>
<td></td>
</tr>
<tr>
<td>B0491.9</td>
<td>291</td>
<td>protein_id:CAA90887</td>
<td>↑</td>
<td>0.78</td>
<td>2.99</td>
<td>2.21</td>
<td></td>
</tr>
<tr>
<td>E02H1.7</td>
<td>3417</td>
<td>nhr-19 Zinc finger, C4 type (two domains)</td>
<td>↑</td>
<td>0.84</td>
<td>3.18</td>
<td>2.34</td>
<td></td>
</tr>
<tr>
<td>C02F12.7</td>
<td>64</td>
<td>tag-278</td>
<td>↑</td>
<td>0.35</td>
<td>3.25</td>
<td>2.90</td>
<td></td>
</tr>
<tr>
<td>C53B4.5</td>
<td>2267</td>
<td>col-119 collagen</td>
<td>↓</td>
<td>0.27</td>
<td>3.89</td>
<td>3.62</td>
<td></td>
</tr>
<tr>
<td>E02H1.7</td>
<td>1846</td>
<td>nhr-19 Zinc finger, C4 type (two domains)</td>
<td>↑</td>
<td>0.10</td>
<td>4.11</td>
<td>4.01</td>
<td></td>
</tr>
<tr>
<td>C03B1.12</td>
<td>785</td>
<td>lmp-1</td>
<td>↓</td>
<td>0.35</td>
<td>4.18</td>
<td>3.83</td>
<td></td>
</tr>
<tr>
<td>ZK965.2</td>
<td>9</td>
<td>col-37 collagen status</td>
<td>↓</td>
<td>0.70</td>
<td>5.09</td>
<td>4.39</td>
<td></td>
</tr>
<tr>
<td>B0365.5</td>
<td>220</td>
<td>eat-6 Na(+)K(+)/APase alpha subunit</td>
<td>↓</td>
<td>0.11</td>
<td>6.74</td>
<td>6.63</td>
<td></td>
</tr>
<tr>
<td>B0491.2</td>
<td>358</td>
<td>CDC2 status</td>
<td>↑</td>
<td>0.71</td>
<td>6.83</td>
<td>6.12</td>
<td></td>
</tr>
<tr>
<td>F42B3.4</td>
<td>427</td>
<td>nmy-2 myosin heavy chain status</td>
<td>↓</td>
<td>0.71</td>
<td>8.28</td>
<td>7.58</td>
<td></td>
</tr>
<tr>
<td>F42D1.2</td>
<td>116</td>
<td>tyrosine aminotransferase</td>
<td>↓</td>
<td>0.89</td>
<td>8.33</td>
<td>7.44</td>
<td></td>
</tr>
</tbody>
</table>

*Shown only are *S. stercoralis*- *C. elegans* orthologs with an absolute difference >2. All *S. stercoralis*- *C. elegans* orthologs are BLAST matches with an E value <0.05. The p value for *S. stercoralis* microarray signal data was <0.01.

**Arrows indicate whether genes had increased (↑) or decreased (↓) expression in *C. elegans* dauer or *S. stercoralis* L3i larva relative to its respective L1 stage.

The values in this column were calculated by taking the absolute value of [fold change *C. elegans* dauer/L1 - fold change *S. stercoralis* L3i/L1]. *C. elegans* expression data were previously obtained by Wang and colleagues [20].

doi:10.1371/journal.pntd.0001039.t001
against the average L1 and L3i intensity signals for each gene (Figure 2).

Results

Identification of differentially expressed genes

A total of 3,571 distinct contigs were studied by this microarray analysis (Table S3 in Supplemental Information Text S1). Using pre-defined cutoffs, 935 contigs were identified as differentially expressed as shown in the volcano plot (Figure 3). Of these, 466 genes were L1 biased (Table S4 in Supporting Information Text S1) and 469 genes were L3i biased (Table S5 in Supporting Information Text S1). Among the 25 most highly expressed L3i genes were the S. stercoralis orthologs of fatty acid/retinol binding protein-1 (contig 115); 11 fold expression difference), a ferritin chain homolog (contig 94; 14 fold expression difference), and one of four putative trehalases (contig 68; 14 fold expression difference). Among the 25 most highly expressed L1 genes were electron transport chain proteins such as NADH dehydrogenase (contig 371; 0.13-fold change); cytochrome b (contig 2328; 0.19 fold change) and cytochrome c oxidase subunit 1 (contig 55; 0.29 fold change). The 25 most highly expressed L1 or L3i genes are listed in Table S6 in Supporting Information Text S1.

Functional analysis of L1 and L3i biased genes

A greater number of L1 (n = 40) than L3i biased (n = 18) genes were putatively involved in transcription (p = 0.004, not Bonferroni adjusted; see Figure 4A,B). A complete listing of these genes is shown in Figure 4B. This finding was also noted in an analysis of classifications based on GO categories (p = 0.01 for ‘transcription’), and manual annotations (p = 0.007 for ‘transcription machinery’), although p values were not <0.05 when Bonferroni-adjusted for multiple comparisons. BLAST matches to SMART and Pfam databases both indicated that the sperm-containing glycoprotein (SCP) domain was found exclusively in the L3i-group (n = 13 orthologs of fatty acid/retinol binding protein-1 (contig

Comparative microarray analysis of S. stercoralis and C. elegans

Four hundred and twenty two of 3,571 S. stercoralis contigs had C. elegans orthologs for which robust microarray signal data were available. When C. elegans and S. stercoralis microarray signals were plotted against each other, a poor and non-significant correlation was found (Spearman rank = 0.06; p = 0.2444, graph not shown). No significant L3i enrichment of S. stercoralis orthologs of C. elegans ‘dauer enriched’ genes was found by GSEA (nominal p-value = 0.034, FDR = 0.56). For an annotated list of the individual genes enriched in each of these categories, refer to Tables S8 and S9 in Supporting Information Text S1. None of the 18 gene sets were enriched in the L1 phenotype.

Correlation between EST and microarray data

A statistically significant positive correlation was found between microarray expression data and EST abundance data (p<0.0001; max R² = 0.26; graph not shown).
A. Functional Analysis of Differentially Expressed S. stercoralis L1 and L2-Biased Genes

B. S. stercoralis L1 and L2-Biased Genes Putatively Involved in Transcription

<table>
<thead>
<tr>
<th>Genes</th>
<th>Positive Identification</th>
<th>Genes</th>
<th>Positive Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td></td>
<td>L2</td>
<td></td>
</tr>
<tr>
<td>562</td>
<td>Transcription factor RPB1 and related Hox domain proteins</td>
<td>565</td>
<td>Transcription regulator MAPKAP1, Cdc42, LIM kinase superfamily</td>
</tr>
<tr>
<td>554</td>
<td>Transcription initiation factor TFIID, subunit TBP and related transcription factors</td>
<td>567</td>
<td>Transcription factor containing Hoxc and Tis domain</td>
</tr>
<tr>
<td>551</td>
<td>Retinoblastoma-binding transcription factor/Snf2-like transport protein</td>
<td>556</td>
<td>RNA polymerase II general transcription factor SBF and related proteins</td>
</tr>
<tr>
<td>530</td>
<td>RNA polymerase II large subunit</td>
<td>550</td>
<td>Catalytic subunit transcription coactivator</td>
</tr>
<tr>
<td>539</td>
<td>Transcription factor MEIS2 and related Hox domain proteins</td>
<td>440</td>
<td>Transcriptional regulator</td>
</tr>
<tr>
<td>537</td>
<td>Transcription initiation factor TFIID, subunit TAF6</td>
<td>446</td>
<td>Transcriptional regulator</td>
</tr>
<tr>
<td>1500</td>
<td>Retinoblastoma nuclear factor I and similar family homeobox receptor</td>
<td>519</td>
<td>Transcription factor SBF and related proteins, contains Y-box DNA binding domain</td>
</tr>
<tr>
<td>1202</td>
<td>Nucleosome compensation complex, subunit MLB</td>
<td>770</td>
<td>ZNF2 box-Zn finger protein</td>
</tr>
<tr>
<td>1277</td>
<td>RNA helicase factor</td>
<td>928</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>1551</td>
<td>Thyroid hormone receptor-associated protein complex, subunit RAP35</td>
<td>1368</td>
<td>Oxidation and related transcription factors</td>
</tr>
<tr>
<td>1557</td>
<td>Retinoblastoma-binding transcription factor/Snf2-like transport protein</td>
<td>1418</td>
<td>Retinoblastoma transcription factor</td>
</tr>
<tr>
<td>1563</td>
<td>GATA-binding DNA-binding (nuclear receptor binding) factor 1</td>
<td>1513</td>
<td>RNA polymerase II, large subunit</td>
</tr>
<tr>
<td>1595</td>
<td>RNA polymerase I, second largest subunit</td>
<td>1521</td>
<td>Retinoblastoma transcription factor</td>
</tr>
<tr>
<td>1615</td>
<td>Retinoblastoma nuclear factor I and similar family homeobox receptor</td>
<td>1719</td>
<td>Retinoblastoma factor MEF2, RNA related transcription factor</td>
</tr>
<tr>
<td>1784</td>
<td>Transcription factor</td>
<td>3681</td>
<td>DNA repair transcription protein Mre11</td>
</tr>
<tr>
<td>1953</td>
<td>Retinoblastoma nuclear factor I and similar family homeobox receptor</td>
<td>3534</td>
<td>Transcriptional coactivator CBP/p300- related domain</td>
</tr>
<tr>
<td>1929</td>
<td>MEIS2 transcription factor</td>
<td>3569</td>
<td>Retinoblastoma transcription coactivator protein</td>
</tr>
<tr>
<td>2021</td>
<td>Transcription factor II DNA binding, subunit B1 and accl domain</td>
<td>3530</td>
<td>DNA repair transcription factor MAT1</td>
</tr>
<tr>
<td>3025</td>
<td>LPAR2 by PE signaling protein</td>
<td>3065</td>
<td>DNA repair transcription factor MAT1</td>
</tr>
<tr>
<td>3585</td>
<td>DNA-binding / DNA-binding domain containing / DNA-repair protein 3</td>
<td>3586</td>
<td>DNA repair transcription factor MAT1</td>
</tr>
<tr>
<td>3599</td>
<td>DNA-binding / DNA-binding domain containing / DNA-repair protein 3</td>
<td>3599</td>
<td>DNA repair transcription factor MAT1</td>
</tr>
</tbody>
</table>

www.plosntds.org 7 May 2011 | Volume 5 | Issue 5 | e1039
Validation of microarray data with qPCR

A positive correlation was found (Spearman rank = 0.4778; p = 0.0449) between average L1 or L3i microarray intensity signals and mean negative delta Ct of qPCR (Figure 2).

Discussion

In this microarray based analysis of differential gene expression between infective and noninfective *S. stercoralis* larvae, we uncovered differences in the expression of genes putatively encoding transcription factors, heat shock proteins and antigens known to be immunoreactive in sera from infected humans. A comparative microarray analysis of our data revealed several differences between *S. stercoralis* L3i and *C. elegans* dauer stage larvae, such as in the expression of genes putatively encoding collagen and myosin. Potential therapeutic and vaccine targets were identified for further study.

L1 larvae appear to be transcriptionally more active

Analogous to their non-dauer *C. elegans* counterparts, actively growing *S. stercoralis* L1 larvae are thought to have higher rates of transcription relative to L3i-stage larvae. This supposition is based on comparisons between *C. elegans* non-dauer biased genes and *S. stercoralis* L1-biased genes that suggest transcriptional conservation of genes involved in early larval growth [6]. Consistent with this finding, we found L1 biased expression of genes putatively involved in transcription. Among the *S. stercoralis* L1-biased genes involved in transcription were transcription initiation factors (contigs 3245, 1037, 680), transcription factors (contigs 1905, 1277, 891, 2023, 2446, 1036, 1794, 592, 2210), and subunits of RNA polymerase (contigs 1505, 3218, 1020, 2917). By contrast, the L3i-biased genes involved in transcription though fewer, included transcriptional regulators (contigs 446, 445, 156) as well as transcription factors (contigs 1521, 519, 836, 167, 1478), implying that L3i larvae are not transcriptionally inactive and may regulate transcription differently. This would be consistent with what is known of *C. elegans* dauer larvae, which express distinct sets of dauer-specific genes at certain time points (dauer exit, for example) [20,21].

L3i biased expression of genes with products that have been shown to be immunoreactive in *S. stercoralis*-infected humans

Not surprisingly, genes encoding *S. stercoralis* antigens known to produce robust antibody responses in infected humans were found to have L3i biased expression by GSEA [17–19]. Two of these genes, IgG immunoreactive antigen (SsIR) and NIE antigen, have been recently employed in serodiagnostic assays with some advantage over crude antigen [19]. The finding that genes with products capable of inducing protective immunity demonstrate stage-biased gene expression supports the further investigation of these genes as vaccine candidates.

Heat shock proteins have been shown to play a critical role in determining parasite survival during stressful conditions because they can bind denatured or misfolded proteins [22,23]. Biased expression of genes encoding heat shock proteins in the *S. stercoralis* L3i relative to L1 larvae, as suggested by GSEA, is consistent with this role. *Hsp*-90 in particular has been identified as a parasitism-central gene based on changes in *S. ratti* gene expression during high immune pressure [22] and is similarly abundantly expressed by *S. stercoralis* L3i larvae.

Sperm containing glycoprotein (SCP) domain exclusively found in L3i larvae

The SCP domain, found exclusively in L3i biased genes, is a conserved domain of unknown function present in a wide range of organisms [24]. Interestingly, it has been found to be present in activation-associated secreted proteins that have been studied as potential vaccine targets in other nematodes [24,25]. Whether overrepresentation of the SCP domain in the L3i group is related to the presence of these secreted proteins is unclear, but activation-associated secreted proteins have been found to be important in many parasitic nematodes in which they have been studied to date.

*C. elegans* dauer and *S. stercoralis* L3i larvae have distinct characteristics

Consistent with previous findings, a striking L3i-*C. elegans* ‘dauer expression signature’ was not uncovered in this comparative microarray analysis [6]. We instead identified genes that are regulated in apparently opposite manners by *C. elegans* dauer and *S. stercoralis* L3i larvae which offer useful clues about the biology of *S. stercoralis* parasitism. L3i biased expression of the putative amy-2 gene (encoding the myosin heavy chain) is consistent with the highly motile nature of L3i larvae which, unlike their dauer counterparts, seek out and initiate infection in a host. Although dauer and L3i larvae both contain a cuticle that enables survival in the environment, the parasitic cuticle has been associated with the ability of infective stages to evade the immune response of the host, and its structure varies from one species to another [26]. Biased expression of genes putatively encoding particular collagens (col-37, col-119) in the L3i but not the *C. elegans* dauer, points to differences in the composition of the parasitic cuticle that could potentially have a role in this regard. In fact, a recent microarray based analysis of the response of the *S. ratti* transcriptome to host immunologic environment notes upregulation of collagen genes by *S. ratti* which is believed to play a protective role for the parasite [27]. *C. elegans* dauer and *S. stercoralis* L3i larvae can survive in the environment even in the absence of a steady source of food. One way by which this occurs is by the development of electron-dense intestinal granules that store non-lipid products [11]. The gene *lmp-1* plays an essential role in this regard for dauer larvae as suggested by RNA interference studies [28]. It is likely that L3i larvae similarly utilize these granules while in the environment. The presence of these granules may additionally explain the darkened color of the radially constricted intestines of L3i larvae, an appearance shared by its dauer counterpart.

A key feature shared by dauer and L3i larvae is the ability to extend the lifespan while in the free-living state. In both *C. elegans* and *S. stercoralis*, the forkhead transcription factor DAF-16 plays a role in regulating dauer diapause, longevity and metabolism [11,29,30]. A downstream target of DAF-16, *egl-10*, is known to be negatively regulated by DAF-16 in *C. elegans* [29]. By contrast, this gene was found to have biased L3i larval expression in *S. stercoralis*. Such discordance is consistent with findings from a prior study that failed to detect a transcriptional profile typical of down-regulated...
insulin-like signaling in long-lived parasitic females of S. ratti [31]. Although the downstream targets of insulin-like signaling have not been fully elucidated in Strungioides species, the apparent upregulation of S-egl-16 in the L3i potentially highlights adaptations at a molecular level that likely underlie the evolution to parasitism. Such adaptations could include alterations in genes controlling metabolic and developmental functions, adaptations of pre-existing genes to encode new functions, and gene duplication and diversification [32]. The apparent lack of a C. elegans dauer-like transcriptional profile in S. stercoralis L3i is also consistent with published findings on the apparent lack of a functions, and gene duplication and diversification [32]. The alterations in genes controlling metabolic and developmental alterations in genes controlling metabolic and developmental adaptations of pre-existing genes to encode new functions, and gene duplication and diversification [32]. The apparent lack of a C. elegans dauer-like transcriptional profile in S. stercoralis L3i is also consistent with published findings on the apparent lack of a functions, and gene duplication and diversification [32].

The true value in identifying these and other genetic determinants of S. stercoralis parasitism lies in whether the products of these genes can induce protective immunity. Indeed, one of the genes identified in our list, the S. stercoralis ortholog of eat-6 Na+H+ATPase, has already been identified as a potential vaccine candidate based on animal experiments [39].

Additional therapeutic targets and immunodiagnostic genes of significance

Contig 1872, a gene with L3i biased expression, encodes an ortholog of C. elegans core subunit of the cytochrome bc1 complex, UCR 2.1 (E-value = 1E-014). This subunit has been shown to be a potential target for antiparasitic drugs based on the finding that in C. elegans, UCR 2.1 is essential for viability and is less related to mammalian UCR-1 than to mitochondrial processing peptidases from other organisms [40]. S. stercoralis transgenesis experiments may prove useful in investigating the question of whether this gene is similarly essential for S. stercoralis larval survival.

In our microarray analysis of S. stercoralis, we found abundant L3i expression of the S. stercoralis ortholog of hsp-90, contig 77 (3 fold expression difference). Interestingly, the hsp-90 inhibitor geldanamycin has been shown to have a macrofilaricidal effect on filarial nematode Brugia pahangi [42]. Hsp-90 has been identified among S. ratti parasitism central genes critical for survival and further studies investigating it as a chemotherapeutic target are warranted.

Contig 1151, which was among the 25 most highly biased L3i genes (11-fold expression difference), corresponds to fatty acid and retinol binding protein-1 (FAR-1; E-value = 1E-016). FAR-like proteins are major secreted products of parasitic nematodes that allow the parasite to scavenge essential nutrients from its host [43]. Depletion of host lipids is thought to be necessary for parasite survival and may additionally impair the host immune response [44]. These proteins have additionally demonstrated stage and gender specificity in other nematodes, most notably in the hookworm Ancylostoma ceylanicum [45]. The immunodiagnostic potential of FAR-like proteins has been assessed in other nematodes, such as Onchocerca volvulus, in a serologic assay based on Ov-20 (FAR-1) [45,46,47]. FAR-1 proteins have been successfully used in a vaccine in animals infected with Ancylostum [43]. These microarray data identify S. stercoralis far-1 as an L3i-biased target that may be a potential vaccine candidate or immunodiagnostic antigen.

Limitations

Approximately one-third of S. stercoralis genes are of unknown function. This finding is consistent with a previous EST analysis that revealed a similar percentage (25%) of S. stercoralis clusters with no significant BLAST alignments [8]. This finding is also consistent with functional genomics analyses of the C. elegans and human genomes where significant numbers of genes of unknown function were identified [48,49]. Some of these unknown sequences may derive from untranslated mRNA regions, which are common in polyA-primed libraries [50]. The complete genome sequence of S. stercoralis is not available to date. Inferred functional annotations of an analogous nematode C. elegans, while useful, may not be directly applicable to S. stercoralis, as suggested by interspecies differences uncovered in the present comparative microarray analysis. Because a number of C. elegans genes did not have S. stercoralis orthologs that were also differentially expressed according to our predefined ‘cutoffs,’ it was difficult to formulate gene lists organized into functional categories with at least 5 contigs. This limited our ability to analyze biochemical or metabolic pathways of potential importance. As our knowledge of the S. stercoralis genome increases, these microarray analyses will likely gain in usefulness and a more direct approach using annotation based on known S. stercoralis gene functions would be even more informative.

Conclusions

DNA microarrays allow for simultaneous analysis of large numbers of genes from two or more biologic conditions. This powerful method of analysis has revolutionized our understanding of the immunopathogenesis of schistosomiasis [51], for example, and has advanced the development of vaccine discovery and therapeutics in parasitology [52,53]. Until now, studies of S. stercoralis have been limited to the analysis of ESTs rather than the full genome sequence. Development of a novel DNA microarray tool for the study of S. stercoralis represents an exciting step forward in our understanding of this parasite.

Supporting Information

Text S1 This file contains supplemental information regarding microarray probe information (Table S1), primer probe sequences used in real-time PCR analysis (Table S2), all contigs (Table S3), L1 biased contigs (Table S4), L3i biased contigs (Table S5), most highly expressed L1 and L3i contigs (Table S6), L3i biased contigs containing sperm containing glycoprotein domain (Table S7), and results of the GSEA for immunoreactive genes (Table S8) and heat shock proteins (Table S9). For the column marked “Manual Annotation,” the following abbreviations were used: em = energy metabolism; exmat = extracellular matrix; cs = cytoskeleton; imm = genes encoding antigens known to be immunoreactive in sera from patients infected with S. stercoralis; met = metabolism; nr = nuclear regulation; pep = protein export machinery; pm = protein modification; prot = proteasome machinery; ps = protein synthesis; st = signal transduction; tf = transcription factor;
Acknowledgments

We thank Guojian Jiang for his expert assistance with microarray hybridizations and NIAID intramural editor Brenda Rae Marshall for assistance with the manuscript preparation.

References