THE ROYAL SOCIETY

Published online 17 January 2002

Putting the *Leishmania* genome to work: functional genomics by transposon trapping and expression profiling

Stephen M. Beverley[1*], Natalia S. Akopyants[1], Sophie Goyard[1], Robin S. Matlib[1], Jennifer L. Gordon[1], Bernard H. Brownstein[1,3], Gary D. Stormo[2], Elena N. Bukanova[1], Christian T. Hott[3], Fugen Li[2], Sandra MacMillan[3], James N. Muo[1], Libbey A. Schwertman[3], Matthew R. Smeds[1] and Yujia Wang[3]

[1]*Departments of Molecular Microbiology,* [2]*Genetics, and the* [3]*WUMS DNA Microarray Facility, Washington University Medical School, 660 S Euclid Avenue, Box 8230, St Louis, MO 63110, USA*

Leishmania are important protozoan pathogens of humans in temperate and tropical regions. The study of gene expression during the infectious cycle, in mutants or after environmental or chemical stimuli, is a powerful approach towards understanding parasite virulence and the development of control measures. Like other trypanosomatids, *Leishmania* gene expression is mediated by a polycistronic transcriptional process that places increased emphasis on post-transcriptional regulatory mechanisms including RNA processing and protein translation. With the impending completion of the *Leishmania* genome, global approaches surveying mRNA and protein expression are now feasible. Our laboratory has developed the *Drosophila* transposon *mariner* as a tool for trapping *Leishmania* genes and studying their regulation in the form of protein fusions; a classic approach in other microbes that can be termed 'proteogenomics'. Similarly, we have developed reagents and approaches for the creation of DNA microarrays, which permit the measurement of RNA abundance across the parasite genome. Progress in these areas promises to greatly increase our understanding of global mechanisms of gene regulation at both mRNA and protein levels, and to lead to the identification of many candidate genes involved in virulence.

Keywords: DNA microarrays; virulence; proteogenomics; trypanosomatid protozoan parasite; gene fusions; *mariner*

1. INTRODUCTION

Leishmania is an important parasite of humans, infecting upwards of 12 million people in tropical and temperate regions of the world. These trypanosomatid protozoans have an obligate digenetic life cycle, alternating between the flagellated promastigote form residing in the gut of the insect vector sandfly, and the intracellular amastigote stage residing within an active phagolysosome of vertebrate macrophages. How *Leishmania* carries out these developmental transitions, and the mechanisms that they employ in surviving within the host and resisting a tremendously hostile array of defenses, are key questions of interest to biologists and clinicians seeking to control these pathogens.

In the past 12 years a variety of genetic tools have been introduced that now permit manipulation of the *Leishmania* genome with a high degree of specificity (Swindle & Tait 1996; Clayton 1999). These include a variety of expression vectors (circular and linear episomes, integrating, regulated and constitutive), methods for gene replacement and the generation of null mutants, transposon mutagenesis and functional genetic rescue, amongst others (figure 1). These methods constitute a powerful genetic 'toolkit', allowing experimenters to take genes identified by various routes and probe their function by both gain and loss-of-function strategies, as well as localization using a variety of tags such as the GFP. One area of particular interest involves the assessment of genes involved in the synthesis of molecules potentially implicated in *Leishmania* virulence and transmission, such as LPG, cysteine proteases and GPI-anchored proteins, by studying knock-out mutants (Turco *et al.* 2001). In these studies a key experimental control was restoration of the gene and wild-type function to the null mutants, one of the central tenets of Falkow's 'molecular Koch's postulates' (Falkow 1988). This is especially important in *Leishmania* as these species have a tendency to spontaneously lose virulence during *in vitro* culture by processes unrelated to the planned mutations.

Complementing our ability to carry out reverse genetic manipulations has been the development of methods for functional genetic rescue, by mass transfection of *Leishmania* mutants or variants with genomic DNA libraries carried in *Leishmania–Escherichia coli* shuttle vectors such as cLHYG, followed by selection for appropriate phenotypes (Beverley & Turco 1998). These methods have

* Author for correspondence (beverley@borcim.wustl.edu).

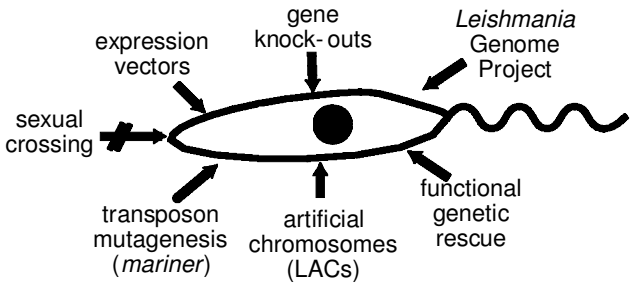

Figure 1. Genetic and genomic tools now available for the study of *Leishmania*.

identified genes important in key parasite processes such as glycoconjugate biosynthesis, peroxisome biogenesis and drug resistance. Such 'forward genetic' methods have the advantage that they rely solely on phenotype, making it probable that the genes obtained will be directly involved in the process under study, often with unanticipated roles or functions that would not have been detected by motif searches or database comparisons.

Unfortunately, the major hurdle for forward genetic approaches in *Leishmania* is that experimentally, these organisms are asexual diploids (Panton *et al.* 1991). While this is easily overcome when making null mutants of specific genes (with two rounds of genetic manipulation), it decreases the frequency of loss-of-function mutations following random mutagenesis to $<10^{-6}$ (Gueiros-Filho & Beverley 1996). Thus when powerful screens or selections are available forward genetics methods are quite valuable, but this requirement limits the widespread application of forward genetic methods presently.

2. THE *LEISHMANIA* GENOME PROJECT

The *Leishmania* genome comprises about 34 Mb and is estimated to encode about 10 000 genes (Blackwell & Melville 1999; Myler & Stuart 2000). An international consortium of researchers is determining the genome sequence of a prototypic species, *Leishmania major*, and completion is anticipated in 2–3 years' time (Ivens & Smith 1997; Myler & Stuart 2000). Already, more than 5 Mb of finished and 20 Mb of unfinished sequence is available, allowing researchers to scan for genes of interest by a variety of approaches. The major challenge now is to develop experimental methods for efficiently searching within this mass of information for genes that play important roles within the parasite. This task and its newly arisen field are often referred to as 'functional genomics'.

One approach well suited for genomic strategies involves the identification of genes whose expression changes during growth or development, or in response to perturbations arising from environmental stimuli, drugs or genetic mutations. This traditionally has been undertaken in a gene-by-gene approach, however methods are now available for performing this on a genome-wide scale. Changes in gene expression can be monitored at the mRNA or protein level. Each of these has advantages: nucleic acid methodologies are robust and easily applied and scaled up, whilst protein methods focus directly on the molecular agents that carry out functional roles. In *Leishmania* and trypanosomes, we know that gene regulation can occur at both the level of mRNA and protein abundance (Graham 1995). An unanswered question is how their reliance on polycistronic transcription and *trans*-splicing to generate mature monocistronic mRNAs (Ullu *et al.* 1996) impacts on the global regulatory mechanisms. This unusual transcriptional mechanism potentially mitigates the need for traditional RNA polymerase II promoters, and potentially increases the reliance on post-transcriptional methods such as processing, RNA stability, translation and protein degradation.

3. mRNA-EXPRESSION PROFILING: A GENOME-WIDE SURVEY OF *LEISHMANIA* GENE EXPRESSION

A powerful method termed 'expression profiling' was introduced in the mid-1990s by Brown and others (Schena *et al.* 1995; Brown & Botstein 1999). With this method, DNAs specific for upwards of 10 000 genes are deposited or synthesized directly onto small slides (yielding 'microarrays'), and then hybridized specifically with fluorescently labelled cDNAs derived from mRNA isolated from organisms or tissues of interest. Quantitative imaging and data processing then allows direct visualization of the mRNA levels. Often, two mRNA sources labelled with different fluors are used simultaneously, and the hybridization ratio is scored. This approach has some advantages since the hybridization conditions for the two probes are identical, making the hybridization ratio an experimentally robust measure of changes in gene expression.

One of the challenges in many parasite systems is obtaining sufficient material for molecular and biochemical analysis. For *Leishmania*, the insect-stage promastigote form can be readily grown in simple culture media. Moreover, as the non-infective logarithmic-phase promastigotes enter the stationary phase, they differentiate into the highly infective metacyclic stage transmitted by the sandfly (Sacks & Perkins 1985) that can be readily purified. For many species of *Leishmania*, methods enabling growth of the amastigote stage axenically are now available, and quantities of parasites may also be harvested from lesions obtained in infected mice. Thus, obtaining sufficient quantities of *Leishmania* RNA from well-defined stages throughout the infectious cycle is relatively straightforward.

With many organisms, expression profiling begins with the construction of a complete set of DNAs identifying every gene or ORF within the genome. Unfortunately, when we began these studies very little of the *Leishmania* genome had been sequenced. As an alternative, we generated a 'shotgun' DNA clone library comprising 10 000 inserts 1–1.5 kb in size obtained from randomly sheared genomic DNA of the Friedlin V1 line of *L. major* (Akopyants *et al.* 2001). To determine the diversity and composition of this DNA collection, we undertook end sequencing and deposited about 10 000 sequences (encompassing more than 4 Mb) in the genome survey sequence section of GenBank (hence we typically refer to this clone collection as our GSS set). These sequences have been a useful resource for gene discovery for many workers.

In Northern blot or microarray-based expression profiling, it is only necessary that the probe (in this case the random shotgun GSS clones) overlap enough of the transcript to hybridize well. The use of 1 kb random probes is especially well suited for trypanosomatid parasites, as the gene density is high, intergenic regions are typically <400 nt, and introns are rare (Akopyants *et al.* 2001). We used several approaches to assess the ability of the GSS to identify *Leishmania* transcripts, that typically average 3.4 kb in length (Myler *et al.* 1999). For example, comparisons against the sequences of *L. major* chromosome 1 and 3 (Myler *et al.* 1999; P. Myler, unpublished data) suggested that one or more of the GSS clones were able to identify at least 63% of known *Leishmania* ORFS or proteins. Other studies suggest that this collection represents about 8000 genes in total (Akopyants *et al.* 2001). Thus, the random shotgun GSS approach provides a relatively inexpensive and rapid method for generating DNA collections suitable for microarray creation, with a reasonably high coverage of the genome. This approach has also been taken in studies of *Trypanosoma brucei* and *Plasmodium falciparum* (El-Sayed *et al.* 2000; Hayward *et al.* 2000). Of course, upon completion of the genome sequence we anticipate the formation of a consortium to generate a specific, complete collection of DNAs corresponding to each *Leishmania* ORF.

The GSS collection forms the core set of DNAs presently under study via expression profiling in our laboratory. Additionally, in these studies we synthesized by PCR DNAs from hundreds of other genes, including ones known to be regulated by Northern blot analysis, all ORFs from chromosome 1, genes of specific interest to our laboratory and various controls. In combination with the 10 000 GSS shotgun clones, this set of approximately 11 000 DNAs has been used to generate a preliminary '11K' microarray. As hybridization probes, we have focused primarily on three sources of mRNAs of *L. major*: log-phase promastigotes, metacyclic promastigotes and lesion amastigotes. Statistical methods and analysis were used to apply the results emerging from these studies to the task of identifying genes whose expression changes at the level of mRNA abundance. For the results described below, every comparison has been performed in triplicate and the genes identified were scored as regulated by at least a factor of two in all three experiments.

Our preliminary results are encouraging. In comparisons of log phase versus metacyclic promastigotes, many genes known to be upregulated in one stage or the other were identified. In log phase this includes β-tubulin, histones and ribosomal proteins, while in metacyclics this includes *HASP*/geneB, *SHERP*/geneD, *META1* and *HSP70*. Notably, the largest number of genes scored as regulated did not show hits to proteins of known function in database searches. Similar results were obtained in comparisons of log-phase promastigotes versus amastigotes. An important part of expression profiling is validation of the microarray results by methods such as Northern blot analysis or quantitative PCR. To date 47 genes have been tested with Northern blot analysis, with 69% showing regulation as predicted.

In these studies we noticed that relatively few genes (only a few per cent) showed changes in expression of more than two-fold between the stages examined. In contrast, analogous growth or developmental transitions in the yeast *Saccharomyces cerevisiae* (log to stationary phase, sporulation) were accompanied by two-fold or greater changes in expression in 18–27% of all genes (DeRisi *et al.* 1997; Chu *et al.* 1998). Why are these numbers so different? One possibility is that in our preliminary studies, we have not detected changes in regulation efficiently by expression profiling. In fact, analyses of the results obtained with control *Leishmania* genes known to be regulated at the mRNA level suggest that some genes were missed. However, this appears unlikely to account for the more than 10-fold difference noted between *Leishmania* and yeast.

Another possibility is that this finding is correct. A number of workers have carried out methods searching for changes in gene expression amongst *Leishmania* stages, using methods such as differential or subtractive hybridization and differential display (Coulson & Smith 1990; Charest & Matlashewski 1994; Pogue *et al.* 1995; Heard *et al.* 1996; Liu *et al.* 2000; Wu *et al.* 2000). While a number of genes showing significant regulation by transcript abundance were found, most workers remarked that they were able to identify only a relatively small number of differentially regulated genes, in agreement with our preliminary microarray results. If this finding holds up, it suggests that for some reason, *Leishmania* relies less upon changes in mRNA abundance due to control-gene regulation, presumably relying more heavily upon regulatory mechanisms at the protein level. Notably, in many organisms, it has become evident that the correlation between mRNA and protein abundance is weak (Gygi *et al.* 1999), and there are many examples of regulation occurring at the protein level rather than mRNA abundance in *Leishmania* and trypanosomes. Indeed, this may be one of the unexpected consequences of the polycistronic transcriptional route for gene expression in trypanosomatid protozoans. We stress that our findings here are preliminary, and studies are now underway in the laboratory to confirm them.

4. PROTEOMICS, TRANSPOSON TRAPPING AND 'PROTEOGENOMICS'

It is widely recognized that to understand organisms we must understand not only the organization of genes and mRNA expression patterns, but also the abundance, modifications and interactions of the encoded proteins. The field of 'proteomics' offers some significant challenges however, as unlike nucleic acids, proteins are relatively fickle and individualistic in their experimental needs. For example, in two-dimensional gels not every protein resolves well, and often each protein 'spot' must be characterized individually by various means.

Curiously, one approach not commonly included in 'proteomics' but used widely by bacteriologists is the use of gene fusions to study regulation at the protein level (Berg *et al.* 1989). In this approach, libraries of genes fused to a convenient reporter protein (such as β-galactosidase or GFP) are generated, and then scored for expression. Fusion libraries can be made by a number of methods, although for various reasons the use of transposons in bacteria *in vivo* or by shuttle mutagenesis *in vitro* have received greatest attention (figure 2). In some eukaryotes it has also been possible to perform transposon

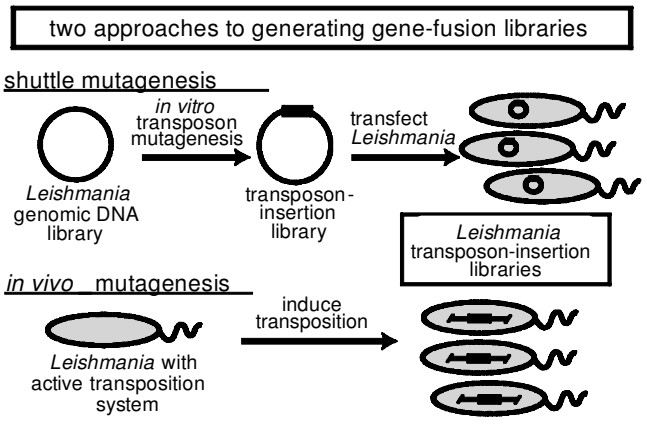

Figure 2. Two strategies for the generation of transposon-mutagenesis libraries in *Leishmania*.

mutagenesis directly *in vivo* with natural or engineered transposons, as in yeast, *Drosophila* and *Caenorhabditis elegans*. Notably, while most gene-fusion libraries emphasize regulation at the mRNA or 'transcriptional' level, one can design fusions that only work at the protein or 'translational' level. Such fusions are ideal for systematically studying expression broadly across the entire 'proteome'. Since bacteriologists began their elegant studies prior to the current 'omics' era, we will refer to the use of libraries of 'translational' gene fusions to study regulation of protein expression as 'proteogenomics'.

The generation of libraries of gene fusions in *Leishmania* poses some challenges. When DNA constructs are introduced into *Leishmania*, they appear to take one of two paths: they may circularize (if not already circular) and replicate autonomously as episomes; or they may integrate into homologous chromosomal loci. Thus far, there have been no examples of non-homologous recombination following transfection of DNA into *Leishmania* (one anticipates that as in *S. cerevisiae*, non-homologous integration (HI) does occur but at frequencies much lower than the homologous route). Thus, one cannot generate gene-fusion libraries *in vivo* by simply transfecting parasites with reporter DNA cassettes, and identifying fusions arising from numerous non-HI events.

Transposon methodologies provide a simple way of generating gene fusions independent of host recombinational pathways. We have shown that it is possible to import the transposon *mariner* from *Drosophila* into *Leishmania* (Gueiros-Filho & Beverley 1997). *Mariner* is a member of a large family of small transposable elements (Tc1/*mariner*) consisting of a transposase gene flanked by inverted repeats (Hartl *et al.* 1997; Plasterk *et al.* 1999). Using parasites, transposase was expressed from a standard *Leishmania* expression vector, where it mediated transposition of suitably engineered transposons from episomal vectors into a variety of chromosomal loci. The utility of this approach in 'gene trapping' was demonstrated by use of an engineered transposon bearing a silent hygromycin B resistance cassette *HYG* lacking a splice acceptor; following transposition and selection for drug resistance, several transposition events were recovered showing activation of *HYG* expression through transposon insertion downstream of a *Leishmania* splice acceptor. In its current incarnation this system has not as yet proven amenable to the generation of large-scale fusion libraries, however, as inter-plasmid events come to dominate the libraries and constitutive expression of transposase raises concerns about secondary events. These are manageable problems and their resolution is currently under study.

As an alternative we have taken a 'shuttle' mutagenesis approach to generate gene-fusion libraries (figure 2). In this approach, transpositions are obtained in large segments of *Leishmania* genomic DNA carried in the shuttle vector cLHYG, and then introduced into *Leishmania* by transfection. The advantages of this approach are that transposition can be readily controlled and is highly efficient, while the disadvantage is that one is limited by the transfection efficiency. In *Leishmania*, the stable transfection efficiency is a respectable value (10^{-4} or better) and it has been possible to generate large libraries (upwards of 50 000 independent cosmid transfectants), making this a feasible approach.

Shuttle mutagenesis can now be performed with a number of available transposon systems (such as Tn5 or Ty1) that act in *E. coli in vivo* or *in vitro*. However, there are some advantages in utilizing transposons such as the Mos1 *mariner* element that can be mobilized both *in vitro* (for shuttle mutagenesis) and *in vivo* directly within *Leishmania*. The Mos1 *mariner* transposase was purified to homogeneity, and an *in vitro* transposition system was developed (Tosi & Beverley 2000). This transposase worked efficiently, with frequencies up to 10^{-3} per DNA. This enabled us to establish the *cis*- and *trans*-acting requirements for optimal transposition; for example (and somewhat unexpectedly), we showed that the 28 bp inverted *mariner* repeats were insufficient in themselves for transposition, and that internal sequences were required. In contrast to the hornfly *mariner* element *Himar1*, where excess transposase is inhibitory, for *Mos1* transposase transposition was simply proportional to transposase concentration. This suggested that when seeking to maximize transposition of Mos1 *mariner* elements *in vivo*, one should strive for maximal transposase expression. The *in vitro* system additionally enabled us to test new candidate transposons for activity prior to more laborious tests *in vivo* (Tosi & Beverley 2000; Goyard *et al.* 2001).

We have now created a variety of new, versatile and highly active *mariner* derivatives, carried on donor plasmids suitable for use in the *in vitro* transposition system and containing a variety of genetic elements of potential interest such as reporter genes, selectable markers, origins of replication, and the like (Goyard *et al.* 2001). These enable one to readily recover *mariner* insertions into genomic DNA by shuttling back to *E. coli*, to sequence DNA rapidly, and to generate gene fusion to both selectable markers and reporter proteins. Included in this *mariner* transposon 'toolkit' are several transposons designed to selectively identify translational gene fusion (figure 3). One example is the transposon /GEP3*, that contains a GFP joined to an *E. coli* promoter driving a PHLEO resistance marker (*PHLEO*) in a way that yields a fusion protein spanning both GFP, promoter and PHLEO. The GFP cassette lacks an initiating ATG codon, and one is not provided by the mariner inverted repeat; thus, in order to obtain GFP–PHLEO expression, transposition into a *Leishmania* ORF providing both an upstream *trans*-splice acceptor site and start codon is required. Similarly, the

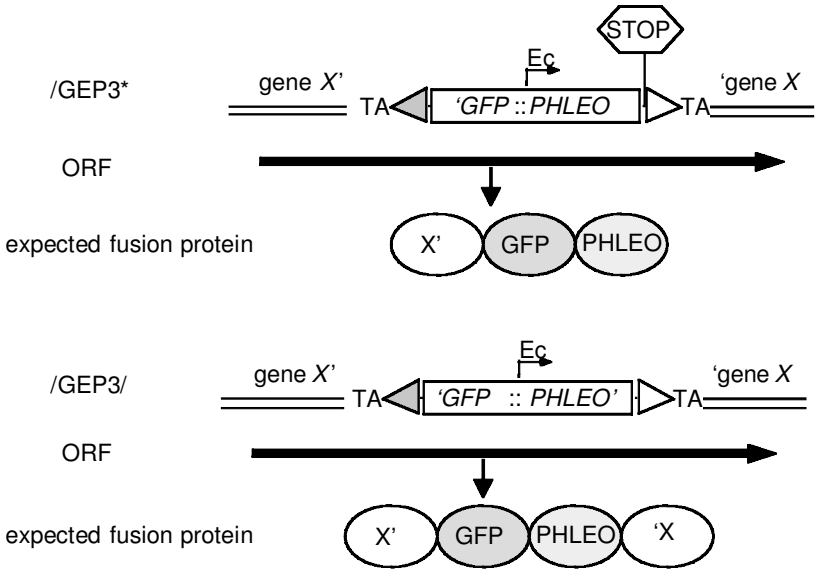

Figure 3. Two examples of Mos1-*mariner* derived transposons able to trap *Leishmania* protein fusions.

transposon /GEP3/ contains the same GFP–*E.coli* promoter–PHLEO protein fusion, but in this case the PHLEO termination codon has been altered in a way that yields an in-frame fusion protein spanning the *mariner* inverted repeats and the internal GFP–PHLEO protein (figure 3). Again, this protein can only be expressed upon insertion into a *Leishmania* ORF. The advantage of transposon /GEP3/ is that unlike /GEP3*, the resultant fusion protein retains both N and C terminal regions of the protein, potentially containing information for proper cellular targeting associated with both domains.

To test the feasibility of the shuttle mutagenesis system, we chose a *Leishmania* cosmid bearing a 40 kb of genomic DNA arising from the H region (Beverley *et al.* 1988). This region of DNA is frequently amplified in drug-resistant parasites and encodes at least two genes implicated in drug resistance: *PGPA*, a member of the MRP family of P-glycoproteins responsible for resistance to arsenite and animonials, and PTR1, a broad spectrum pteridine reductase implicated in antifolate resistance (Callahan & Beverley 1991; Bello *et al.* 1994; Ouellette *et al.* 1996). Furthermore, it has been suggested that other genes within the H region may be implicated in resistance to other compounds such as primaquine and terbinafine (Ellenberger & Beverley 1989). Since the sequence of this region is available in GenBank, it offered a good venue in which to test shuttle mutagenesis as well as to probe the role of other H-region genes.

In our shuttle mutagenesis protocol, we first created large pools of transpositions of /GEP3* or /GEP3/ into the H-region cosmid. These pools were transfected into *Leishmania* and selected for resistance to hygromycin B, the vector marker present in cLHYG. These *Leishmania* transfectant pools were then selected for resistance to phleomycin (PHLEO+) or screened for expression of GFP by flow cytometry (GFP+). In these studies we determined the optimal numbers for the cosmid and transfectant pools, developed methods for recovery of the transposon-bearing cosmids from *Leishmania*, and visualized the presence of the predicted fusion proteins by Western blotting with anti-GFP antisera. Little difference was found between the two transposons in terms of their ability to identify *Leishmania* ORFs.

We were rapidly able to identify transpositions into seven of the ten ORFs present in the H-region cosmid, but despite repeated further attempts, transpositions into the remaining three ORFs were not recovered as gene fusions. Possibly these three ORFs are 'cold spots' for *mariner* insertion, although previous work with *mariner in vitro* and *in vivo* suggests there is little target specificity. We tested this with one of the 'missing' ORFs (PTR1) and showed that there were insertions present in the primary transposition pool, suggesting this was not the cause. Alternatively, it may be that the missing ORFs are expressed at a low level, or that they do not encode *bona fide* parasite ORFs. Again for PTR1, we know this explanation is incorrect, since this protein is expressed by *Leishmania* and we have through other approaches engineered active PTR1–GFP–PHLEO fusions. Thus transposon mutagenesis may not be able to recover every candidate *Leishmania* ORF. Nonetheless, from these and other studies it is clear that transposon mutagenesis is a rapid way of identifying most *Leishmania* ORFs and generating reporter gene fusion, with far less effort than required in traditional gene-by-gene methods.

Currently the regulation of the GFP fusion proteins during *Leishmania* development is under examination. In these studies we selected and screened for expression specifically in log-phase promastigotes and so we set this as the reference point. Notably, two out of eight GFP–PHLEO fusions tested showed at least 10-fold down-regulation in stationary-phase promastigotes. Thus this approach allows monitoring of regulation at the protein level as anticipated. In the future, it is intended to expand these studies to broader, genome-wide screens employing whole cosmid libraries as transposition substrates, and to incorporate protocols involving specific expression of GFP or PHLEO in the amastigote and metacyclic parasite stages. In this regard, preliminary efforts employing a *Leishmania donovani* cosmid library are encouraging. Lastly, one can combine successive screens for PHLEO or GFP expression to recovery fusion-bearing parasites

showing highly specific expression patterns. In prokaryotes this method is termed 'differential fluorescence induction' (Valdivia & Falkow 1996) and its application in *Leishmania* is now feasible.

In summary, expression profiling and transposon trapping provide two complementary approaches to the study of gene expression in *Leishmania*. It is anticipated that these and other studies will give a better genome-wide picture of mRNA and protein expression, and explain their regulation in the key stages of the *Leishmania* infectious cycle. The challenge then will be to learn how these changes translate into the ability of the parasite to persist and cause disease, and how to use this knowledge to develop new tools for the control of this deadly pathogen.

The authors thank Deborah E. Dobson for discussions and comments on the manuscript. Supported by National Institutes of Health grant nos NIH AI29646 (S.M.B) and HG00249 (G.D.S), DOE ER61606 (F.L.), the Institute Pasteur (S.G.) and an NIH Institutional Training Grant in Genomic Science 5 T32 HG00045 (R.M.). The Microarray Facility is supported by funds from the Siteman Cancer Center and a Danforth Foundation Gift to Washington University Medical School.

REFERENCES

Akopyants, N. S., Clifton, S. W., Martin, J., Pape, D., Wylie, T., Li, L., Kissinger, J. C., Roos, D. S. & Beverley, S. M. 2001 A survey of the *Leishmania major* Friedlin strain V1 genome by shotgun sequencing: a resource for DNA microarrays and expression profiling. *Mol. Biochem. Parasitol.* **113**, 337–340.

Bello, A. R., Nare, B., Freedman, D., Hardy, L. & Beverley, S. M. 1994 PTR1: a reductase mediating salvage of oxidized pteridines and methotrexate resistance in the protozoan parasite *Leishmania major*. *Proc. Natl Acad. Sci. USA* **91**, 11 442–11 446.

Berg, C. M., Berg, D. E. & Groisman, E. A. 1989 Transposable elements and the genetic engineering of bacteria. In *Mobile DNA* (ed. D. E. Berg & M. M. Howe). pp. 879–925. Washington, DC: American Society for Microbiology.

Beverley, S. M., Ellenberger, T. E., Iovannisci, D. M., Kapler, G. M., Petrillo-Peixoto, M. & Sina, B. J. 1988 Gene amplification in *Leishmania*. In *The Biology of Parasitism* (ed. P. T. Englund & A. Sher), pp. 431–448, *MBL Lectures in Biology, vol. 9*. New York: Alan Liss Publishers.

Beverley, S. M. & Turco, S. J. 1998 Lipophosphoglycan (LPG) and the identification of virulence genes in the protozoan parasite *Leishmania*. *Trends Microbiol.* **6**, 35–40.

Blackwell, J. M. & Melville, S. E. 1999 Status of protozoan genome analysis: trypanosomatids. *Parasitology* **118**, S11–S114.

Brown, P. O. & Botstein, D. 1999 Exploring the new world of the genome with DNA microarrays. *Nat. Genet.* **21**, 33–37.

Callahan, H. L. & Beverley, S. M. 1991 Heavy metal resistance: a new role for P-glycoproteins in *Leishmania*. *J. Biol. Chem.* **266**, 18 427–18 430.

Charest, H. & Matlashewski, G. 1994 Developmental gene expression in *Leishmania donovani*: differential cloning and analysis of an amastigote stage-specific gene. *Mol. Cell Biol.* **14**, 2975–2984.

Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P. O. & Herskowitz, I. 1998 The transcriptional program of sporulation in budding yeast. *Science* **282**, 699–705.

Clayton, C. E. 1999 Genetic manipulation of kinetoplastida. *Parasitol. Today* **15**, 372–378.

Coulson, R. M. & Smith, D. F. 1990 Isolation of genes showing increased or unique expression in the infective promastigotes of *Leishmania major*. *Mol. Biochem. Parasitol.* **40**, 63–75.

DeRisi, J. L., Iyer, V. R. & Brown, P. O. 1997 Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686.

Ellenberger, T. E. & Beverley, S. M. 1989 Multiple drug resistance and conservative amplification of the H region in *Leishmania major*. *J. Biol. Chem.* **264**, 15 094–15 103.

El-Sayed, N. M., Hegde, P., Quackenbush, J., Melville, S. E. & Donelson, J. E. 2000 The African trypanosome genome. *Int. J. Parasitol.* **30**, 329–345.

Falkow, S. 1988 Molecular Koch's postulates applied to microbial pathogenicity. *Rev. Infect. Dis.* **10**, S274–S276.

Goyard, S., Tosi, L. R. O., Gouzova, J., Majors, J. & Beverley, S. M. 2001 New Mos1 *mariner* derivatives and transposition assays suitable for the recovery of gene fusions *in vivo* and *in vitro*. *Gene* **280**, 97–105.

Graham, S. V. 1995 Mechanisms of stage-regulated gene expression in Kinetoplastida. *Parasitol. Today* **11**, 217–223.

Gueiros-Filho, F. J. & Beverley, S. M. 1996 Selection against the dihydrofolate reductase-thymidylate synthase (*DHFR-TS*) locus as a probe of genetic alterations in *Leishmania major*. *Mol. Cell. Biol.* **16**, 5655–5663.

Gueiros-Filho, F. J. & Beverley, S. M. 1997 Trans-kingdom transposition of the *Drosophila* element *mariner* within the protozoan *Leishmania*. *Science* **276**, 1716–1719.

Gygi, S. P., Rochon, Y., Franza, B. R. & Aebersold, R. 1999 Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **19**, 1720–1730.

Hartl, D. L., Lohe, A. R. & Lozovskaya, E. R. 1997 Modern thoughts on an ancyent marinere: function, evolution, regulation. *A. Rev. Genet.* **31**, 337–358.

Hayward, R. E., DeRisi, J. L., Alfadhli, S., Kaslow, D. C., Brown, P. O. & Rathod, P. K. 2000 Shotgun DNA microarrays and stage-specific gene expression in *Plasmodium falciparum* malaria. *Mol. Microbiol.* **35**, 6–14.

Heard, P. L., Lewis, C. S. & Chaudhuri, G. 1996 *Leishmania mexicana amazonensis*: differential display analysis and cloning of mRNAs from attenuated and infective forms. *J. Eukaryot. Microbiol.* **43**, 409–415.

Ivens, A. C. & Smith, D. F. 1997 Parasite genome analysis. A global map of the *Leishmania major* genome: prelude to genomic sequencing. *Trans. R. Soc. Trop. Med. Hyg.* **91**, 111–115.

Liu, K., Zinker, S., Arguello, C. & Salgado, L. M. 2000 Isolation and analysis of a new developmentally regulated gene from amastigotes of *Leishmania mexicana mexicana*. *Parasitol. Res.* **86**, 140–150.

Myler, P. J. & Stuart, K. D. 2000 Recent developments from the *Leishmania* genome project. *Curr. Opin. Microbiol.* **3**, 412–416.

Myler, P. J. (and 15 others) 1999 *Leishmania major* Friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proc. Natl Acad. Sci. USA* **96**, 2902–2906.

Ouellette, M., Haimeur, A., Leblanc, E., Grondin, K., Legare, D., Kundig, C. & Papadopoulou, B. 1996 New mechanisms of drug resistance in leishmania. *Trop. Med. Int. Hlth* **1**, A33–A34.

Panton, L. J., Tesh, R. B., Nadeau, K. C. & Beverley, S. M. 1991 A test for genetic exchange in mixed infections of Leishmania major in the sand fly *Phlebotomus papatasi*. *J. Protozool.* **38**, 224–228.

Plasterk, R. H., Izsvak, Z. & Ivics, Z. 1999 Resident aliens: the Tc1/mariner superfamily of transposable elements. *Trends Genet.* **15**, 326–332.

Pogue, G. P., Lee, N. S., Koul, S., Dwyer, D. M. & Nakhasi, H. L. 1995 Identification of differentially expressed *Leishmania donovani* genes using arbitrarily primed polymerase chain reactions. *Gene* **165**, 31–38.

Sacks, D. L. & Perkins, P. V. 1985 Development of infective

stage Leishmania promastigotes within phlebotomine sand flies. *Am. J. Trop. Med. Hyg.* **34**, 456–459.

Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. 1995 Quantitative monitoring of gene expression patterns with a complementary DNA microarray [see comments]. *Science* **270**, 467–470.

Swindle, J. & Tait, A. 1996 Trypanosomatid genetics. In *Molecular biology of parasitic protozoa* (ed. D. F. Smith & M. Parsons), pp. 6–34. Oxford: IRL Press.

Tosi, L. R. & Beverley, S. M. 2000 *cis* and *trans* factors affecting Mos1 *mariner* evolution and transposition *in vitro*, and its potential for functional genomics. *Nucleic Acids Res.* **28**, 784–790.

Turco, S. J., Spath, G. F. & Beverley, S. M. 2001 Is lipophosphoglycan a virulence factor? A surprising diversity between Leishmania species. *Trends Parasitol.* **17**, 223–226.

Ullu, E., Tschudi, C. & Gunzl, A. 1996 *Trans*-splicing in trypanosomatid protozoa. In *Molecular biology of parasitic protozoa* (ed. D. F. Smith & M. Parsons), pp. 115–133. Oxford: IRL Press.

Valdivia, R. H. & Falkow, S. 1996 Bacterial genetics by flow cytometry: rapid isolation of *Salmonella typhimurium* acid-inducible promoters by differential fluorescence induction. *Mol. Microbiol.* **22**, 367–378.

Wu, Y., El Fakhry, Y., Sereno, D., Tamar, S. & Papadopoulou, B. 2000 A new developmentally regulated gene family in Leishmania amastigotes encoding a homolog of amastin surface proteins. *Mol. Biochem. Parasitol.* **110**, 345–357.