An oligonucleotide microarray for transcriptome analysis of *Schistosoma mansoni* and its application/use to investigate gender-associated gene expression

Jennifer M. Fitzpatrick a, David A. Johnston b, Gary W. Williams c, Debbie J. Williams c, Tom C. Freeman c, David W. Dunne a, Karl F. Hoffmann a, *

a Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK
b Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK
c Rosalind Franklin Centre for Genomics Research, Hinxton, Cambridge CB10 1SB, UK

Received 14 September 2004; received in revised form 12 December 2004; accepted 4 January 2005

Abstract

Global profiling transcriptomes of parasitic helminths offers the potential to simultaneously identify co-ordinately expressed genes, novel genetic programs and uniquely utilized metabolic pathways, which together provide an extensive and new resource for vaccine and drug discovery. We have exploited this post-genomic approach to fabricate the first oligonucleotide DNA microarray for gene expression analysis of the parasitic trematode *Schistosoma mansoni*. A total of 17,329 *S. mansoni* DNA sequences were used to design a microarray consisting of 7335 parasite elements or approximately 50% of this parasite’s transcriptome. Here, we describe the design of this new microarray resource and its evaluation by extending studies into gender-associated gene expression in adult schistosomes. We demonstrate a high degree of reproducibility in detecting transcriptional differences among biologically replicated experiments and the ability of the microarray to distinguish between the expression of closely related gene family members. Importantly, for issues related to sexual dimorphism, labour division, gamete production and drug target discovery, 197 transcripts demonstrated a gender-biased pattern of gene expression in the adult schistosome, greatly extending the number of sex-associated genes. These data demonstrate the power of this new resource to facilitate a greater understanding into the biological complexities of schistosome development and maturation useful for identifying novel intervention strategies.

Keywords: *Schistosoma*; Helminth; DNA microarray; Gene expression

1. Introduction

Parasite expressed sequence tag (EST) and genomic sequencing projects have proven to be an invaluable resource for parasite gene discovery and have led to the identification of numerous putative gene products [1–3]. Despite the large availability of DNA sequence information, true potential of this resource will only be realized upon the assignment of gene function within an actual biological and cellular context, thus leading to the possible functional annotation of many important parasitic genomes. Towards this goal, several investigators have developed DNA microarrays to probe and begin to elucidate the role of specific gene products in the lifestyle, pathogenicity and fundamental biology of multiple parasites [4–6]. This approach, in combination with continued wide ranging genomic and EST sequencing has brought together genomic and functional-genomic data

© 2005 Published by Elsevier B.V.

Keywords: *Schistosoma*; Helminth; DNA microarray; Gene expression
to reveal new insights into the complex pathogenic parasitic lifestyle.

Gene discovery and functional analysis of schistosomes has also been facilitated by a variety of approaches [6,7]. These studies have revealed cDNA microarrays to be particularly useful and reliable in identifying gene-associated transcripts in both *Schistosoma mansoni* [5] and *S. japonicum* [7]. However, while small sample size subsets proved fruitful for the initiation of these investigations, utilizing a high-throughput approach on a whole genome scale will dramatically increase the transcriptional understanding of parasite sexual biology where current thinking proposes that males and females evolved to maximize independent functional roles important to the survival of the parasite. Presently, the *S. mansoni* genome contains approximately 14,000 predicted genes, and like many other parasites, most of these genes display no database homology and therefore, have no functional annotation [8–10]. Assigning some putative function or association, based on expression profiling by DNA microarray analysis, may provide some of the most promising research areas for elucidating the molecular basis of parasite biology.

Here we describe the design, fabrication and validation of a new DNA microarray for schistosome transcriptome analysis based on the use of long oligonucleotide probes. Long oligonucleotide DNA microarrays were constructed as highly sensitive alternatives to cDNA microarrays due to inherent advantages throughout production and experimental use [11]. The DNA microarray contains 7335 *S. mansoni* probes covering approximately 50% of the total estimated gene complement and was used here to expand upon our previous studies of schistosome conjugal biology. The experiments described provide a dual function: (1) to specifically characterize the fabricated oligonucleotide DNA microarray allowing sensitive, reproducible gene expression results to be generated by multiple users and laboratories; and (2) to further elucidate the expression profiles of male and female adult parasites with the goal of expanding our knowledge relating to sexual maturation, sexual dimorphism, labor division and gamete production. Ultimately, investigations into *S. mansoni* transcriptional mechanisms will likely generate new insights into the development and maintenance of this helminth’s dioecious lifestyle, leading to the identification of novel drug targets or vaccine candidates.

2. Materials and methods

2.1. Parasites

Adult male and female *S. mansoni* (NMRI Puerto Rican strain) were perfused from percutaneously infected mice at 7 weeks after challenge with independent batches of 250 cercariae each shed from albino *Biomphalaria glabrata*. After perfusion, both immature and mature worms were counted and sex-separated. Miracidia used to infect *B. glabrata* were hatched from eggs collected from mouse livers 7 weeks after infection [12].

2.2. Design of *S. mansoni* DNA oligonucleotide probes

*S. mansoni* DNA elements chosen for oligonucleotide design were selected from EST sequences available from public databases as of June 28, 2002, full-length mRNA and genomic DNA (gDNA) sequences (using NCBI Entrez limits excluding ESTs, STGs, GSSs, TPs, patents and working drafts) available from public databases as of April 2003 and from one full-length mRNA sequence identified in our laboratory (AY267032—*S. mansoni* arginase). The 16,815 EST sequences were clustered using the CAP3 DNA sequence assembly program [13] into 2076 contigs (representing more than one EST sequence) and 5049 singletons (representing only one EST sequence) for a total of 7125 unique DNA sequence clusters. In addition to these 7125 non-redundant EST clusters, 513 full-length mRNA (some redundancy with respect to EST clusters) and gDNA sequences were included to bring the total number of DNA sequences used as templates for oligonucleotide design to 7638. CAP3 has previously been shown to be tolerant of sequencing errors resulting from single pass sequencing and is effective at differentiating between closely related gene family members [14].

Putative sequence homology of each schistosome DNA element was assigned using the web-based Basic Local Alignment Search Tool (BLASTx) [15] searching against the NCBI protein non-redundant (nr) database. BLASTx hits with an Expect-value (E-value) of ≤10−05 were considered significant and the corresponding NCBI protein nr designation was used to annotate the EST contigs and singletons. BLASTx hits with an E-value of ≥10−10 were not considered significant and therefore the corresponding schistosome DNA elements obtaining these scores were annotated as 'UNKNOWN'. In a further attempt to annotate these unknown schistosome DNA sequences, they were compared against the *S. mansoni* EST database compiled by Verjovski-Almeida et al. [9] using BLASTn, where DNA elements generating BLASTn bit score values of ≥200 were considered a match. Subsequent annotation was provided then by the EST database assigned through additional BLASTx searches of NCBI (using their applied criteria [9]). Sequence similarity assignment was performed for two reasons: (1) to annotate the DNA sequence representations deposited on each DNA microarray; and (2) to identify which DNA strand to select for oligonucleotide design. For BLASTx searches with no significant similarity, the sense strand was selected for oligonucleotide design, unless there was a poly-T tract incorporated at one end of the parent DNA sequence (anti-sense strand was used in this case). Therefore, a small percentage of oligonucleotides deposited on this DNA microarray may have represented the non-coding DNA strand. The labelling procedure used in this study [16] (Klenow incorporo-
All oligonucleotides were 50 bases in length, modified by a 6-C linked 5’ amino modification and synthesized at 200 nmol scale (Illinova, SD, CA). Multiple hierarchical tests were performed on each of the 7638 DNA sequences to select the optimum 50-mer oligonucleotide representative for each contig/singleton/mRNA/gene. The European Molecular Biology Open Software Suite (EMBOSS) [17] hosted at SourceForge (http://emboss.sourceforge.net/) was used for most of the analyses. Initially, each sequence was required to be a minimum of 50 bases in length. Those sequences passing this first test were then subjected to a series of selection criteria. Using the EMBOSS program ‘dan’, regions were identified that were within the required GC content of 30–50%. The program ‘palindrome’ was used to identify regions with no gross secondary structure (defined as having seven or more bases capable of forming a perfect hairpin structure). The program ‘fake’ was used to identify regions that had no ambiguous nucleotide codes. Additionally, using the program ‘RepeatMasker’ (http://www.repeatmasker.org), regions were identified that contained no low complexity sections, simple repeats or sequences matching the S. mansoni SR2 retrotransposon sequence [18]. Using BLASTs, each potential oligonucleotide source region was then compared against all 7638 S. mansoni sequences and areas with no matches of 12 or more bases were identified. If this BLAST criterion proved too restrictive to allow regions in a sequence to be used as oligonucleotides, then it was progressively relaxed to ≥20 and then ≥40 bases, respectively. Starting from the 5’ end of each sequence, a maximum of four non-overlapping regions of sequence with a length of 50 bases or more that passed the selections criteria were identified. The 5’ end of these possible oligonucleotides was determined using the EMBOSS program ‘dan’, and the one from each sequence with a Tm closest to 72 °C was selected for oligonucleotide design. Originating from the initial 7638 DNA sequences, 7214 passed all of these tests and were used to design 50-mer oligonucleotides. The remaining 424 DNA sequences that failed were again subjected to the same tests, but with the % GC content relaxed to 25–50%. Out of these 424 DNA sequences, 121 could generate a 50-mer oligonucleotide. The remaining 303 sequences failed these tests due to length (<50 bases), high ambiguity or a GC content outside of the required ranges. All sequences were analyzed using the InterProScan package [19] and Gene Ontology (GO) [20] terms were extracted from the output of this analysis using a perl script. A total of 7335 50-mer S. mansoni oligonucleotides were designed and, along with control oligonucleotides (Bacillus subtilis-specific and Arabidopsis thaliana-specific DNA elements), represented the oligonucleotide DNA microarray probes used in this study (Table 1). Oligonucleotide DNA microarrays were printed on CodeLink Activated Slides (Amine-Binding Slides) for Genomics Research (Hinxton, UK) at a concentration of 250 ng μl⁻¹.

Table 1

<table>
<thead>
<tr>
<th>Description</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clustering</td>
<td>16,815</td>
</tr>
<tr>
<td>Contigs</td>
<td>2076</td>
</tr>
<tr>
<td>Singletons</td>
<td>5040</td>
</tr>
<tr>
<td>Genes</td>
<td>513</td>
</tr>
<tr>
<td>Sequences</td>
<td>7638</td>
</tr>
<tr>
<td>Microarrays</td>
<td>8160</td>
</tr>
<tr>
<td>Oligonucleotides</td>
<td>7335</td>
</tr>
<tr>
<td>Control Sequences</td>
<td>84</td>
</tr>
<tr>
<td>Control Sequences (or full-length mRNAs)</td>
<td>120</td>
</tr>
<tr>
<td>Sequences displaying significant similarity</td>
<td>4621</td>
</tr>
<tr>
<td>Sequences displaying significant similarity</td>
<td>3605</td>
</tr>
<tr>
<td>Sequences</td>
<td>2714</td>
</tr>
<tr>
<td>Sequences displaying significant similarity to hypothetical proteins</td>
<td>1156</td>
</tr>
<tr>
<td>Sequences submitted for BLAST analysis</td>
<td>3605</td>
</tr>
<tr>
<td>Sequences</td>
<td>2591</td>
</tr>
<tr>
<td>Sequences</td>
<td>2030</td>
</tr>
<tr>
<td>Sequences displaying significant similarity to hypothetical proteins</td>
<td>1156</td>
</tr>
<tr>
<td>Genes</td>
<td>161</td>
</tr>
<tr>
<td>Oligonucleotide microarray sequence information</td>
<td></td>
</tr>
</tbody>
</table>
3. Results

3.1. *S. mansoni* sequence analysis and fabrication of a novel oligonucleotide microarray to profile schistosome gene expression

Clustering of the parasite EST sequences contained in *NCBI* protein nr databases as of June 2002 (using the CAP3 program) and inclusion of full-length genes as of April 2003, led to a collection of 7638 sequences available for oligonucleotide design. Selective criteria for design of optimal oligonucleotides enabled 7335 *S. mansoni* oligonucleotides (duplicated on microarray to 14,670) to be synthesized and arrayed (Table 1). BLASTx manual annotation of each contig/EST revealed 4621 sequences displaying no significant similarity using BLASTx p-value (E) criterion of $\leq 10^{-50}$, representing $\sim 63\%$ of the total *S. mansoni* sequences. Therefore, from the initial BLASTx annotation, 2714 sequences displayed significant similarity to known genes in NCBI ($\sim 37\%$ of total *S. mansoni* sequences). The 4621 sequences without similarity were then submitted to BLASTn-analysis at the University of Sao Paulo/ONCA. *S. mansoni* EST database [9,24] and hits with a bit score value of $\geq 200$ were merged with the sequences to create longer contigs. These new contigs were then re-submitted to the *NCBI* nr protein database using BLASTx in an attempt to annotate them. This left 2591 sequences still with no significant similarity to entries deposited in NCBI (using same BLASTx criteria as previously described). In contrast, 2030 sequences gained some annotation through this second database search, 1156 of which having significant similarity to hypothetical proteins (874 of these containing some functional annotation). Therefore, the total number of *S. mansoni* sequences used for oligonucleotide design displaying some annotation is approximately 65% (49% without hypothetical proteins). Additionally, 1140 sequences were potentially presented in the databases in the anti-sense direction and therefore the final oligonucleotide was designed from the sense strand (see supplementary information). Therefore, from the initial BLASTx annotation, 2591 sequences still with no significant similarity using BLASTx

3.2. Adult gender-associated transcripts are reproducibly detected by fabricated *S. mansoni* oligonucleotide microarrays

As multi-gene families are represented throughout the *S. mansoni* genome (cathepsins, superoxide dismutase, dynein...
As quality control experiments illustrated the reproducible nature of this S. mansoni DNA oligonucleotide microarray, analyzing the differences in gene expression between adult female and adult male parasites initially validated this resource. Here, all transcripts identified as differentially expressed displayed Log2 normalized gene expression ratios outside of the 90% confidence interval in three out of five independent, biologically replicated experiments. A hierarchical clusterogram (single linkage analysis/Euclidian distance correlation) shown in Fig. 2 displays all transcripts identified as being highly differentially expressed in either male (red: positive ratio) or female (green: negative ratio) parasites. These male/female bimodal comparisons revealed 141 unique genes highly expressed in the female when compared to the male, and 86 genes highly expressed in the male when compared to the female. This considerably increases the number of known sex-associated transcripts in S. mansoni adult parasites. Importantly, dye-swap experiments showed a high correlation with data passing these strict filtering criteria and therefore suggest minimal dye effect on the prediction of differentially expressed transcripts. Twenty-four oligonucleotides (16 classified as having no significantly sequence similarity) highly expressed in female (green: negative ratio) parasites. These male/female bimodal comparisons revealed 141 unique genes highly expressed in the female when compared to the male, and 86 genes highly expressed in the male when compared to the female. This considerably increases the number of known sex-associated transcripts in S. mansoni adult parasites. Importantly, dye-swap experiments showed a high correlation with data passing these strict filtering criteria and therefore suggest minimal dye effect on the prediction of differentially expressed transcripts. Twenty-four oligonucleotides (16 classified as having no significantly sequence similarity) highly expressed in female (green: negative ratio) parasites. These male/female bimodal comparisons revealed 141 unique genes highly expressed in the female when compared to the male, and 86 genes highly expressed in the male when compared to the female. This considerably increases the number of known sex-associated transcripts in S. mansoni adult parasites.
Fig. 1. *S. mansoni* oligonucleotide DNA microarrays are capable of generating reproducible expression results and can distinguish transcriptional differences among gene family members. (A) Hybridization performance of multiple oligonucleotides spanning different intragenic regions of two superoxide dismutase homologues. Five oligonucleotides spanning the three extracellular superoxide dismutase exons (Ex-SOD, M27529) reveal similar gene expression profiles (for all five biological replicates averaged) and strongly support the female-enriched expression of this SOD homolog (verified by gene-specific RT-PCR). (B) Cytosolic SOD (Cyt-SOD, M97298) did not display a statistically significant gender association as determined from hybridization information originating from four oligonucleotides spanning two exons and the 3′-UTR. No signal intensity (NS) was observed for a Cyt-SOD specific oligonucleotide designed over an intron. Boxes represent exons, solid lines represent non-coding DNA elements and dashed lines represent positions of designed oligonucleotides for the 5.1 kb Ex-SOD and the 8.5 kb Cyt-SOD genes. RT-PCR conditions and derivation of expression ratio (Ex. ratio) are described in Section 2. (C) Statistical analysis of biologically reproduced DNA microarray hybridization experiments. Scatterplots compare the Log$_2$-calibrated ratio generated for each oligonucleotide probe (mean of two replicate spots) from one biological batch of adult worm material to the Log$_2$-calibrated ratio generated for the same oligonucleotide probe (mean of two replicate spots) from a different biological batch. The correlation coefficient values from two representative comparisons, $R = 0.848$ and 0.816, indicate a high degree of agreement between biologically replicated experiments. Scatterplots display oligonucleotide probes that passed the initial filtering criteria (signal intensity greater than 1 S.D. above the mean of negative control elements) for quality data in each biological batch/hybridization comparison and include 4524 for experiments 3 and 5, and 4645 elements for experiments 1 and 5. Lines represent the line of regression (centre line) and the predicted 99% confidence intervals of the plotted data.
Fig. 2. *S. mansoni* oligonucleotide microarrays can reproducibly detect gender-enriched gene expression profiles in sexually mature *S. mansoni* male and female adult parasites. Single-linkage hierarchical clustering (using Euclidean distance correlation matrices) of all gender-enriched transcripts passing filtering criteria from five independent experimental replicates and dye-swap experiments. Representative gender-enriched transcripts verified by RT-PCR (Fig. 3) are listed next to the clustergram along with their unique identifier (contig ID, accession number or name). All other gender-enriched transcripts identified in this study can be found in accompanying supplementary information. Increasing shades of red depict greater expression in adult male parasites, increasing shades of green represent greater expression in adult female parasites and grey represents gene expression data absent for that particular replicate experiment.

These transcripts include those previously identified in our laboratory [5,7] as well as female eggshell proteins (p48/chorion/34 kDa) [28–33], female-specific 800 protein [34], ferritin-1 [35], ORF-Rf2 [36], glutathione peroxidase [37] and adenylosuccinate lyase [38] amongst others. These positive controls, confirming predicted patterns of gene expression, again reinforced the usefulness of this oligonucleotide microarray to reveal new and interesting gender associations. A large number of novel associations were also observed, revealing both adult female and adult male expression biases. These associations included the female-enriched expression of an extracellular superoxide dismutase and tyrosinase 1 and 2, although all of these have previously been reported as female-associated in *S. japonicum* [7], and *S. mansoni* for tyrosinase [5]. Further novel female-associated transcripts included an acyl-CoA diacylglycerol acyltransferase (DGAT), acyl-CoA cholesterol acyltransferase (ACAT), UDP-GlcNac:Alpha-6-D-Mannoside Beta-1,2-N-Acetylgalactosaminyltransferase II, multiple histidine-rich proteins, serine-threonine kinases, a large amino acid transporter (also previously observed in *S. japonicum*), anti-inflammatory protein-16, stathmin-like protein and purine-nucleoside phosphorylase. Furthermore, a significant proportion of all oligonucleotides identified as differentially expressed in the female showed no significant identity with sequences contained in the NCBI protein nr databases. Analysis revealed that this list of female-associated transcripts contained 39 such sequences and an additional 25 sequences annotated as putative hypothetical proteins.
brought a variety of new information available in this area of schistosome biology. Muscle, components of the tegument and cytoskeletal elements dominate this set of male-associated transcripts, and include myosin (multiple oligonucleotides representing multiple subunit chains), paramyosin, tropomyosin, alpha-actinin, annexin, fimbrin, microtubule-associated protein 1B, echinocandin, Sm30, tropomin I and Sm20. Other interesting male-associated transcripts include those inferred to be a rydodine receptor, procollagen-lysine 2-oxoglutarate 5-dioxygenase 1 precursor (lysyl hydroxylase 1), two collagens, an extracellular superoxide dismutase (contig1384; a different isoform to that detected as female-expressed) and the high voltage-activated calcium channel beta subunit Cva11.1, believed to be partially responsible for praziquantel sensitivity [39,40] (although this transcript did not repeat in the dye-swap experiment, subsequent RT-PCR analysis (Fig. 3B) suggests a definite male bias for its expression). Additionally, 16 sequences with no significant similarity to genes in NCBI and 24 hypothetical proteins were also identified.

See supplementary information for full list of differentially expressed transcripts for both female and male adult parasites.

3.4. Verification of gender-biased transcripts by RT-PCR

Confirmation of microarray expression data by an independent method of analysis demonstrated the precision of the transcriptional profiling facilitated by this new oligonucleotide DNA microarray resource. Gene-specific primers were used in an RT-PCR reaction to confirm the expression profiles detected by microarray analysis. RT-PCR results showed the exact same pattern of predicted gene expression revealed by microarray hybridization. Sixteen randomly selected female-associated (by DNA microarray analysis) transcripts were tested and shown to be strongly female expressed by RT-PCR in comparison to the male using the a-tubulin gene as an internal standard and sample control [27] (Fig. 3A). In addition, the 16 male transcripts chosen for confirmation were also identified as more heavily expressed in the male when compared to the adult female by RT-PCR analysis (Fig. 3B). One microarray oligonucleotide representing glutathione peroxidase (SCMGPX1A; L37762) was designed over an intronic region of the gene (and therefore displayed no expression values). As predicted by previous studies [37] and shown here by RT-PCR, this transcript was expressed no expression values). As predicted by previous studies [37] and shown here by RT-PCR, this transcript was strongly female expressed (although this transcript did not repeat in the dye-swap experiment) and showed the same pattern of expression as in the DNA microarray analysis, thus demonstrating the usefulness of both the microarray resource and the filtering criteria applied. Transcripts denoted with * did not reproduce male-associated expression in dye-swap experiments (did not pass strict filtering criteria); however, were revealed as highly expressed in the male when the original dye combination was used.

4. Discussion

DNA microarray analysis is now a well-established functional genomics tool for the global analysis of gene expression [41,42]. The oligonucleotide microarray fabricated here represents ~50% of the estimated S. mansoni gene complement [9]. The sequences utilized for oligonucleotide design are proportionally representative of the entire S. mansoni genome, since the relative percentages of sequences displaying significant similarity to known genes in the databases is similar to that estimated by large-scale EST sequencing efforts [9]. Furthermore, the distribution of sequences between distinct biological functions and processes reflects the transcriptome of this parasite as a whole and suggests diverse roles in numerous biological functions for independent sequences. Since the sequences examined here represent a range of important processes and originate from diverse developmental stages, this established DNA microarray has major implications for probing different biological questions.

Although DNA microarray technology is relatively widely used, this is the first time long DNA oligonucleotides have been employed for large-scale profiling of gene expression in S. mansoni. Consequently, in order to characterize, optimize and assess this post-genomics tool, a set of experiments profiling the transcriptome of mature adult schistosomes was performed, similar to previous studies [5,7]. While sexually mature, adult male and female parasites transcribe the majority of genes to a similar degree, it is anticipated that numerous genes will be differentially expressed, as the sexes are morphologically, functionally and chromosomally distinct. Having evolved from hermaphroditic ancestors, it is likely that differential gene expression has driven sexual dimorphism and labour division within the genus and ultimately led to cooperative conjugal biology as a highly successful means of maximizing parasite transmission. Although highly effective for continuing the lifecycle, intravascular conjugal biology resulting in egg production by adult female parasites also contributes to host-mediated, inflammatory, circumoval immune responses. If uncontrolled, these immune responses can precipitate a series of pathological complications resulting in severe morbidity and mortality in infected individuals [43]. Since the transcriptional basis of adult sexual maturation and egg-production remains relatively unknown, any information gained will be crucial to the elucidation of the specific processes involved. To be classified here as differentially expressed between the sexes, genes had to pass independent statistical criteria: (1) expression values had to be significantly above background levels in order to remain in the dataset (as calculated from negative control elements); and (2) genes had to be outside of the 90% confidence interval of the entire range of expressed data in three of five independent biologically replicated hybridizations and passed the identical criteria in multiple dye-swap experiments using pooled RNA samples from independent batches. These criteria enabled a total of 117 female and 80 male genes to be reproducibly and confidently identified as differentially expressed.
Fig. 3. RT-PCR analysis confirmed gender-enriched DNA microarray data for all transcripts tested, illustrating a high correlation between global gene expression profiling and individual gene transcript abundance. One microgram of pooled (five experimental replicates) S. mansoni total RNA (DNase I treated (Ambion Inc.)) was used in a RT reaction to prime cDNA synthesis as described in Section 2. Unique identifier (Unique ID) and BLASTx NCBI nr protein database annotation of amplified transcript, sequence of each PCR primer pair, cycle number, expected product size and annealing temperatures for each verified cDNA are listed in Table 2, supplementary information. (A) RT-PCR confirmation of randomly selected female-enriched transcripts identified by DNA microarray analysis. (B) RT-PCR confirmation of randomly selected male-enriched transcripts identified by DNA microarray analysis. Amplification of constitutive α-tubulin served as an internal control for all PCR reactions. (*) indicates female-enriched expression of a glutathione peroxidase ortholog determined solely by RT-PCR as oligonucleotides for SCMGPX1 deposited on DNA microarray were designed over intronic regions. (‡) denotes two transcripts that did not pass filtering criteria for differential expression in the dye-swap experiments. BLASTx E-values for amplified gene products are given to indicate sequence similarity.
pressed, revealing a massive amount of new molecular data.

Many positive controls confirmed the validity of these results, including the female-associated eggshell proteins, as previously demonstrated in *S. mansoni* [5] and *S. japonicum* cDNA [7] microarray experiments.

Investigating the putative function of many of these gender-associated transcripts has revealed a number of interesting and previously unexplored patterns of expression. It is now clear that the vast majority of male-associated transcripts are intimately involved in the structural organization of the parasite (supplementary information) which contrasts directly with those transcripts highly expressed in the adult female, some having a presumed role in reproductive and egg-laying mechanisms. Structural elements expressed differentially by the male include components of the muscular system, tegumental proteins and the underlying cytoskeleton. This apparent ‘division of labour’ between the genders [44-45] now has an extensive transcriptional basis. The male ensures the survival of the egg-laying female by providing physical support and musculature to aid feeding [46], physiological transportation within the vasculature [47] and potential extracellular maturation factors [48,49]. The female is thus allowed to concentrate energy expenditure on egg-production, as reflected by the large number of transcripts involved in reproduction and development.

Differentially expressed transcripts of special interest were revealed in both genders (a number of which were subsequently confirmed by RT-PCR analysis, Fig. 3). Evidence exists to the importance of regulated intracellular calcium levels to parasitism by *S. mansoni*; *CaV* signalling is absolutely required for muscular contraction (important more so in the male given their extensive muscular scaffold and infrastructure). Multiple voltage-gated Ca²⁺ channels have been identified within schistosomes, including both α- and β-subunits of *S. mansoni* (and *S. japonicum*) [50,51]. Here, microarray analysis showed the expression of SmCaVα₁ to be significantly greater in the male than the female (in direct contrast to SmCaVα₂, which appears to be similarly expressed between the two genders). Although in one dye combination, SmCaVα₁ did not demonstrate a significantly strong male association, RT-PCR analysis of this transcript (Fig. 3) confirmed the original microarray results and suggested that dye-swap experiments can be influenced at multiple levels and should be interpreted with caution. Praziqanthel (PZQ), the current drug of choice for the treatment of schistosomiasis, is believed to instigate disruption of Ca²⁺ homeostasis within the parasite, although the exact mechanism is unknown [52-53]. It has become clear in recent years that female parasites are much less sensitive to the action of PZQ than males [54], the SmCaVα₁ subunit being intimately involved in the conference of sensitivity [39]. Moreover, these DNA microarray results also revealed a male-associated expression of desmoyokin (AHNAC protein [55]). Although AHNAC may possess multiple functions within a cell, it has been shown to interact specifically with the β-subunits (1, 2 and 2a) of the L-type Ca²⁺ channel and with F-actin of human heart muscle, mediating directly Ca²⁺ signal transduction [56]. It is possible then that high levels of SmCaVα₁ and AHNAC in the male may be instrumental in their increased sensitivity to PZQ observed when compared to females. Numerous other important Ca²⁺-binding proteins (CaBP) were additionally identified as being male-associated, including calpain (calcium-activated neutral protease (CANP)) [57], calmodulin [58], Sm20 [59] and a ryanodine receptor [60].

From the list of differentially expressed transcripts, females express a higher proportion of enzymes than do males. Included here as specifically female-associated is the expression of two endoplasmic reticulum neutral lipid synthesiss enzymes acyl-CoA:diacylglycerol acyltransf erase 2-like (DGAT2-like) and acyl-CoA cholesterol-acyltransferase-1 (ACAT1). ACAT covalently joins cholesterol and fatty acyl-CoA molecules to form cholesterol esters [61-63] and DGAT catalyzes a similar reaction to generate triglycerides, using di-acylglycerol as the acyl group acceptor [64,65]. Subsequent sequence analysis and database queries identified a further two potential DGAT1 and DGAT2-like molecules in *S. mansoni*. RT-PCR analysis (in comparison to α-tubulin) revealed a female-associated expression bias for each of these four genes (data not shown). *S. mansoni* is known to be unable to synthesize fatty acids and sterols de novo [66]. The specific role of these enzymes in the female at this time is still unclear, although the generation of triglyceride stores within the parasite is likely to represent a major function. Since schistosomes do not generate energy/ATP through the β-oxidation of fatty acids [67] (despite much of the enzymatic capacity for this being encoded within the genome, data not shown), the parasites may use neutral lipids as stores to guard against high intracellular (toxic) levels of free fatty acids and cholesterol. In addition, the adult female may utilize neutral lipids during egg production.

Concomitant with the egg-laying process, the female consumes a large excess of red blood cells in comparison to the male [68]. The female therefore requires the by-products of haemoglobin digestion to be metabolized, eliminated and stored to reduce direct toxicity and lipid peroxidation. The high expression of the enzymes superoxide dismutase and aspartate aminotransferase, together with the previously known female-associated cathepsins and aspartic proteases [69] and ferritin-1 (stores Fe³⁺ in non-toxic form) [35] in the female, as shown by both DNA microarray analysis and RT-PCR, suggests these enzymes may share a functional role in this process.

The DNA microarray analysis described in this study and combined with previous studies [5,7] revealed a far greater number of transcripts displaying female-associated expression than male. This may be directly indicative of the enhanced metabolic requirements of extensive egg-production, but in addition, may also reflect the presence of actual egg transcripts within the female sample pool. Current investigations utilizing this oligonucleotide microarray to longitudinally analyze the development of the parasite within the host will identify such transcripts and enable a more detailed...
This study suggests that the design and use of an oligonucleotide microarray to profile the S. mansoni transcriptome, is both a feasible and an efficient strategy to examine transcriptional differences between parasite life-stages, time-points, drug challenges and physiological conditions, for example. This oligonucleotide DNA microarray is presently being employed by other laboratories interested in different aspects of schistosome biology and will facilitate joint and collaborative efforts towards a better understanding of this important pathogen. Future investigations into individual functions, specifics and potential interactions of these newly-defined and potentially fundamentally-important genes will further delineate their differential role(s) in the conjugal biology of schistosomes (several of these functional studies are currently ongoing). Highlighting the potential functional properties of unknown genes, by virtue of their expression profiles, has provided priorities and avenues for further research. Whole genome annotation, in conjunction with further DNA microarray analyses, along with other techniques, such as signal-sequence trap [70], RNAi [71,72], proteomics [73] and investigations of large-scale protein–protein interactions [74] provides prospects to dramatically improve the current understanding of the molecular basis of pathogenicity and pathology by this complex and important major parasite of humans.

Acknowledgements

This work was supported by means of a Wellcome Trust Career Development grant awarded to KFH and a TDR grant awarded to KFH. We thank Maureen Laidlaw and Frances Jones for excellent technical help with cloning and Andrew McArdle for computer support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.molbiopara.2005.01.007.

References

[22] MIAMEexpress (http://www.ebi.ac.uk/miameexpress/).
[24] Schistosoma mansoni EST Project. OpenEGI (http://openegi.ig.etl.go.jp/schistosoma/).


