From genomes to vaccines: *Leishmania* as a model

Renata Almeida¹, Alan Norrish¹, Mark Levick¹, David Vetrie², Tom Freeman²†, Jaak Vilo³, Alasdair Ivens², Uta Lange¹, Carmel Stober¹, Sharon McCann¹ and Jenefer M. Blackwell*¹

¹Cambridge Institute for Medical Research, University of Cambridge School of Clinical Medicine, Wellcome Trust/MRC Building, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2XY, UK
²The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
³EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK

The 35 Mb genome of *Leishmania* should be sequenced by late 2002. It contains approximately 8500 genes that will probably translate into more than 10 000 proteins. In the laboratory we have been piloting strategies to try to harness the power of the genome–proteome for rapid screening of new vaccine candidate. To this end, microarray analysis of 1094 unique genes identified using an EST analysis of 2091 cDNA clones from spliced leader libraries prepared from different developmental stages of *Leishmania* has been employed. The plan was to identify amastigote-expressed genes that could be used in high-throughput DNA-vaccine screens to identify potential new vaccine candidates. Despite the lack of transcriptional regulation that polycistronic transcription in *Leishmania* dictates, the data provide evidence for a high level of post-transcriptional regulation of RNA abundance during the developmental cycle of promastigotes in culture and in lesion-derived amastigotes of *Leishmania major*. This has provided 147 candidates from the 1094 unique genes that are specifically upregulated in amastigotes and are being used in vaccine studies. Using DNA vaccination, it was demonstrated that pooling strategies can work to identify protective vaccines, but it was found that some potentially protective antigens are masked by other disease-exacerbatory antigens in the pool. A total of 100 new vaccine candidates are currently being tested separately and in pools to extend this analysis, and to facilitate retrospective bioinformatic analysis to develop predictive algorithms for sequences that constitute potentially protective antigens. We are also working with other members of the *Leishmania* Genome Network to determine whether RNA expression determined by microarray analyses parallels expression at the protein level. We believe we are making good progress in developing strategies that will allow rapid translation of the sequence of *Leishmania* into potential interventions for disease control in humans.

Keywords: *Leishmania*; genome; vaccines

1. INTRODUCTION

Leishmaniasis remains a major public health problem throughout the tropics. The current outbreak of kala-azar or visceral leishmaniasis caused by *Leishmania donovani sensu strictu* in eastern and southern Sudan has taken its toll on an impoverished, war-stricken population, killing tens of thousands and depopulating vast areas of southern Sudan. The calamity brought kala-azar to the forefront as one of the greatest epidemics of the 20th century (Perea et al. 1989; De Beer et al. 1990). Post kala-azar dermal leishmaniasis following drug treatment causes additional problems, as do the dramatic and debilitating cutaneous forms of disease that occur in the tropical regions of both the Old and New Worlds. Even in sub-tropical regions where disease has not traditionally caused a major problem in humans, sub-clinical status is elevated to clinical disease in association with HIV infection (Rosenthal et al. 2000). There are no vaccines in routine use, and chemotherapy still relies principally on antimonial-based drugs first used in the early 20th century.

The 35 Mb genome of *Leishmania* should be sequenced by late 2002. It contains approximately 8500 genes that it is assumed will translate into more than 10 000 proteins. The question is: how rapidly can we harness the power of the genome–proteome to learn more about the biology of the host–parasite interaction and to develop new drugs and new vaccines? In our laboratory we have focussed on DNA-vaccine technology to develop higher throughput screens for potential new vaccine candidates. Initially the following two questions were considered.

(i) Could we narrow down the number of genes to be tested as vaccines by prescreening for genes expressed in forms of the parasite that invade and survive in the vertebrate host?

(ii) Could we test DNA vaccines in pools?
2. EXPRESSION PROFILING USING cDNA MICROARRAYS

The analysis of gene expression at the RNA level using microarray techniques can provide a global genetic perspective on biological processes important in parasite survival and host–parasite interactions. In Leishmania, most gene transcription is polycistronic. Transcript abundance measured using microarray techniques will therefore reflect post-transcriptional regulation of gene expression through processes including regulation of mRNA stability. The extent to which transcript abundance reflects expression at the protein level has yet to be determined. The first step was to determine whether significant differences in transcript abundance could be measured at the RNA level using microarray techniques.

In the sandfly vector, Leishmania undergoes developmental changes that lead to the generation of infective metacyclic forms transmitted to the vertebrate host. This can be mimicked in culture as promastigotes forms of the parasite pass through logarithmic to stationary phases of the growth cycle. To study the developmental changes in transcript abundance, total RNA was prepared from...
promastigotes at days 3, 5, 7, 10 and 13 of culture and from LAs for *L. major LV39*.

As a pilot microarrays were produced by spotting PCR-amplified inserts from 2091 randomly selected full-length cDNAs from spliced leader libraries prepared from different developmental stages (days 3, 7 and 10 cultured promastigotes *L. major LV39*; LAs from *L. major* Friedlin) (Coulson et al. 1996; Levick et al. 1996). These clones had already been used for EST one-pass 5' sequence analysis (Levick et al. 1996; R. Almeida, A. Norrish and J. M. Blackwell, unpublished data). A complete list of the EST sequences, and the cDNA libraries from which they were derived, can be obtained at http://www.ebi.ac.uk/parasites/leish.html. Sequence-cluster analysis shows that the 2091 genes represent 1094 unique transcripts, that is, there was around 50% redundancy in the clones selected for random-sequence analysis. For the clustering, a Blastn analysis of all *L. major* ESTs against the original set (Levick et al. 1996) in the dbEST database was performed. All hits with probabilities \(<1\times10^{-10}\) were initially ranked in descending order of quality. Sequences were then placed in 1094 ‘pools’ (some contain only one entry) on the basis of these data. *Leishmania* sequences with potentially interesting Blast similarities to constituents of the non-redundant nucleotide, dbEST or Swissprot protein sequence databases comprised 30% of the total EST dataset. These hits represent the ‘top’ non-self hit obtained for each sequence (the data have been parsed: probability cut-off of \(1e^{-05}\)). Out of these only 23% have significant matches to genes of known functions such that they can be assigned a putative function. Seventy percent of the clones have no sequence identity to any other sequences in the public databases. The 1094 unique genes represent approximately 12% of the total estimated number of *Leishmania* genes.

To screen the microarrays, a single round of reverse transcription was used to tag RNA transcripts for a particular developmental stage with Cy3. To compare changes in transcript abundance across different developmental stages tested on different microarray chips we had to decide on a control RNA that could be labelled with Cy5 and competitively hybridized with Cy3-labelled ‘test’ cDNA on every chip. In the initial experiment looking for expressed genes upregulated in expression in amastigotes (figure 1), control RNA was used from a single time-point: day 5 cultured promastigotes of *L. donovani* LV9. More recently, a pool of RNA across different developmental stages was used for a single outgroup ‘control’ species; for example, a pool of RNA from actively growing cultures from days 3, 5 and 7 for *L. amazonensis* has been used to compare changes in transcript abundance for different developmental stages across *L. donovani* and *L. major*. Full details of the results of these experiments will be presented elsewhere (Almeida et al., unpublished data). All of the data will be deposited in the public domain (http://www.ebi.ac.uk/parasites/leish.html) with links that will allow the data to be re-interrogated online at the European Bioinformatics Institute (http://www.ep.ebi.ac.uk/EP/EPCLUST7/). To compare changes in transcript abundance across developmental stages the data were analysed using both hierarchical and partition-based clustering algorithms (Brazma & Vilo 2000) (http://industry.ebi.ac.uk/~vilo/Expression-Profiler/). To understand how the clustering programs make such comparisons, it is useful initially to look at a scatter plot for Cy3 against Cy5 spot intensities for a single developmental stage within a single experiment. Since we want to identify amastigote-expressed genes for potential vaccine development, the scatter plot for RNA harvested from LA of *L. major LV39* (Cy3) against the *L. amazonensis* RNA pool (Cy5) is of direct relevance (figure 2). Each spot on the scatter plot represents a single cDNA clone. To declare a significant change in transcript abundance between the developmental stages, the ratio of intensity of the Cy3 ‘test’ RNA : Cy5 ‘control’ RNA needs to be greater than +2 or less than −2, that is, there needs to be more than a two-fold enhancement or reduction in transcript abundance comparing ‘test’ and ‘control’ RNAs. On the scatter plot this includes the spots above and below the dotted lines. It should be noted, however, that two different spots may have the same absolute fluorescence intensity on the Cy3 axis, but can be \(+2\) -fold and \(<-2\)-fold relative to the control pool.

![Figure 2. Scatter plot showing fluorescence intensity on the x- and y-axes for Cy5-labelled control RNA from the control pool against Cy3-labelled RNA from *L. major* amastigotes. The x- and y-axes have already been adjusted to eliminate background fluorescence, so every spot on the graph is expressed above the background. The individual points on the scatter plot represent the 2091 independent clones on the microarrays. The dotted lines above and below the ratio 1:1 line indicate the cut-offs for two-fold upregulation or two-fold downregulation relative to the control pool. Two dots are annotated to demonstrate that two clones may have the same absolute fluorescence intensity on the Cy3 axis, but can be \(+2\)-fold and \(<-2\)-fold relative to the control pool.](image-url)
expressed in amastigotes. The stringency of the cut-off above this can be determined according to user requirements; for example, in the initial targeting of amastigote-expressed genes as vaccine candidates we might choose to set a cut-off a log-fold above the background. However, since we cannot be sure that transcript abundance relates directly to protein abundance, it might be prudent to sample and test genes with both high- and low-transcript abundance.

To obtain an overall profile of the changes in transcript abundance for each gene throughout the developmental cycle we first used partition-based clustering asking the clustering algorithm to cluster our 2091 genes around 10 ‘seed’ clones that it chose as representative of the range of different expression profiles. Figure 1 shows this analysis for the experiment in which the different developmental stages for *L. major* LV39 were each compared against day 5 *L. donovani* promastigote RNA. The 10 seeds are shown in figure 1f. The program iteratively compares each of the seeds to its next closest matching expression profile, recalculating the cluster means or centres as it goes. The final cluster centres are shown in figure 1f. The resulting clusters are shown in figure 1a–i, j. The program iteratively compares each of the seeds to its next closest matching expression profile, recalculating the cluster means or centres as it goes. The final cluster centres are shown in figure 1f. The resulting clusters are shown in figure 1a–i, j. The profiles show log₂ ratios for fold-increases above and below the control RNA (log₂ of the ratio 1/0). The biggest change in expression relative to control for any genes within a particular genes within the cluster between day 5 promastigotes and amastigotes. The clusters in figure 1b, c, h, k are of particular interest because they contain clusters of genes that appear to be strongly upregulated in abundance in

Table 1. Identity of clones from clusters of genes upregulated in amastigotes of *L. major* relative to the day 5 promastigote control. ((a) The cluster depicted in figure 1e contains all the redundant copies of the tryparedoxin peroxidase gene; (b) the cluster depicted in figure 1k contains six out of seven redundant clones of the nucleoside diphosphate kinase B gene. The output of *Eclust* provides the clone name, accession number, and top matches and scores from BLAST analyses, permitting identification of clones corregulated with these genes.)

(a) tryparedoxin peroxidase cluster

<table>
<thead>
<tr>
<th>Clone</th>
<th>Accession</th>
<th>Score</th>
<th>Log2</th>
<th>Protein</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td>6381</td>
<td>AA728310</td>
<td>32</td>
<td>5.90E-12</td>
<td>TR:O61000</td>
<td><em>Crithidia fasciculata</em> tryparedoxin peroxidase</td>
</tr>
<tr>
<td>lm0f30</td>
<td>T67335</td>
<td>32</td>
<td>1.40E-16</td>
<td>TR:Q96673</td>
<td><em>Trypanosoma cruzi</em> tryparedoxin peroxidase</td>
</tr>
<tr>
<td>lm0h08</td>
<td>H64196</td>
<td>32</td>
<td>3.80E-40</td>
<td>TR:O61000</td>
<td><em>C. fasciculata</em> tryparedoxin peroxidase</td>
</tr>
<tr>
<td>lm0f30</td>
<td>T67335</td>
<td>32</td>
<td>1.40E-16</td>
<td>TR:Q96673</td>
<td><em>T. cruzi</em> tryparedoxin peroxidase</td>
</tr>
<tr>
<td>l085</td>
<td>AA680724</td>
<td>32</td>
<td>3.00E-45</td>
<td>SW:TDX</td>
<td><em>T. brucei</em> thioredoxin peroxidase</td>
</tr>
<tr>
<td>lm370</td>
<td>T93440</td>
<td>145</td>
<td>1.10E-65</td>
<td>SW:METK</td>
<td><em>Leishmania infantum</em> S-adenosylmethionine synthetase</td>
</tr>
<tr>
<td>lm5k08</td>
<td>W88371</td>
<td>56</td>
<td>2.00E-32</td>
<td>TR:Q9N6Y7</td>
<td><em>L. major</em> probable nucleoside diphosphate kinase B</td>
</tr>
<tr>
<td>s033</td>
<td>AA680974</td>
<td>55</td>
<td>1.10E-25</td>
<td>SW:IF51</td>
<td>SCHPO initiation factor EF-5A-1</td>
</tr>
<tr>
<td>m007</td>
<td>T67314</td>
<td>294</td>
<td>3.90E-11</td>
<td>SW:H2A</td>
<td><em>P. falciparum</em> histone H2A</td>
</tr>
<tr>
<td>lm0k14</td>
<td>W88375</td>
<td>218</td>
<td>6.50E-31</td>
<td>TR:Q99658</td>
<td><em>L. major</em> possible ubiquitin fusion protein</td>
</tr>
<tr>
<td>s515</td>
<td>AA728294</td>
<td>34</td>
<td>1.10E-44</td>
<td>TR:P90633</td>
<td><em>Leishmania braziliensis</em> elongation factor EF-1A</td>
</tr>
<tr>
<td>581m</td>
<td>AA728297</td>
<td>187</td>
<td>3.40E-54</td>
<td>TR:Q9NE87</td>
<td><em>L. major</em> hypothetical L3640.07</td>
</tr>
<tr>
<td>lm6k27</td>
<td>W883895</td>
<td>162</td>
<td>2.90E-21</td>
<td>TR:Q96938</td>
<td><em>Drosophila melanogaster</em> CG3203 protein</td>
</tr>
<tr>
<td>lm393</td>
<td>T93450</td>
<td>311</td>
<td>3.50E-55</td>
<td>TR:Q9W3W8</td>
<td><em>T. brucei</em> embryoid body precursor</td>
</tr>
<tr>
<td>362d</td>
<td>AA780465</td>
<td>276</td>
<td>6.10E-20</td>
<td>TR:Q97213</td>
<td><em>L. major</em> L830.9 protein</td>
</tr>
</tbody>
</table>

(b) nucleoside diphosphate kinase B cluster

<table>
<thead>
<tr>
<th>Clone</th>
<th>Accession</th>
<th>Score</th>
<th>Log2</th>
<th>Protein</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td>lm4k01</td>
<td>AA060749</td>
<td>56</td>
<td>2.20E-57</td>
<td>TR:Q9N6Y7</td>
<td><em>L. major</em> nucleoside diphosphate kinase B</td>
</tr>
<tr>
<td>168c</td>
<td>A0784527</td>
<td>56</td>
<td>1.30E-47</td>
<td>TR:Q9N6Y7</td>
<td><em>L. major</em> nucleoside diphosphate kinase B</td>
</tr>
<tr>
<td>lme064</td>
<td>T67339</td>
<td>56</td>
<td>7.70E-57</td>
<td>TR:Q9N6Y7</td>
<td><em>L. major</em> nucleoside diphosphate kinase B</td>
</tr>
<tr>
<td>664c</td>
<td>A0784534</td>
<td>56</td>
<td>6.60E-103</td>
<td>TR:Q9U1E1</td>
<td><em>L. major</em> nucleoside diphosphate kinase B</td>
</tr>
<tr>
<td>q972</td>
<td>AA680908</td>
<td>56</td>
<td>4.40E-70</td>
<td>TR:Q9N6Y7</td>
<td><em>L. major</em> nucleoside diphosphate kinase B</td>
</tr>
<tr>
<td>6531</td>
<td>AA728315</td>
<td>56</td>
<td>2.30E-26</td>
<td>TR:Q9N6Y7</td>
<td><em>L. major</em> nucleoside diphosphate kinase B</td>
</tr>
<tr>
<td>438c</td>
<td>M034692</td>
<td>82</td>
<td>2.20E-71</td>
<td>TR:Q9NK14</td>
<td><em>L. major</em> mitotic centromere associated kinesin 2</td>
</tr>
<tr>
<td>214a</td>
<td>AA741623</td>
<td>92</td>
<td>1.10E-25</td>
<td>SW:H4</td>
<td><em>Chlamydomonas reinhardtii</em> histone H4</td>
</tr>
<tr>
<td>292d</td>
<td>A0784608</td>
<td>92</td>
<td>1.60E-17</td>
<td>TR:O62586</td>
<td><em>Blepharisma undulans</em> histone H4</td>
</tr>
<tr>
<td>6831</td>
<td>AA728322</td>
<td>92</td>
<td>7.70E-13</td>
<td>TR:Q9U9A5</td>
<td><em>Leishmania tarentolae</em> histone H4</td>
</tr>
</tbody>
</table>

4 x 60S ribosomal proteins
1 x other ribosomal proteins
7 x unknowns

Phil. Trans. R. Soc. Lond. B (2002)
the amastigote stage of the parasite life cycle, and hence might be good vaccine candidates. Looking within these clusters, it can be seen that all redundant clones of the trypanoxdoxin peroxidase gene fall within cluster 1 (table 1a), while six out of seven redundant clones of the nucleoside diphosphate kinase B gene fall within cluster 1i (table 1b). This is reassuring, since it demonstrates remarkable within-experiment reproducibility that extends over all redundant sets of clones within all of our experiments (data not shown). We can then begin to ask questions about what other genes that might be related in function and also be performed to determine which of these genes responds to changes in experimentally induced oxidative stress, thus helping to delimit which genes might contribute to antioxidant pathways.

3. FROM GENES TO VACCINES

The most direct route to taking our putative amastigote-expressed genes into vaccine testing was to employ a DNA-vaccine strategy. DNA vaccination has already
proved successful in vaccinating mice against cutaneous
L. major infection (Gurunathan et al. 1997, 1998; Hand-
man et al. 2000; Fragaki et al. 2001). A number of experi-
ments (Norrish et al., unpublished data) have been carried
out in our laboratory to refine the DNA-vaccine strategy,
in particular to see whether it was possible to reduce both
the number and dose of vaccines given. This would facil-
itate a higher throughput of large numbers of potential vac-
cine candidates in a murine screen, and allow us to think
about pooling strategies to try to identify new vaccine can-
didates more efficiently. Figure 3 shows the results of one
such experiment. Figure 3a shows the normal profile for
L. major LV39 infection in susceptible BALB/c compared
with genetically resistant CBA/Ca mice. Figure 3b shows
that vaccination with 100 µg vector DNA alone mimics
the BALB/c infection-only profile, whereas vaccination
with a single dose of 25 µg DNA of a protective antigen
lmd29 (plus 75 µg vector-only DNA) administered s.c.
6 weeks prior to challenge infection provides significant
protection. Figure 3c depicts the infection profiles for eight
pools of 4 × 25 µg DNA for 16 unknown test vac-
cines from a 4 × 4 array. Two vaccine pools show some
protection relative to the vector-only control; the remain-
ing six pools gave the same infection profile, or showed
exacerbation of disease, relative to the vector-only control.
Figure 3d shows that the vaccine at the intersec of the
two protective pools, designated R61, is protective when
administered on its own. This strategy was therefore suc-
cessful in identifying a new vaccine candidate from the 16
test vaccines in the pools. However, it was also found that
protection was lost when 25 µg lmd29 was administered
in a pool with 25 µg each for three members of one of the
non-protective pools (data not shown). Hence, there may
be antigens within pools that exacerbate disease and mask
the potential protective effect of other members of the
pool. In current experiments this analysis is being
extended by testing the 20 pools obtained in a 10 × 10
array, alongside each of the 100 vaccine candidates tested
separately. One advantage of testing all the vaccine candi-
dates separately is that, with the full sequence of all the
vaccine candidates to hand, it will be possible to perform
a retrospective bioinformatic analysis to see if a predictive
algorithm can be devised for what constitutes a successful
vaccine candidate for leishmaniasis.

4. CONCLUSIONS

In the laboratory we have been piloting experiments
designed to see whether it is possible to develop global
approaches to identify amastigote-expressed genes and
carry out large-scale vaccine testing directly utilizing the
output of the genome project. The findings thus far have
been encouraging, but there is much work ahead to har-
ness the full power of the genome–proteome. In relation to
microarray analysis, it is necessary to determine whether
expression profiling at the RNA level bears any relation to
expression of the protein encoded by the RNA. To exam-
ine this we are working with other members of the Leish-
mania Genome Network to carry out parallel analyses of
developmental changes in the transcriptome and the prote-
ome using material harvested from the same cultures of
parasites. We are also taking some of our protective anti-
gens to see if we can improve vaccine efficiency using a
DNA-vaccine prime and protein boost strategy (McShane
et al. 2001) employing attenuated Salmonella typhimurium
or Vaccinia virus as the protein-producing vaccine. Clearly
the real challenge will be to see how rapidly we can pro-
cceed from animal experimentation to phase I and II trials
in humans, thus proving the real worth that knowledge of
parasite genomes will probably prove to be in relation to
human health.

Laboratory work on parasite genome analysis and vaccine
development was supported by Glaxo Group Research Ltd,
the UNDP–World Bank–World Health Organization Special
Programme for Research and Training in Tropical Diseases
(TDR) the Commonwealth Scholarship Commission in the
UK, and the British Medical Research Council.

REFERENCES

Alphhey, M. S., Bond, C. S., Tetaud, E., Fairlamb, A. H. &
Hunter, W. N. 2000 The structure of reduced tryptophan
peroxidase reveals a decamer and insight into reactivity of
Differential expression of Leishmania major beta-tubulin
genes during the acquisition of promastigote infectivity.
Mol. Biochem. Parasitol. 82, 227–236.
De Beer, P., Harith, A. E., Van Groothuyst, M. & Winkler, A.
1990 Outbreak of kala-azar in the Sudan. Lancet 335,
224.
de la Rosa, A., Williams, R. L. & Steeg, P. S. 1995
Nm23/nucleoside diphosphate kinase: toward a structural
and biochemical understanding of its biological functions.
Bioessays 17, 53–62.
European Bioinformatics Institute http://www.ep.epi.ac.uk/
EP/EPCLUST/
Fragaki, K., Sufiia, I., Ferrua, B., Rousseau, D., Le Fichoux,
Y. & Kubar, J. 2001 Immunisation with DNA encoding
Leishmania infantum protein papLe22 decreases the fre-
quency of parasitic episodes in infected hamsters. Vaccine
19, 1701–1709.
Gale, M., Carter, V. & Parsons, M. 1994 Cell-cycle-specific
induction of an 89 kDa serine/threonine protein kinase
Nucleoside diphosphate kinase from human erythrocytes.
Structural characterisation of the two polypeptide chains
responsible for heterogeneity of the hexameric enzyme.
J. Biol. Chem. 266, 8784–8789.
Vaccine requirements for sustained cellular immunity to an
Gurunathan, S., Sacks, D. L., Brown, D. R., Reiner, S. L.,
Charest, H., Gleichenhau, N. & Seder, R. A. 1997 Vaccin-
ination with DNA encoding the immunodominant LACK
parasite antigen confers protective immunity to mice infected
Handman, E., Nooromahmadi, A. H., Curtis, J. M.,
Baldwin, T. & Sjolander, A. 2000 Therapy of murine
cutaneous leishmaniasis by DNA vaccination. Vaccine
18, 3011–3017.
http://industry.epi.ac.uk/~vilo/Expression-Profiler/
http://www.epi.ac.uk/parasites/leish.html
Lees-Miller, S. P. 1996 The DNA-dependent protein kinase,
74, 503–512.
An expressed sequence tag analysis of a full-length, spliced-leader cDNA library from *Leishmania major* promastigotes.


