Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular & Biochemical Parasitology 129 (2003) 103–114
MOLECULAR & BIOCHEMICAL PARASITOLOGY

Evaluation of differential gene expression in *Leishmania major* Friedlin procyclics and metacyclics using DNA microarray analysis

Alka Saxena [a], Elizabeth A. Worthey [a], Shaofeng Yan [a,1], Aaron Leland [a], Kenneth D. Stuart [a,b], Peter J. Myler [a,b,c,*]

[a] *Department of Pathobiology, Seattle Biomedical Research Institute, 4 Nickerson Street, Seattle, WA 98109-1651, USA*
[b] *Department of Pathobiology, University of Washington, Washington, WA 98195, USA*
[c] *Division of Biomedical and Health Informatics, University of Washington, Washington, WA 98195, USA*

Received 30 January 2003; received in revised form 11 April 2003; accepted 11 April 2003

Abstract

The various *Leishmania* species are flagellated protozoans, responsible for a wide spectrum of human diseases. The sequence of the *L. major* genome is nearing completion and a large proportion of the identified genes have yet to be ascribed functions. DNA microarrays containing PCR-amplified DNA from a random amplified genomic library of *L. major* Friedlin (LmjF) [Mol. Biochem. Parasitol. 113 (2001) 337] were hybridized with fluorescent probes made from *L. major* Friedlin RNA from five time-points during differentiation from procyclics to metacyclics. The data were normalized for background and probe intensity and the relative abundance of RNA for each spot was calculated. Almost 15% (1387/9282) of the DNAs showed statistically significant ($P < 0.01$) changes in expression (1.1–5-fold) during the transition, with 1.16% (108) showing up-regulation at two or more time-points and 0.14% (13) showing down-regulation. Northern blot analyses of selected genes confirmed these results. These studies confirmed the stage-specific expression of several known genes, as well as identifying a number of novel genes that are up-regulated in either procyclics or metacyclics.

Keywords: *Leishmania major*; Gene expression; Procyclic; Metacyclic; DNA microarray

1. Introduction

Leishmania, a trypanosomatid parasite, causes a wide spectrum of infections, ranging from self-curing ulcers to often-fatal visceral diseases. It alternates between sandfly and mammalian hosts and has three distinct developmental stages. *Leishmania* exist as motile flagellated, non-infectious, procyclic promastigotes in the midgut of the sandfly host, where it undergoes rapid division [2]. During metacyclogenesis, the lipophosphoglycan on the surface of the promastigotes is modified and the parasites migrate from the alimentary canal of the sandfly host to the proboscis where they form non-dividing, infective metacyclics forms [3]. These are transmitted to the mammalian host, when the sandfly takes a blood meal. Once in the human hosts, metacyclics enter resident skin macrophages. Within the acidic environment of phagolysosomes, these are transformed into non-flagellated, round, amastigotes [4]. The amastigotes undergo rapid division by binary fission; the macrophage lyses, and amastigotes are subsequently released to infect other macrophages. Metacyclogenesis can be mimicked in axenic culture. Cultured promastigotes during an early-log phase (2–5 days after sub-culture) are non-infective, but as the culture approaches stationary phase (9–11 days), the promastigotes rapidly gain virulence and are capable of infecting mouse peritoneal cells in vitro [2].

The changes in the biochemistry and morphology of *Leishmania* from one lifecycle stage to another is most likely the result of programmed changes in the gene expression,

Abbreviations: cDNA, complementary DNA; CPN-10, chaperonin 10; DNA, deoxyribonucleic acid; dATP, deoxyadenosine tri phosphate; dCTP, deoxycytidine triphosphate; dGTP, deoxyguanosine triphosphate; dTTP, deoxythymidine triphosphate; dUTP, deoxyuridine triphosphate; EST, expressed sequence tags; GSS, genome sequence survey; kb, kilo base; mRNA, messenger RNA; ug, microgram; mM, millimolar; μl, microliter; ORF, open reading frame; PCR, polymerase chain reaction; pmoles, pico moles; RNA, ribonucleic acid; RT-PCR, reverse transcription PCR; *r*, coefficient of correlation; SSC, sodium saline citrate; SDS, sodium dodecyl sulphate; 2xYT, yeast tryptone

* Corresponding author. Tel.: +1-206-284-8846; fax: +1-206-284-0313.
E-mail address: peter.myler@sbri.org (P.J. Myler).
[2] Present address: Department of Pathology, Dartmouth-Hitchcock Medical Center, NH 03756, USA.

0166-6851/03/$ – see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0166-6851(03)00100-2

as a response to the changes in the external environment of the parasite [5]. The *Leishmania* genome-sequencing project is progressing rapidly, with the goal of having the complete genome sequenced within the next year. However, of the thousands of new genes that have been identified, only a small proportion have ascribed functions [6]. Thus, there is a need to carry out functional studies to expand the knowledge and understanding of, and finally, control this obligate endoparasite of mammals.

The goal of the present study was to use DNA microarrays to analyze the global changes in gene expression as procyclic promastigotes differentiate in vitro into metacyclic promastigotes. The results of this study confirmed the changes in mRNA abundance for several known genes, as well as identifying several hundred additional genes whose expression changes during metacyclogenesis. This study establishes the utility of genome-wide RNA expression profiling in *Leishmania* and identifies numerous genes with potential roles in metacyclogenesis.

2. Materials and methods

2.1. Array construction

The *Leishmania* microarrays were constructed by arraying 10,464 DNA fragments obtained by PCR-amplification of the inserts from genome survey sequence (GSS) clones from a random shotgun library of *L. major* Friedlin (MHOM/JL/80/Freidlin) [1]. In addition to the clones from this library, PCR-amplified fragments from several previously characterized genes known to undergo stage-specific expression (*A2*, *gene B*, and *SW3*) as well as α-*tubulin* were included in the arrays. The average insert size of the fragment was 1–2 kb, and they represent at least 80% of the LmjF genome. Bacterial samples from the glycerol stocks of the GSS clones arrayed in 384-well plates were inoculated to 96-well plates containing 100 μl of 2xYT per well. After overnight incubation, 1 μl of the bacterial aliquot was taken for direct PCR amplification using 2 units of Taq polymerase and 50 pmol of M13 forward and reverse primers per 100 μl reaction. The PCR products were purified using Arrayit 96-well PCR purification kits (Tele Chem Inc.), then quantitated by Hoechst fluorescence. The quality and quantity of amplified inserts were analyzed by agarose gel electrophoresis and the identity of ∼10% of the fragments was confirmed by DNA sequencing. The PCR products were then transferred to 384-well plates, dried and resuspended in 3 × SSC, before arraying onto poly-lysine coated glass slides using an Omnigrid microarrayer (GeneMachines). The microarrays were designed in two formats: the first had the entire library along with the first half of the library in duplicate and the second contained the entire library along with second half of the library in duplicate. Thus, the entire library is represented three times by combining one slide from each format. Cot-1 DNA, salmon sperm DNA, poly A, and 3 × SSC were spotted in duplicate on each slide as negative controls.

2.2. Leishmania strain and growth condition

Promastigotes of *L. major* MHOM/IL/80/Friedlin (LmjF) were cultured at 27 °C in RPMI 1640 supplemented with 10% heat inactivated fetal bovine serum (Life Technologies) and 5% each of sodium pyruvate, minimal essential amino acid solution, MEM non-essential amino acids solution and 0.2% glucose. The parasites were sub-cultured routinely every fourth day. For differentiation into metacyclics, a fresh sub-culture was grown for 11 days and cells harvested at days 2, 5, 7, 9 and 11. A growth curve was plotted to ascertain when the cells entered stationary phase (between day 7 and day 9). The expression of the metacyclic-specific gene *meta1* [7] was determined at each time-point by RT-PCR and found to be highest in day 9 (data not shown). Examination of second and ninth day cultures using peanut agglutinin (PNA) [8] indicated that they contained mostly procyclic and metacyclic promastigotes, respectively.

2.3. RNA isolation

Cells were collected by centrifugation at 4000 rpm, homogenized by Trizol reagent (Gibco BRL) and frozen at −70 °C until RNA extraction. The RNA was further purified of contaminating genomic DNA or phenol by using RNAeasy Midi kit (Qiagen, Valencia, CA). The RNA samples were ethanol-precipitated, washed once in 70% ethanol, and re-dissolved in water. RNA was quantitated using spectrophotometer. Examination of purified total RNA by gel electrophoresis revealed prominent 18S and 24Sα and 24Sβ ribosomal bands indicating that the RNA was not degraded.

2.4. Preparation of labeled cDNA

Fluorescently labeled cDNA copies of the total RNA pool were prepared by indirect incorporation of fluorescent nucleotide analogs after the first strand reverse transcription reaction. cDNA was prepared by using 30 μg of total RNA, 5 μg oligo (dT) primer, 25 mM each of dATP, dCTP and dGTP; 15 mM of dTTP, 10 mM of amino allyl dUTP, 10 mM DTT and 200 units reverse transcriptase (Superscript II, Gibco-BRL) in reaction buffer provided by the manufacturer. The RNA and primer were heated to 70 °C for 10 min and then snap-chilled on ice, before adding other components of the reaction. The reaction was incubated at 42 °C for 2 h. The RNA was then degraded and reaction neutralized. The cDNA products were concentrated using Microcon-30 microconcentrators (Amicon) and coupled to 2 nmol of either Cy3-dUTP or Cy5-dUTP in the presence of sodium bicarbonate by incubating the reaction at room

temperature for 1 h. The reactions were then quenched, combined and purified.

2.5. Microarray hybridization

The combined cDNA probes were applied to the array slides in the presence of hybridization buffer consisting of 3 × SSC, 0.3% SDS and 1.5 μg of poly A, 5 μg of salmon sperm DNA and 2 μg of Cot-1 DNA. Hybridization was carried out in a hybridization chamber (Tele Chem Inc.), submerged in a 63 °C water bath for 12–16 h. The microarrays were washed at room temperature for a minute each in 1 × SSC/0.03% SDS (wash 1) and 1 × SSC (wash 2); for 20 min in 0.2 × SSC (wash 3) and 10 min in 0.05 × SSC (wash 4). The arrays were spun dry in a centrifuge and scanned using a GenePix Pro 4000 scanner (Axon Instruments, South San Francisco, CA) to determine the fluorescent intensities of the two dyes for each spot.

2.6. Data analysis

The data was extracted and initially analyzed using the GenePix Pro 3.0 software supplied with the scanner. Local background was subtracted from the intensity value of each spot on the array. Spots were manually examined to assess their quality and those that exhibited poor quality or were saturated were discarded from further analysis. Poor quality spots were removed if they were very small; were irregularly shaped; or the total pixel intensity of the spot was 55% lower than the median background intensity at both wavelengths. Data was exported as 16-bit TIFF files, and further analyzed using the statistical analysis software package, GenePlus version 1.2 (Enodar Biologic Corporation, WA) [9]. Within-array normalization was based on linear regression fit, using a least square line and adjusting the slope such that all the points lie on the diagonal line. Normalization between arrays was carried out using multiplicative and additive heterogeneity factors calculated for each of the arrays. In order to compare expression levels between RNAs, the $\log_{10}$-transformed Cy5/Cy3 ratios were calculated from the normalized values. The ratios from reverse-labeled experiments were reciprocated before analysis. This software uses a modified Bonferroni's correction for multiple testing [10] to estimate *Z*-score for each transcript, which are then translated into *P* values to measure the significance of findings.

2.7. Northern blot analysis

Briefly, total RNA (10 μg) was size fractionated by electrophoresis and analyzed by Northern blotting as previously described [11]. Hybridization with radiolabeled PCR probes, generated by random priming, was carried out at 42 °C in the presence of Ultra-hyb (Ambion Inc.). The filters were washed and signal quantitated using a Storm Phosphorimager.

2.8. Annotation

Where the identity of the GSS clone had been confirmed by sequencing of the entire insert (~10% of spots), this sequence was used to predict the location of genes within each spot. For the remainder of spots, the GSS end-sequence was compared against sequence generated from the *L. major* Friedlin genome project in order to identify the sequence spotted on to the array. All annotation was performed using the Artemis software [12]. The current *L. major* codon usage table (obtained from http://www.kazusa.or.jp/codon/) and the sequence were imported into Artemis in order to produce the codon scores (in a sliding window) for each of the forward and reverse reading. The Institute for Genome Research (TIGR)'s Glimmer (Gene Locator and Interpolated Markov Modeler) package version 2.0 [13,14], which uses interpolated Markov models (IMMs) to identify the coding regions within a DNA sequence and distinguish them from non-coding DNA, was also used to predict coding regions within the sequence. Glimmer was trained with previously identified protein-coding genes from *L. major* Friedlin chromosome 1 and 4. The gene predictions obtained from Glimmer 2.0 were then parsed into a format that could be imported into Artemis. An in-house adaptation of Fickett's algorithm Testcode, which identifies protein-coding sequences by plotting a measure of the non-randomness of the composition at every third base [15], was also used to identify potential protein coding regions. Finally, predictions from an adaptation of Genescan, which predicts the location of probable genes by Fourier analysis [16] were also parsed and examined using Artemis.

The amino acid sequences predicted from each putative gene within these sequences were used to carry out a local blastp search of the non-redundant protein database and tblastn searching of a kinetoplastid-specific nucleotide database. These searches [17] were run through Artemis. Generally, hits with blast scores of >50 and *e*-values of $<1 \times e^{-6}$ were considered potentially significant, although some exceptions were made upon visual inspection of the alignments. Each protein sequence was then searched against numerous collections of protein motifs and families (SWISS_PROT Release 39.27, ProDom version 2001.2, SMART version 3.3, PROSITE Release 16.46, Pfam Version 6.6, PRINTS Release 31.0, Domo Version 2.0, and BLOCKs Release 13.0), as well as being blasted against the database of Clusters of Orthologous Groups of proteins (COGs; [18]), and NCBI's Conserved Domain Database and Search Service, version 1.54 [19]. These analyses allowed the putative functions to be identified for many of the genes spotted onto these arrays.

3. Results

Changes in mRNA abundance during metacyclogenesis were examined by genome-wide expression profiling

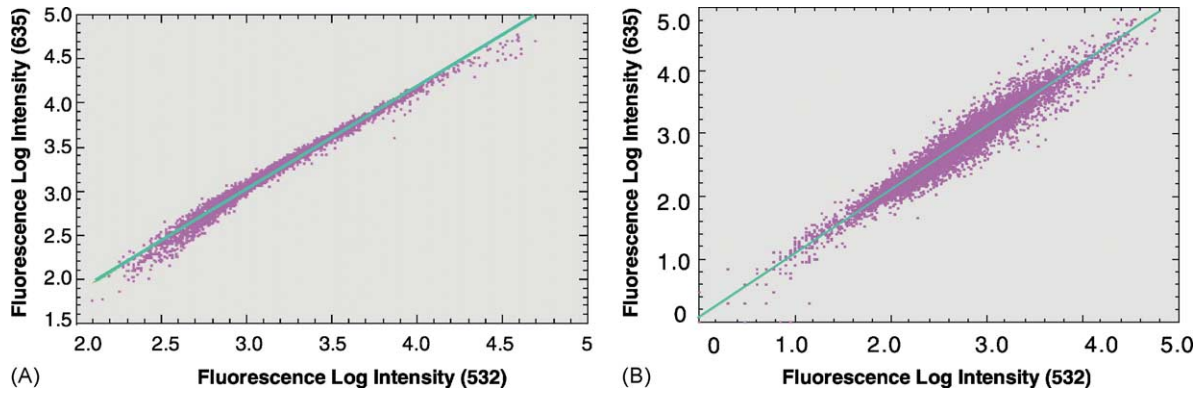

Fig. 1. Reproducibility of microarrays. A $\log_{10}$ plot of Cy3 (532 nm) vs. Cy5 (635 nm) calibrated fluorescent response from representative hybridizations for (A) Cy3-labeled day 2 RNA vs. Cy5-labeled day 2 RNA, and (B) Cy3-labeled day 2 RNA vs. Cy5-labeled day 9 RNA. The line of best fit is shown in both panels.

using DNA microarrays containing PCR amplified fragments from more than 10,000 GSS clones of *L. major* Friedlin (MHOM/JL/80/Freidlin) [1]. Multiple replicates of all hybridizations were performed to account for sample heterogeneity and possible variation due to hybridization. A series of replicate experiments using the same RNA sample were done initially to obtain an estimate of the accuracy and precision of the system. Microarrays were hybridized with fluorescently labeled Cy3 and Cy5 cDNA, both prepared from day 2 (procyclic) RNA. This theoretically should give a log expression ratio (Cy5/Cy3) of 1 for all the elements arrayed on to the slide. A $\log_{10}$ plot of Cy3 (532 nm) versus Cy5 (635 nm) calibrated fluorescent response from a representative hybridization is shown in Fig. 1A. The plot shows tightly packed distribution of most genes along the line of best fit, with a regression correlation coefficient (r) of 0.99. When a similar experiment was conducted with Cy3-labeled day 2 cDNA and Cy5-labeled day 9 (metacyclic) cDNA, the distribution was not as tightly packed with a fraction of the points deviating from the line of best fit (Fig. 1B), indicating possible differential expression.

In order to compare the relative contributions of experimental and biological variation, we compared the results above, with those obtained from hybridization of cDNAs from RNA extracted from the same day 2 and day 9 cultures (Fig. 2A) with those from different day 2 and day 9 cultures (Fig. 2B). The correlation coefficient (r) for the two duplicate experiments conducted with cDNAs from the same RNAs was 0.794 (Fig. 2A), while RNA extracted from different cultures showed r of 0.787 (Fig. 2B). Thus, the biological variation was similar to the experimental variation.

To study the temporal changes in the gene expression as procyclics differentiate into metacyclics, we carried out microarray hybridization experiments with RNA isolated from different stages of the life cycle (days 2, 5, 7, 9 and 11). For all these studies, day 2 procyclics were used as the reference day of growth. The comparison of day 2 (procyclic) and day 9 (metacyclic) cultures was carried out using eight

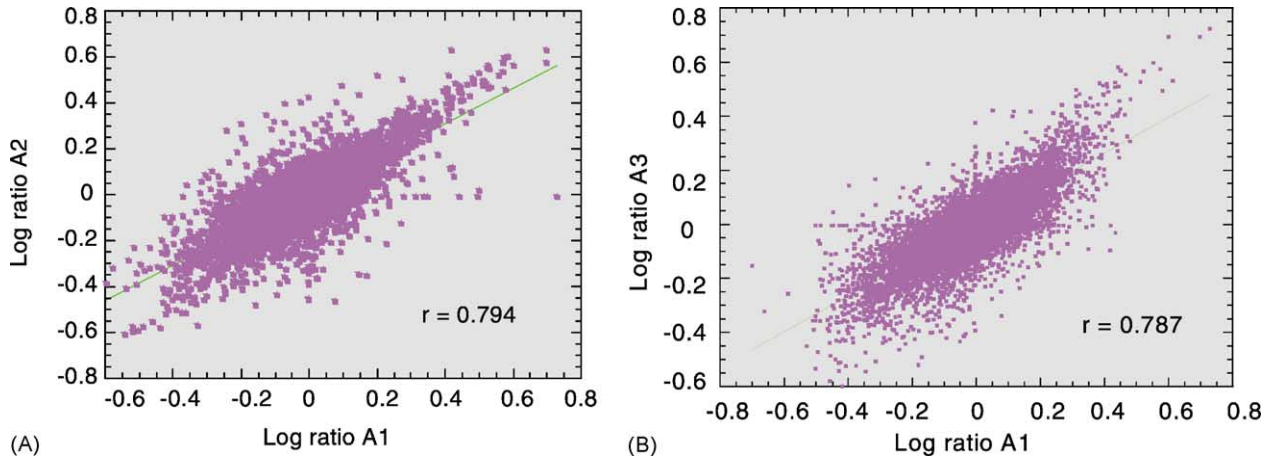

Fig. 2. Comparison of expression ratios in duplicate experiments. Log_{10} expression ratios (Cy5:Cy3) are plotted for arrays probed with Cy5-labeled day 9 RNA vs. Cy3-labeled day 2 RNA. (A) Duplicate hybridizations (A1 and A2), using cDNA from the same pair of day 2 and day 9 RNA samples. (B) Hybridizations (A1 and A3), using cDNAs from different pairs of RNA samples.

separate hybridizations, representing 12 replicates for each of the 10,479 DNA elements spotted on the microarray slides. After removing elements with more than two missing data points, data from 9282 DNA elements remained for subsequent statistical analysis using GenePlus (Enodar Biologic Corporation, WA). This software estimates a correction factor to account for sample-specific heterogeneity, which is used to normalize the data [9]. The corresponding standard errors for each gene are calculated using estimating equation theory [20,21]. *Z*-scores for each gene are then computed as the ratio of mean difference between the two groups/stages for each gene, divided by the standard error for the corresponding gene. To measure the significance of the findings, *Z*-scores are translated into *P*-values under asymptotic normality. To address the multiple comparison issue, the threshold is adjusted for declaring genes differentially expressed using a modified Bonferroni's correction [10]. The higher the *Z* score, the greater is the confidence that the transcript is differentially expressed between two stages.

Of the 9282 DNA elements analyzed, 5.1% (472) showed a statistically significant ($P < 0.01$) increase in expression in day 9 metacyclics, while 1.5% (141) showed significantly higher expression in day 2 procyclics (Table 1). Of the former, 21.3% (101) showed 2–5-fold higher mRNA abundance in metacyclics, while the remaining 78% (371) showed 1.10–1.99-fold higher mRNA abundance. Of the 141 DNA elements with higher expression in procyclics, 29.5% (42) showed of 2–3.3-fold lower mRNA abundance in metacyclics, while 70% (99) showed 1.2–1.99-fold lower mRNA abundance.

Similar analyses were carried out on the results of four hybridizations, representing six replicates, each comparing day 2 procyclic RNA with RNAs from day 5, day 7, and day 11. These data were analyzed on a pair-wise basis using GenePlus; and 17, 2, and 38 DNA elements, showed significantly higher mRNA abundance at day 5, day 7, and day 11, respectively, when compared to day 2; while 8, 5, and 9 showed significantly lower mRNA abundance ($P < 0.01$) (Table 1). When the combined data from the five time-points were analyzed by regression analysis using a simple linear model, 194 DNA elements showed a significant ($P < 0.01$) linear increase in mRNA abundance over for the entire time course and 85 showed a linear decrease (Table 1). Regression analysis fitting to a quadratic curve revealed 268 DNA elements that showed an initial increase and subsequent decrease in expression patterns and 290 showed an initial decrease and subsequent increase (Table 1). Further analysis of this data showed that a total of 857 DNA elements showed higher mRNA abundance in at least one time-point during the process of procyclic differentiation into metacyclics while 530 DNA elements showed lower mRNA abundance. Of these 1387 DNA elements, 121 (108 higher and 13 lower) showed significant differential expression at two or more time points (Table 1).

Northern blot analysis were performed using 13 DNA fragments that showed significant differential mRNA abundance during metacyclogenesis, as well as three that showed no significant difference in mRNA abundance between procyclics and metacyclics. As indicated in Fig. 3, we found an excellent correlation between the microarray and Northern blot results. Ten DNA elements which demonstrated 1.5–4.9-fold higher mRNA abundance in day 9 metacyclics by microarray analysis all showed similar, but not identical, differences in transcript abundance (1.3–6.4-fold) by Northern analysis (Fig. 3A). Three DNA elements with lower (0.5–0.6-fold) mRNA abundance in metacyclics (i.e. 1.5–1.9-fold higher expression in day 2 procyclics) by microarray analysis showed similar lower mRNA abundance by Northern blot analysis (Fig. 3B). Finally, three DNA elements which showed no statistically significant differences in mRNA abundance by microarray analysis showed only marginal (0.8–1.0-fold) differences by Northern blot analysis (Fig. 3C). It is interesting to note that this was true even of the one DNA element (lm18f12) that showed an apparent (but not statistically significant) 0.5-fold mRNA abundance in metacyclics by microarray analysis.

When the data for the day 2 versus day 9 hybridizations (which included six replicates for each dye-cDNA combination) were analyzed by comparing the $\log_{10}$ ratio for the forward-labeled experiments (day 2 procyclics labeled with Cy3 and day 9 metacyclics labeled with Cy5) with reciprocal-labeled experiments (day 2 procyclics labeled with Cy5 and day 9 metacyclics labeled with Cy3), we observe a clear indication of dye-bias (Fig. 4). In the absence of dye-bias, the scatter plot of normalized data (Fig. 4B) should show a linear trend clustered around the diagonal. However, a "hump" along the horizontal zero axes (indicated by the box in Fig. 4) suggests that a small fraction of DNA elements have a dye-bias associated with them. Statistical testing between forward- and reverse-labeling, using GenePlus, showed that 21 (0.2%) of the 9282 DNA

Table 1
DNA fragments with differential expression between procyclics and metacyclics ($P < 0.01$)

	2v5	2v7	2v9	2v11	Linear	Quadratic	1 time point	≥2 time point	Total
Higher in metacyclics	17	2	472	38	194	290[a]	749	108	857
Higher in procyclics	8	5	141	9	85	268[b]	517	13	530
Total	25	7	613	47	279	558	1266	121	1387

[a] An initial decrease in gene expression followed by an increase.
[b] An initial increase followed by a decrease in gene expression.

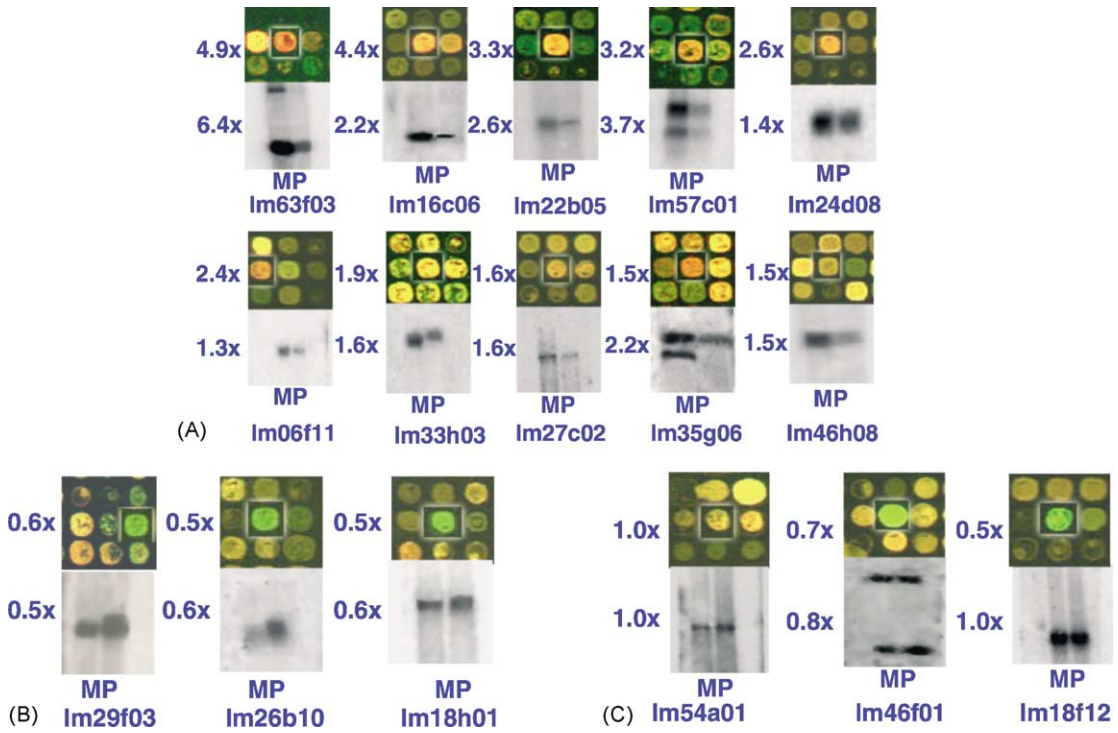

Fig. 3. Northern blot confirmation of microarray results. The results are shown from representative microarrays (top panels) and Northern blots (bottom panels) for DNA elements with: (A) up-regulation in day 9 metacyclics; (B) down-regulation in metacyclics; and (C) no significant regulation. The DNA element is indicated by the boxed spot in the top panels and the name below the bottom panels. In the bottom panels P denotes day 2 (procyclic) RNA and M denotes day 9 (metacyclic) RNA. The ratio of mRNA level in day 9 RNA to day 2 RNA as determined by both methods is indicated to the left of each panel.

elements which could be analyzed showed a significant ($P < 0.01$) dye-bias. Of these 21 DNA elements, five showed apparently higher mRNA abundance in metacyclics, three showed apparently lower mRNA abundance in metacyclics (i.e. higher in procyclics), and the remaining 13 showed no significant difference between stages. Northern blot analysis of seven of these elements confirmed the misleading effect of dye-bias. Of three elements with apparently higher mRNA abundance in metacyclics by microarray analysis, only one (lm66b12) showed higher mRNA levels by Northern analysis, while one (lm01e05) was equally expressed in both stages, and in the other (lm03d08) transcript levels were

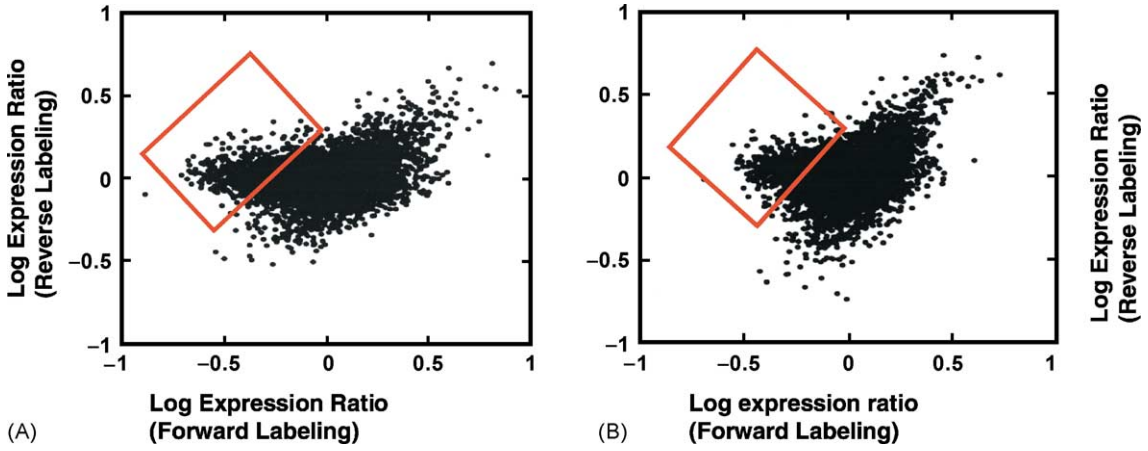

Fig. 4. Dye-bias identified by dye-swap analysis. The $\log_{10}$ expression ratios for 6 arrays probed with Cy3-labeled day 2 RNA vs. Cy5-labeled day 9 RNA (Fwd labeling) are plotted against those for 6 arrays probed with Cy5-labeled day 2 RNA vs. Cy3-labeled day 9 RNA (Rev labeling), (A) before normalization and (B) after normalization. The boxed regions demarcate elements with potential dye-bias.

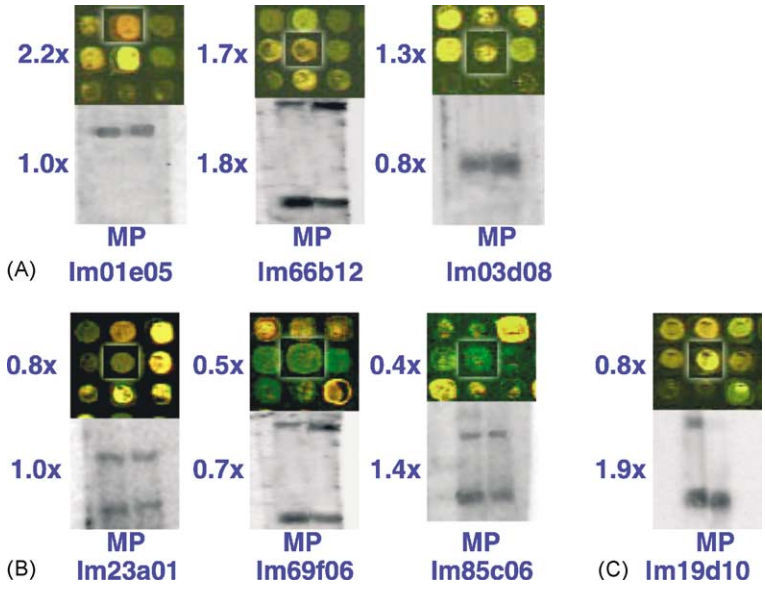

Fig. 5. Northern blot analysis of DNAs with dye-bias. The results from representative microarrays (top panels) and Northern blots (bottom panels) are shown for DNA elements with: (A) up-regulation in day 9 metacyclics, (B) down-regulation in metacyclics and (C) no significant regulation. The ratio of mRNA level in day 9 RNA to day 2 RNA as determined by both methods is indicated to the left of each panel. Symbols used are the same as Fig. 3.

actually higher in procyclics (Fig. 5A). Even in first case, one of the two transcripts detected was more abundant in procyclics. Of the three DNA elements with apparently lower mRNA abundance in metacyclics, one (lm23a01) showed equal abundance of transcript in both stages by Northern analysis, while the other two (lm85c06 and lm69f06) showed the presence of two transcripts, at least one of which was more abundant in metacyclics. Another element (lm19d10) with significant dye-bias demonstrated no significant difference in mRNA abundance according to the microarray analysis, but showed the presence of two transcripts by Northern analysis, one of which was higher in procyclics and the other higher in metacyclics. It is interesting to note that several of the DNA elements, which showed significant dye-bias, detected two transcripts in Northern blots.

Of the 613 DNA elements that showed significant ($P < 0.01$) differential mRNA abundance in the day 2 versus day 9 hybridizations, 405 had sequence available from either one or both ends, and 313 of these showed matches against the LmjF sequence database (Table 2). Twenty-one of these were removed as redundant, since they represented sequences contained in other elements. Low quality sequences (34) were also removed from further analysis.

Elements, which contained more than one, predicted protein-coding ORF (23) and elements with substantial inter-ORF sequence (39) were not classified further, since these may hybridize with two adjacent mRNAs. The remaining 196 represented DNA sequences containing a single protein-coding gene. Of the 133 genes potentially up-regulated in metacyclics, 63 (47.3%) could be classified into one of 12 functional categories on the basis of sequence similarity to known genes (Table 3). Similarly, only 25 (39.6%) of the 63 genes potentially down-regulated in metacyclic (higher expression in procyclics) could be classified (Table 4).

The genes up-regulated in metacyclics are involved in cell growth and division, metabolism, protein destination, protein synthesis, signal transduction, transcription and RNA processing, as well as the surface antigens. Several protein kinases and phosphatases are also upregulated in metacyclics. In eukaryotes, these mediate the phosphorylation and dephosphorylation of serine, threonine and tyrosine residues in proteins and function as control switches in cellular networks. Changes in protein phosphorylation during parasite life cycle suggest that these enzymes play an important role in parasite differentiation, virulence and cell division [22,23]. Genes involved in cell growth and division, cellular organization/biogenesis, and transport are down-regulated in metacyclics. Several specific genes, such as *MKK2* (lm61d12), which encodes a MAP kinase kinase

Table 2
Identification of differentially expressed genes

	Procyclics	Metacyclics	Total
No. of DNA elements	141	472	613
Dye-bias	3	5	8
Sequence available	96	309	405
Sequence match against LmjF	75	238	313
Redundant sequence	2	19	21
Low quality sequences	4	30	34
Elements with two genes	3	20	23
Intergenic sequences	3	36	39
Elements with single gene	63	133	196
Classified genes	25	63	88
Unclassified genes	38	70	108

Table 3
Identity and classification of genes up regulated in metacyclics

Category	GSS number	Putative function	Fold-regulation	*P*-value
Cell growth, division and DNA synthesis	lm29a11	Kinesin-like protein	1.42	0.0005
	lm38b05	Chromosome segregation ATPase	1.7	0.0005
	lm71e12	DNA-dependent RNA polymerase, subunit N	2.1	1.31E−07
	lm95g06	Cyclin	1.3	0.009
Cell rescue, defense, death and aging	lm01a06	PSA-2	1.2	0.0002
	lm54c11	PPG3	1.5	0.002
	ln03b08	Sodium stibogluconate resistance protein	1.4	0.0001
	lm51d04	P-glycoprotein E	1.5	0.0002
	lm83a02	P-glycoprotein	1.7	0.007
Cellular organization/biogenesis	lm54a03	D-Alanine-D-alanine ligase	1.3	0.002
	lm90h11	Flagellar calcium-binding protein	2.2	3.82E−06
Energy generation	lm26f07	Tryparedoxin	1.3	0.0002
	ln03c01	ATPase	1.5	2.39E−05
	lm99g09	Vacuolar-type H+ ATPase subunit C	1.4	0.001
	lm17b08	Thioredoxin	1.4	0.0003
Metabolism	lm03h06	Coenzyme A transferase	1.5	0.0002
	lm26a07	Membrane-bound acid phosphatase	1.6	0.0001
	lm27a12	Methionine synthase	3.3	3.70E−06
	lm39f12	Aldose 1-epimerase	1.9	0.0002
	lm26a09	Long chain polyunsaturated fatty acid elongation enzyme	1.4	0.005
	lm35f03	Carnitine *o*-acyltransferase	1.6	0.002
	lm03h10	Amidohydrolase	1.4	0.0004
	lm10a12	PLP-dependent methyltransferase	1.5	0.0001
	lm81b07	Diacylglycerol acyltransferase	2.3	7.4E−07
	lm03f12	Dehydrogenase	1.5	0.001
Signal transduction	ln03h07	Protein kinase	1.5	0.002
	lm09a04	GAF domain containing protein	1.6	0.001
	lm16d07	Phosphatidylinositol-3′ kinase	2.2	0.005
	lm34d12	Serine/threonine specific protein phosphatase	1.2	0.001
	lm58c08	Serine/threonine protein kinase	1.4	0.0003
	lm61d12	Mitogen-activated protein kinase kinase	2.5	1.00E−05
	lm25e01	Tetratricopeptide repeat containing protein	1.7	0.0004
	lm74g05	Serine/threonine protein phosphatase	2.2	0.0003
Protein synthesis	lm61d10	Ribosomal protein L6	1.6	0.0003
	lm69b01	Translation initiation factor IF-2	2.1	9.90E−06
	lm74c10	Ribosomal protein L26	1.4	0.008
	lm98a10	Ribosomal protein L37A	1.8	5.70E−05
	lm46c04	Ribosomal protein L21E	1.9	0.0003
Structural RNA	lm27c02	16SrRNA	1.7	0.003
Transport	lm76a02	ABC transporter	1.7	0.008
	lm62e09	Amino acid permease	1.41	0.001
	lm78c08	Pteridine transporter	1.6	0.0002
	lm44c01	Cobalamin-independent methionine synthase	2.9	9.58E−07
	lm97f10	PLP-dependent aminotransferase	2.1	2.49E−05
	lm47h02	Pentose-5-phosphate-3-epimerase	1.7	5.94E−05
Protein destination	lm09b02	RJS/HERC2-like protein	1.7	0.006
	lm10f01	Calpain-like protease	1.5	0.001
	lm15f07	HSP70	1.8	2.05E−06
	lm33h06	Glycoprotein endopeptidase	1.3	0.008
	lm52a06	Co-chaperonin CPN-10	1.5	0.0001
	lm59c10	26S proteasome, non-ATPase regulatory subunit	1.4	7.73E−07
	lm67g05	HSP70-related protein ORP150RP	1.7	0.003
	lm22d10	Cyclophilin (TcCYP)	1.8	3.50E−06
	lm61f04	DNAJ molecular chaperone	1.4	0.001
	ln09b02	Proteosome regulatory non-ATPase subunit	1.5	0.0008
	lm30b08	Calpain-like protein	1.9	5.7E−05
Transcription and RNA processing	lm71a11	Poly A binding protein 1	1.9	3.43E−05
	lm19h10	mRNA capping enzyme	2.2	2.40E−05
	lm44f11	RNA binding protein	1.4	0.0006
	lm62h08	Poly A export protein PAXP	1.4	0.005
	lm03a08	SNF2 family helicase	1.3	0.002
	lm59d02	ATP-dependent chromatin remodeling protein	1.4	0.009
	lm16c06	Mat-1 protein	4.4	1.19E−06

Table 4
Identity and classification of genes up-regulated in procyclics

Category	GSS number	Putative function	Fold-regulation	*P*-value
Cell growth division and DNA synthesis	lm13f05	PolA family member	2.11	0.00002
	lm24b02	Cyclin	1.822	0.009
	lm26b10	Chromosome segregation ATPase	1.94	0.00004
	lm43f02	Chromosome segregation ATPase	1.41	0.007
	lm75h09	Possible kinesin-related protein	1.2	0.001
Cell rescue, defense, death and aging	lm43a05	DNA binding protein	2.96	0.0003
	lm80e11	ATPase involved in DNA repair	1.327	0.0009
	lm55d05	AMA1 protein	1.4	0.0001
Cellular organization/biogenesis	lm18h01	Paraflagellar rod protein	1.31	0.0001
	lm24f04	Dynein ATPase	1.896	0.0047
	lm26b12	Microtubule-associated protein ATPase	2.12	0.007
	lm43g01	Axonemal dynein	2.24	0.0007
	lm57b10	Alpha coat protein	1.25	0.0008
	ln10f09	Axonemal dynein	2.3	0.001
Energy generation	lm24f06	Proline dehydrogenase	1.348	0.001
	lm78c09	NADH-ubiquinone oxidoreductase	1.34	0.001
Intracellular trafficking	lm23b09	Beige protein homolog	1.664	6.3E−05
	lm55a05	Signal peptidase like protein	2.1	0.002
Protein destination	lm46b01	T-complex protein 1 (gamma subunit) homolog	1.864	0.00002
	lm73c12	Calpain-like protease	1.579	0.006
Protein synthesis	lm51b12	DEAD box RNA helicase	1.893	0.003
Transcription and RNA processing	lm43d04	TFIIIB	1.66	0.007
Transport	lm36d06	Long chain fatty acyl CoA synthetase LCFACAS3	1.38	0.0001
	lm48e06	Calcium motive P-type ATPase	1.32	0.0001
	lm78c11	ABC-type multidrug/protein/lipid transport system, ATPase component	1.3	0.0005

[24], *gene B* (PCR product) [25], *A2* gene (PCR product) [26] and *CPN-10* (lm52a06) (Zamora-Veyl F.B. et. al.; Gene Bank Accession no. AF394959) have been previously reported to be up-regulated in metacyclics or amastigotes. The *Mat-1* gene (lm16c06), conserved in the genus *Leishmania* [27], which is selectively expressed in non-dividing infective stage of *L. major* [8] was also upregulated in metacyclics.

4. Discussion

DNA microarray analysis was used to examine the temporal changes in gene expression as procyclic promastigotes differentiate into metacyclics. Since DNA microarrays allow examination of gene expression on a genome-wide scale, these studies revealed substantial new information about the dynamics of transcript abundance during this differentiation process. However, a complete view of gene expression was not possible, since the *Leishmania* genome is not yet completely sequenced and annotated, and the arrays used do not represent every gene present in the *Leishmania* genome. Nevertheless, our data indicates that 5.1% of the genes showed statistically significant higher relative mRNA abundance in metacyclics, while 1.5% showed higher relative mRNA abundance in procyclics. In addition, another 9.2% showed an increase in transcript abundance (i.e. up-regulation) and 5.7% showed a decrease in transcript abundance (i.e. down-regulation) at one or more intermediate points during metacyclogenesis (Table 1). These numbers probably represent a conservative estimate of the number of genes that undergo changes in gene expression, since we used very stringent cut-off criteria ($P < 0.01$) for the statistical analysis.

The use of rigorous statistical methods, rather than simple fold-changes, is now becoming the standard for analysis of DNA microarray experiments [28] since the latter does not take into account the variability of measurements being considered [29] and can lead to substantial false positives and false negatives [30]. This can be seen in the case of lm18f12, which was apparently (but not statistically significantly) down-regulated by 2.1-fold according to the microarray study, but was found to be equally expressed in procyclics and metacyclics by Northern blot analysis (Fig. 3C). Conversely, the use of *P*-value estimates allowed the identification, with high confidence, of many genes that were regulated by less than two-fold. Indeed, only 23% of the 613 elements that showed statistically significant ($P < 0.01$) regulation between day 2 procyclics and day 9 metacyclics showed regulation of two-fold or more. Confirmation of several of these results by Northern blot analysis (Fig. 3A and B) suggests that these represent real changes in mRNA abundance. It should also be noted that these values

represent the change in abundance of each mRNA relative to the total RNA concentration, rather than a change in actual intracellular concentration of individual mRNAs. Nevertheless, further confidence in the validity of a substantial portion of these candidates can be taken from the finding that 121 (8%) showed significant regulation at more than one time point (Table 1). This number is likely an under-estimate of the true number of stage-regulated genes, since the relatively stringent cut-off probably lead to substantial false negatives, especially for time points with only six replicates.

Studies detailing developmental transitions in *Saccharomyces cervisiae* (log to stationary phase sporulation) showed a two-fold or greater change in expression of 27% [31] of yeast genes, whereas only 1.5% (143/9282) of the LmjF elements analyzed showed changes of this magnitude between procyclics and metacyclics. In addition, the largest changes observed during procyclic-to-metacyclic differentiation were only five-fold, quite small compared to those observed in other systems. However, other studies that have been carried out to look at developmental changes in *Leishmania*, have also reported that there are only a small number of genes that are developmentally regulated at the level of transcript abundance [32–34]. Thus, it appears that metacyclogenesis results in only modest (but measurable) changes in mRNA levels, perhaps a reflection of the apparent absence of transcriptional regulation in these organisms [35]. Nevertheless, more substantial changes in transcript abundance have been observed during promastigote-to-amastigote differentiation in *L. donovani* using similar microarray analyses (unpublished data). Thus, the biological significance of these relatively subtle changes in mRNA abundance during metacyclogenesis is not yet clear, as they may not accurately represent changes in protein expression during this transition. It is becoming clear that in most of the organisms there may be a poor correlation between transcript and protein abundance [36], although in yeast, it has been recently suggested that there is a close association between cellular mRNA content, regulated also at the transcriptional level, to its efficiency of translation mediated by a fine tuning of codon usage strategy [37].

A single microarray experiment is subject to considerable variability, because of high noise-to-signal ratios, differences in hybridization conditions, and biological variation between cell populations [38]. In our hands, replicate experiments show a coefficient of correlation of ∼0.8, with little difference between experiments using the sample RNA samples and those using RNAs from independent cultures. Thus, the variation due to experimental errors appears to be of the same order as that due to biological variation. In addition, the efficiency with which Cy3 and Cy5 dyes can bind to a cDNA may vary depending on a number of different factors, such as the properties of the dye or the sequence of the transcript. Comparison of day 2 versus day 9 hybridizations showed the presence of a substantial "hump" along the horizontal zero axis of the plot of $\log_{10}$ ratio for the forward-labeled experiments with reciprocal-labeled experiments (Fig. 4), indicating that a substantial number of DNA elements may show dye-bias. However, this proved to be statistically significant ($P < 0.01$) in only a small fraction (0.2%) of cases. Nevertheless, in several of these cases, as indicated by the Northern blot data in Fig. 5, the observed dye-bias may have caused false-positives if not recognized.

Identification of stage-regulated genes was complicated by the use of random shotgun GSS clones to construct the arrays, since a small proportion (∼10%) of the elements contain inter-ORF region or more than one predicted protein-coding ORF and probably hybridize with adjacent mRNAs. In addition, a number of the elements have not yet been sequenced, and some that have been do not match annotated sequences within the, as-yet incomplete, LmjF sequence database. It is also likely that some regions of the genome (and hence genes from these regions) will not be represented in the GSS array, since it has only ∼80% coverage of the genome [1]. Nevertheless, until the LmjF genome is completed and annotated, these arrays provide a valuable source of identifying biologically important genes. Indeed, these experiments have identified more than 100 genes, which are up- or down-regulated during the transition from procyclics to metacyclics. These include genes that have been previously reported, such as the *PSA-2* [39], *MAP kinase kinase* [24], *gene B* [40], *Mat-1* [27], and *A2* [26] genes. In general, metacyclics appear to down-regulate some of the genes involved in cell growth and division, cellular organization/biogenesis, and transport; and up-regulate genes involved in cell growth and division, metabolism, protein destination, protein synthesis, signal transduction, transcription and RNA processing. Down-regulation of cell growth and division genes, particularly those involved in DNA replication, is not surprising since metacyclics are no longer actively dividing. However, the up-regulation of genes involved in RNA processing, protein synthesis and protein synthesis suggests that they are actively synthesizing new proteins, perhaps in preparation for differentiation into amastigotes. This is further supported by the up-regulation of genes involved in signal transduction and metabolism. It will be interesting to see whether these genes are also up-regulated in amastigotes. However, the majority of regulated genes belonged to the unclassified category, indicating that they have as-yet unknown function. It will be interesting to study these genes as they might encode proteins with functions specific to *Leishmania* and may provide novel targets for therapeutic intervention.

These experiments clearly demonstrate the effectiveness of using DNA microarrays for genome-wide analysis of differential gene expression in *L. major*. We have also used the LmjF arrays for hybridization with RNA from *L. donovani* and found that the majority (>85%) of the elements have similar signal to that with LmjF RNA, indicating that they will be suitable for analysis of differential gene expression in *L. donovani* (unpublished data). Indeed, preliminary experiments comparing promastigotes and amastigotes from this species indicate that substantial changes in gene

expression can be seen and many of these appear to have a greater magnitude than that observed in the present study. Similar results have been reported by others using DNA microarrays containing a smaller number of EST clones [41].

Acknowledgements

We would like to thank Dr. Stephen Beverley (Washington University) for the kind gift of the LmjF GSS clones. We also thank Drs. Jeff Delrow and Helmut Zarbl (Fred Hutchison Cancer Research Institute) for the assistance in printing the microarray slides, and Drs. Lue Ping Zhao and Chun Cheng (Fred Hutchison Cancer Research Institute) for their help with the statistical analysis. This work was supported by PHS grants AI17375 to KDS and AI47234 to PJM from the National Institutes of Health.

References

[1] Kopyants NS, Clifton SW, Martin J, Pape D, Wylie T, Li L, et al. A survey of the *Leishmania major* Friedlin strain V1 genome by shotgun sequencing: a resource for DNA microarrays and expression profiling. Mol Biochem Parasitol 2001;113:337–40.
[2] Sacks DL, Perkins PV. Identification of an infective stage of *Leishmania* promastigotes. Science 1984;223:1417–9.
[3] Sacks DL, Hieny S, Sher A. Identification of cell surface carbohydrate and antigenic changes between non-infective and infective developmental stages of *Leishmania major* promastigotes. J Immunol 1985;135:564–9.
[4] Wiese M. A mitogen-activated protein (MAP) kinase homologue of *Leishmania mexicana* is essential for parasite survival in the infected host. EMBO 1998;17:2619–28.
[5] Flinn HM, Smith DF. Genomic organization and expression of differentially regulated gene family from *Leishmania major*. Nucleic Acids Res 1992;20:755–62.
[6] Myler PJ, Beverley SM, Cruz AK, Dobson DE, Ivens AC, McDonagh PD, et al. The *Leishmania* genome project: new insights into gene organization and function. Med Microbiol Immunol 2001;190:9–12.
[7] Nourbakhsh F, Uliana SRB, Smith DF. Characterisation and expression of a stage-regulated gene of *Leishmania major*. Mol Biochem Parasitol 1996;76:201–13.
[8] Brodin TN, Heath S, Sacks DL. Genes selectively expressed in the infectious (metacyclic) stage of *Leishmania major* promastigotes encode a potential basic-zipper structural motif. Mol Biochem Parasitol 1992;52:241–50.
[9] Thomas JG, Olson JM, Tapscott SJ, Zhao LP. An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. Genome Res 2001;11:1227–36.
[10] Hochberg Y. A sharper Bonferroni procedure for multiple test of significance. Biometrika 2002;75:800–2.
[11] Lodes MJ, Merlin G, deVos T, Ghosh A, Madhubala R, Myler PJ, et al. Increased expression of *LD1* genes transcribed by RNA polymerase I in *Leishmania donovani* as a result of duplication into the *rRNA* gene locus. Mol Cell Biol 1995;15:6845–53.
[12] Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualisation and annotation. Bioinformatics 2000;16:944–5.
[13] Salzberg SL, Delcher AL, Kasif S, White O. Microbial gene identification using interpolated Markov models. Nucleic Acids Res 1998;26:544–8.
[14] Salzberg SL, Pertea M, Delcher AL, Gardner MJ, Tettelin H. Interpolated Markov models for eukaryotic gene finding. Genomics 1999;59:24–31.
[15] Fickett JW. Recognition of protein coding regions in DNA sequences. Nucleic Acids Res 1982;10:5303–18.
[16] Tiwari S, Ramachandran S, Bhattacharya A, Bhattacharya S, Ramaswamy R. Prediction of probable genes by Fourier analysis of genomic sequences. Comput Appl Biosci 1997;13:263–70.
[17] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–10.
[18] Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, et al. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res 2001;29:22–8.
[19] Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Res 2002;30:281–3.
[20] Godambe VP. An optimum property of regular maximum likelihood estimation. Ann Math Stat 1960;31:1208–12.
[21] Prentice RL, Zhao LP. Estimating equations for parameters in means and covariances of multivariate discrete continuous responses. Biometrics 1991;47:825–39.
[22] Mukhopadhyay NK, Saha AK, Lovelace JK, Da Silva R, Sacks DL, Glew RH. Comparison of protein kinase and acid phosphotase activitiesof five species of *Leishmania*. J Protozool 1988;35:601–7.
[23] Cool DE, Blum JJ. Protein tyrosine phosphatase activity in *Leishmania donovani*. Mol Cell Biochem 1993;127/128:143–9.
[24] Li S, Wilson ME, Donelson JE. *Leishmania chagasi*: a gene encoding a protein kinase with a catalytic domain structurally related to MAP kinase kinase. Exp Parasitol 1996;82:87–96.
[25] Flinn HM. Expression of a hydrophilic surface protein in infective stages of *Leishmania major*. Mol Biochem Parasitol 1994;65:259–70.
[26] Charest H, Matlashewski G. Developmental gene expression in *Leishmania donovani*: differential cloning and analysis of an amastigote-stage-specific gene. Mol Cell Biol 1994;14:2975–84.
[27] Marin M, Muskus C, Ramirez JR, Arbelaez LF, Alzate JF, Berberich C. The gene encoding the metacyclogenesis-associated transcript Mat-1 is conserved in the genus *Leishmania* and shows a tendency to form dimers upon protein expression. Parasitol Res 2000;86:431–5.
[28] Newton MA, Kendziorski CM, Richmond CS, Blattner FR, Tsui KW. On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data. J Comput Biol 2001;8:37–52.
[29] Pavlidis P, Noble WS. Analysis of strain and regional variation in gene expression in mouse brain. Genome Biol 2001;2:RESEARCH0042.
[30] Kerr MK, Churchill GA. Bootstrapping cluster analysis: assessing the reliability of conclusions from microarray experiments. Proc Natl Acad Sci USA 2001;98:8961–5.
[31] DeRisi JL, Iyer VR, Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 1997;278:680–6.
[32] Wu Y, El Fakhry Y, Sereno D, Tamar S, Papadopoulou B. A new developmentally regulated gene family in *Leishmania* amastigotes encoding a homolog of amastin surface proteins. Mol Biochem Parasitol 2000;110:345–57.
[33] Burchmore RJS, Landfear SM. Differential regulation of multiple glucose transporter genes in *Leishmania mexicana*. J Biol Chem 1998;272:29118–26.
[34] Beetham JK, Myung KS, McCoy JJ, Donelson JE. Glycoprotein 46 mRNA abundance is post-transcriptionally regulated during development of *Leishmania chagasi* promastigotes to an infectious form. J Biol Chem 1997;273:17360–6.

[35] Belli S. Chromatin remodelling during the life cycle of trypanosomatids. Int J Parasitol 2000;30:679–87.
[36] Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 1999;19:1720–30.
[37] Pavesi A. Relationships between transcriptional and translational control of gene expression in *Saccharomyces cerevisiae*: a multiple regression analysis. Mol Evol 1999;48:133–41.
[38] Lee ML, Kuo FC, Whitmore GA, Sklar J. Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. Proc Natl Acad Sci USA 2000;97:9834–9.
[39] Handman E, Osborn AH, Symons F, Van Driel R, Cappai R. The *Leishmania* promastigote surface antigen 2 complex is differentially expressed during the parasite life cycle. Mol Biochem Parasitol 1995;74:189–200.
[40] Rangarajan D, Gokool S, McCrossan MV, Smith DF. The gene B protein localises to the surface of *Leishmania major* parasites in the absence of metacyclic stage lipophosphoglycan. J Cell Sci 1995;108:3359–66.
[41] Almeida R, Norrish A, Levick M, Vetrie D, Freeman T, Vilo J, et al. From genomes to vaccines: *Leishmania* as a model. Philos Trans R Soc Lond B Biol Sci 2002;357:5–11.