ELSEVIER

Molecular & Biochemical Parasitology 123 (2002) 115–123

MOLECULAR & BIOCHEMICAL PARASITOLOGY

www.parasitology-online.com

Analysis of stage-specific gene expression in the bloodstream and the procyclic form of *Trypanosoma brucei* using a genomic DNA-microarray

Susanne Diehl [a], Frank Diehl [a], Najib M. El-Sayed [b], Christine Clayton [c], Jörg D. Hoheisel [a,*]

[a] *Division of Functional Genome Analysis, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 506, 69120 Heidelberg, Germany*
[b] *The Institute for Genomic Research (TIGR), 9712 Medical Center Drive, Rockville, MD, USA*
[c] *Zentrum für Molekularbiologie Heidelberg (ZMBH), Im Neuenheimer Feld 282, 69120 Heidelberg, Germany*

Received 16 April 2002; accepted in revised form 25 June 2002

Abstract

A microarray comprising 21 024 different PCR products spotted on glass slides was constructed for gene expression studies on *Trypanosoma brucei*. The arrayed fragments were generated from a *T. brucei* shotgun clone library, which had been prepared from randomly sheared and size-fractionated genomic DNA. For the identification of stage-specific gene activity, total RNA from in vitro cultures of the human, long slender form and the insect, procyclic form of the parasite was labelled and hybridised to the microarray. Approximately 75% of the genomic fragments produced a signal and about 2% exhibited significant differences between the transcript levels in the bloodstream and procyclic forms. A few results were confirmed by Northern blot analysis or reverse-transcription and PCR. Three hundred differentially regulated clones have been selected for sequencing. So far, of 33 clones that showed about 2-fold or more over-expression in bloodstream forms, 15 contained sequences similar to those of VSG expression sites and at least six others appeared non-protein-coding. Of 29 procyclic-specific clones, at least eight appeared not to be protein-coding. A surprisingly large proportion of known regulated genes was already identified in this small sample, and some new ones were found, illustrating the utility of genomic arrays. © 2002 Published by Elsevier Science B.V.

Keywords: *Trypanosoma brucei*; Microarray; Regulation; Gene; Expression

1. Introduction

The life cycle of *Trypanosoma brucei* involves adaptation to a variety of conditions in the host and Tsetse fly. The successive changes in morphology, biochemistry and plasma membrane proteins, some of which also involve cell cycle arrest, are still very poorly understood. Many of these changes are likely to be directed by changes in mRNA abundance and translation, and over two decades of effort have now been expended in the identification of stage-specific mRNAs. Since of technical limitations, however, most of the transcripts identified have been rather abundant, and the regulation studied has mainly been restricted to the rapidly-dividing long slender bloodstream and procyclic forms.

Given the relatively small size of the *T. brucei* genome, there is a good prospect of a complete exploration of its genome by microarray analyses. This would allow identification of lower-abundance regulated transcripts, and (using amplification methods) the study of the transcriptome of the less accessible forms found in the Tsetse fly. The ideal solution for the analysis of mRNAs would be to construct microarrays by PCR-amplification of open reading frames. This is currently not possible as the genome is incomplete, and might also prove prohibitively expensive. As an alternative, EST clones have already been used [1]. These are, however, limited by the fact that for complete coverage normalised cDNA libraries from all different stages would be required (which is probably technically

* Corresponding author
E-mail address: j.hoheisel@dkfz.de (S. Diehl).

0166-6851/02/$ - see front matter © 2002 Published by Elsevier Science B.V.
PII: S0166-6851(02)00138-X

impossible). A further alternative is to perform genome-wide expression studies on genomic instead of gene-specific fragments. Such arrays have several intrinsic advantages. Overall genome representation is usually good in shotgun libraries with comparatively little variation across the genome. Thus, even if randomly selected clone inserts are used as probes, there should be good coverage and relatively little redundancy. Also, not only coding but also intergenic regions can be studied. Insert amplification can be performed with a single primer pair and functional analyses can actually precede sequencing. In trypanosomes, about 50% of the genome are coding sequence, open reading frames have an average size of 1–2 kb and nearly always lack introns [1]. Genes are clustered densely in polycistronic transcription units separated by short intergenic regions of 0.5–1 kb. Thus, for a short random genomic fragment of about 2 kb the likelihood to contain sequence that is represented either as mRNA or structural RNA is relatively high.

We have produced DNA-microarrays containing more than 20 000 PCR-products of about 2 kb long random genomic fragments of *T. brucei* strain TREU927/4. We describe here the results of an initial validation of the approach, in which gene expression in bloodstream and procyclic trypanosomes was monitored.

2. Materials and methods

2.1. Genomic shotgun library

The TIGR shotgun library was made of *T. brucei* strain TREU927/4 and consists of fragments in the 2–2.5 kb size range cloned into the *Sma*I site of plasmid vector pUC18. An aliquot of the library was transformed into electro-competent *Escherichia coli* DH10b. Randomly selected colonies were transferred into 384-well microtitre plates containing LB-medium supplemented with 100 μg ml^{-1} ampicillin and 10% HMFM freezing medium [2], grown overnight at 37 °C and stored at −80 °C.

2.2. Genomic *T. brucei* microarray

From shotgun clones in overall 56 microtitre plates, 21 024 PCR-products were generated using as primers d(TTGTAAAACGACGGCCAGTG) and d(GCGGATAACAATTTCACACAGGA), which are flanking the multiple cloning site of vector pUC18. All products were checked individually by standard agarose gel electrophoresis. Also, several known trypanosome genes, including *VSG*, *EP1*, *GAPDH*, and *PGK* were amplified by gene-specific primers. In addition, 384 human cDNA clones of the IMAGE library [3] were added as negative controls.

The fragments were spotted onto poly-L-lysine coated glass slides following the spotting and postprocessing protocol described in detail elsewhere [4]. Spotting was done with an SDDC-2 DNA Micro-Arrayer from Engineering Services Inc. (Toronto, Canada) and SMP-3 pins (TeleChem International Inc., Sunnyvale, USA). Since the spotting buffer contained 1.5 M betaine, humidity control could be avoided during the process. Subsequently, the slides were UV-irradiated with 60 mJ total energy (UV-crosslinker UVC-500, Hoefer Scientific Instruments, San Francisco, USA).

2.3. Parasites

T. brucei brucei clone TREU927/4 had been maintained as glycerol stock at −80 °C. Bloodstream and trypanosomes were grown in HMI-9 [5] supplemented with10% foetal calf serum, and procyclic trypanosomes were grown in MEM-Pros (DTM) supplemented with Hemin and 10% foetal calf serum [6]. Cells were always harvested in the log phase of growth: below 1×10^6 cells per ml for bloodstream forms and 2×10^6 cells per ml for procyclics.

2.4. Sample preparation, labelling and hybridisation

Pelleted cells were lysed in 1 ml Trizol (Invitrogen, Karlsruhe, Germany) per 10^6 cells and total RNA was extracted following the manufacturers protocol. Probes for hybridisation were generated from 60 μg of total RNA by incorporation of fluorophor-labelled dCTP during first-strand cDNA-synthesis. The reaction mix contained 3 moles of either Cy3- or Cy5-dCTP (Amersham Biosciences, Freiburg, Germany), 4 μg oligo-$dT_{(15)}$ primer, 0.5 mM each of dATP, dGTP, dTTP and 0.2 mM dCTP, 100 mM dithiothreitol and 400 units of Superscript II reverse transcriptase (Invitrogen) in the buffer provided by the manufacturer. Incubation was at 42 °C for 2 h. Subsequently, the RNA was hydrolysed by the addition of 10 μl 0.1 M NaOH and an incubation at 65 °C for 10 min. The solution was neutralised by addition of the same volume of 0.1 M Tris–HCl, pH 7.5. The samples were purified with the QIAquick PCR-purification kit (Qiagen, Hilden, Germany). The incorporation rate of fluorophor was then determined by photometric measurement. The samples were dried in vacuum and the pellets dissolved in 60 μl hybridisation buffer (50% formamide, 3 × SSC, 1% SDS, 5 × Denhardt's reagent and 5% dextran sulfate) [7]. Hybridisation to the array was performed under a glass cover-slip (25 × 60 mm). The slides were kept in a humidified chamber (TeleChem International Inc.) in a waterbath at 45 °C for 16 h. The arrays were washed at room temperature in 150 mM sodium chloride, 15 mM

sodium citrate pH 7.5 ($1 \times$ SSC) for 2 min and then briefly rinsed in $0.2 \times$ SSC.

2.5. Image acquisition and data analysis

The signal intensities were determined with a laser-scanner (ScanArray 5000, Packard BioScience, Dreieich, Germany). Images were acquired by performing subsequent scans with the Cy3- and Cy5-channel at a resolution of 5 μm. The resulting images were analysed using GenePix software (Axon Instruments, Union City, USA), generating numerical values for the signal intensities at each spot. Only spots in which more than 80% of the pixels had a signal above background of at least twice the standard deviation (S.D.) of the local background were considered in subsequent data analysis.

Normalisation was done by two different methods. In all cases, the overall signal intensity on all spots was used to normalise individual experiments. In addition, spiking probes made from lambda phage had been added to some samples prior to labelling, permitting an independent cross-check of the results. The significance of the observed variations in transcript levels was analysed by statistical procedures described in [8]. The median of the signal intensities obtained on each fragment was calculated and the significance of variations was assessed by two stringency criteria. The highly stringent ‘min–max separation’ is calculated by taking the minimum distance between all data points of the two conditions. The less stringent criteria, called ‘standard deviation separation’, is defined as the difference of the means of either data set diminished by one S.D. [8]. Some statistically insignificant differences in expression that were confirmed by Northern blot analysis or RT-PCR (see below) were also included in the final Table of results. Data quality assessment, normalisation and analysis were performed with a software package [9,10] (www.dkfz-heidelberg.de/tbi/services/mchips) written in MATLAB (Mathworks, Natick, USA).

2.6. Sequence analysis

Sequences of the clones were obtained by long reads from one or both ends performed by a commercial provider (Genotype AG, Hirschhorn, Germany). Contigs were built and corrected, where possible, by matching to the *T. brucei* genome project database at http://www.ebi.ac.uk. The resulting final sequences were compared with others through searches of the non-redundant protein and DNA databases at the National Center for Biotechnology Information (NCBI, NIH, Bethesda, USA) and EMBL (Hinxton) using the BLASTN, BLASTX and BLASTP algorithms.

2.7. Confirmation of transcriptional variations

Cases of differential transcription were confirmed by RT-PCR and Northern blot analyses. RT-PCR was performed using gene-specific primers. Primers were designed using the PRIMER3 program (http://www.genome.wi.mit.edu/cgi-bin/primer/primer3.cgi/primer3_www.cgi) and obtained from MWG-Biotech (Ebersberg, Germany). About 50 μg of total RNA was purified by precipitation, adding to the RNA samples an equal volume of 5 M LiCl, 0.1 M Tris–HCl (pH 7.4). Precipitation was overnight at 4 °C, followed by centrifugation at $12\,000 \times g$. The pellet was washed with 70% ethanol and resuspended in diethylpyrocarbonate-treated water. For cDNA synthesis, 5 μg of purified total RNA were reverse transcribed at 42 °C for 1 h in 50 μl of 500 μM dNTPs 10 μM dithiothreitol, 1.25 μM $dT_{(15)}$ primer, 40 U RNAsin (Promega, Mannheim, Germany) and 75 units of Superscript Reverse Transcriptase II (Invitrogen). Reactions were diluted 10-, 100- and 1000-fold. PCR amplifications with 2 μl of each dilution were performed in a total volume of 20 μl of 50 μM dNTPs, 0.5 μM of each oligonucleotide primer, 1.75 mM magnesium chloride and 1 U of Taq-DNA polymerase (Qiagen) using MJ-Research P200 thermocyclers. To avoid saturation or plateau effects, PCR was limited to 22 cycles.

3. Results

To construct the *T. brucei* microarray, we started with a previously-constructed random-shear shotgun genomic library containing inserts of 2–4 kb. About 21 024 different inserts were PCR-amplified, then spotted onto glass slides coated with poly-L-lysine. With an average length of 2.25 kb, the fragments represent a 1.3-fold coverage of the *T. brucei* genome. Due to the size selection process prior to the shotgun cloning, the fragments are rather similar in length. This is advantageous for microarray analysis, since overall PCR amplification yields are influenced by size differences. Also, variation in the efficiency of priming events is minimised, since an identical primer pair was used in all reactions. As positive controls, several cDNA clones representing known genes such as histones and *EP1* were added to the array, while rRNA fragments and 384 human cDNA clones were used as negative controls.

For the detection of fragments that contain genes which are differentially transcribed in the bloodstream and procyclic stages of the parasite, six independent experiments were performed. RNAs were prepared from logarithmically growing cells and were made separately for each hybridisation. We initially performed hybridisations in which the procyclic RNA and the bloodstream RNA were labelled with Cy3 and Cy5,

respectively. To prevent bias caused by preferential label incorporation into particular sequences, the dyes were swapped between the two types of RNA. In addition, we performed hybridisations with procyclic RNA labelled either with Cy3 or Cy5 in order to determine the intrinsic variation between two samples made from identical culture conditions. No significant variations caused by this dye-swapping were observed.

Approximately 75% of the arrayed fragments contained a transcribed sequence that was detectable by microarray analysis in either bloodstream, procyclic or both stages. Analysis of the human IMAGE cDNA controls revealed that unspecific hybridisation to non-homologous sequence did not occur, since none of these fragments produced a signal that was above the background signal (data not shown). A scatter plot of relative expression ratios of all detected signals is shown in Fig. 1. Unregulated data points are to be found on the main diagonal, and regulated ones are outside the additional diagonals.

In total, approximately 300 fragments contained sequence that appeared to be differentially expressed. The significance of the observed variations in transcript levels were analysed by statistical procedures [8]. Initially, only significantly differentially transcribed genes were taken into account. Subsequently, also some apparent but statistically insignificant differences in expression were examined further by Northern blot analysis to confirm that the results were genuine. At the time of array construction, sequenced clones of this size were not available, so to check their identity, we have started sequencing the differentially transcribed clones. Most were sequenced from both sides, usually producing up to 1 kb of sequence information from either end. Where possible, the sequence of the complete insert was constructed by incorporating available sequence from the genome database.

The results of the analysis for the first 70 successfully-sequenced clones are displayed in Table 1, and are summarised in Fig. 2. Of the clones that were preferentially expressed in bloodstream forms, half contained sequence that was strongly related to (and would, therefore, hybridise with) mRNAs derived from the VSG expression sites. These included not only ESAG and GRESAG genes, but also a variety of sequences related to (but not identical to) known VSGs. It is quite likely that many of these genes—especially the VSG genes—are not actually represented in bloodstream-form mRNA, but are detected because of cross-hybridisation with related expressed sequences. Of the other bloodstream form mRNAs, the differential expression of GPIPLC [11], invariant surface glycoprotein [12], EP locus transcription terminator [13], and alternative oxidase genes had all been previously documented [14]. Interesting newcomers included sequences likely to encode an NADH dehydrogenase subunit, an amino

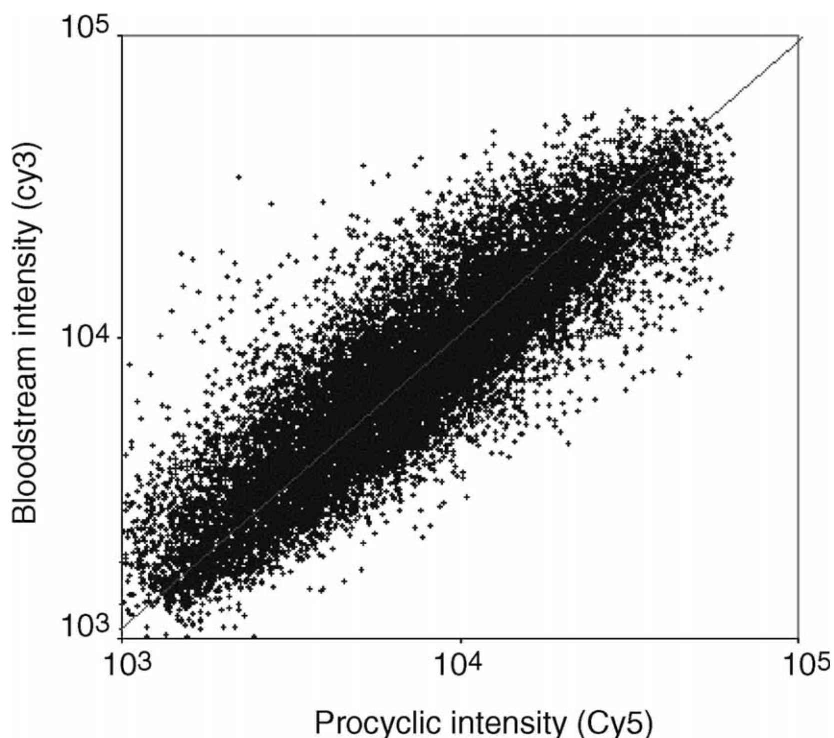

Fig. 1. Scatterplot comparing the hybridisation of bloodstream and procyclic form RNA probes with the arrayed DNAs. Initially spots were removed when, for both the Cy3 and Cy5 signals, intensities were more than the background +2 S.D. in 80% or more pixels. Spots in which more than 20% of the pixels were saturated were also excluded. The values were normalised to the sum of the signal intensities, and any additional spots where the signals were saturated (normalised level over 65 000) or under 1000 were excluded.

Table 1
List of differentially transcribed genes

Id	96w	Acc #	Putative identity	BF/PF	Conf.
Genes similar to expression-site-associated genes or VSGs					
01_13_02_21	F04	AJ438524, AJ438525	ESAG7	−9.9±4.0	
01_13_08_21	F09	AJ438529, AJ438530	65 KDa invariant surface glycoprotein or VSG [12]	−3.9±1.3	
01_09_03_14	D06	AJ438508*	VSG	−3.8±1.6	
01_14_11_22	G03	AJ438547, AJ438548	ESAG8	−3.1±0.9	
01_13_02_24	F05	AJ438526*	Could be VSG	−2.7±1.7	
01_05_07_12	B11	AJ438495, AJ438496	VSG promoter	−2.6±0.9	
01_16_10_17	H03	AJ438560*	VSG	−2.5±0.8	
01_17_09_13	H06	AJ438564*	VSG	−2.4±0.9	
01_17_02_05	H04	AJ438561, AJ438562	ESAG2	−2.3±0.7	
01_12_07_23	E12	AJ438520, AJ438521	VSG	−2.3±0.8	
01_14_01_03	G01	AJ438545*	VSG	−2.1±0.7	
01_13_14_11	F11	AJ438543*	ESAG4	−1.9±0.6	
01_13_14_16	F12	AJ438544*	VSG	−1.9±0.5	
01_03_11_08	A12	AJ438538*	GRESAG4/ESAG4	−1.7±0.2	
01_15_13_06	G10	AJ438556*	GRESAG2/ESAG2	−1.6±0.5	PCR
Non-expression-site genes up-regulated in bloodstream forms					
01_11_10_21	E09	AJ438518*	Hypothetical protein in many species, belongs to Pfam-B_9175, function unknown	−3.3±1.3	
01_05_09_17	B12	AJ438498	Protein (fragment) *L.major* Q9N7K2	−2.9±0.9	
01_07_05_01	C10		*T. brucei* GPI-PLC CAB60085 plus downstream 900 bp	−2.7±1.0	Northern PCR
01_01_01_12	A01	AJ438532*	*T. brucei* 75kDa invariant surface glycoprotein [12]	−2.5±0.7	
01_03_15_24	B02	AJ438539*	Putative CAAXprenylprotease, *T.cruzi* AF252543 [24]	−2.3±1.0	PCR
01_17_09_13	H07	AJ438565*	Ribosomal P0 protein *T. congolense* AB056702	−2.3±1.0	
01_13_06_09	F07		*T. brucei* alternative oxidase U52964 [14]	−2.2±0.7	PCR
01_02_01_13	A07		*T.brucei* DNA for procyclin PARP A transcription terminator Z36932, best match chromosome 10. [13]	−2.1±0.6	
01_05_05_21	B08		Axonemal dynein heavy chain (*T. brucei*, chromosome 10)	−2.0±0.8	
Up-regulated in bloodstream forms and lacking similarities to genes from other species in databases					
01_18_05_13	H09	AJ438567, AJ438568	No ORFs found	−4.6±1.3	
01_16_06_04	G12	AJ438559	ORF but no significant matches	−3.2±1.3	Northern
01_15_09_24	G07	AJ438552*	No ORFs found	−3.1±1.2	
01_02_09_01	A08, A09	AJ438533	Repetitive, no ORFs, possibly chromosome 10	−2.6±1.1	
01_04_09_03	B05	AJ438540*	ORF but no significant matches	−2.6±0.9	
01_05_14_21	C01	AJ438499*	No ORFs found	−2.4±0.9	
01_05_05_22	B09	AJ438492, AJ438542	No ORFs found	−2.0±0.5	Northern (faint)
01_07_07_23	D01	AJ438502*	ORF but no significant matches	−1.9±0.6	
01_11_06_21	E07	AJ438515	Short ORFs, no significant matches	−1.9±0.6	
EP/GPEET locus					
01_12_11_18	F03		*T. brucei* PARP-A locus X52584	2.3±0.8	PCR
01_16_09_01	H02		*T. brucei* PARP-A locus X52584	3.9±1.1	
Polymerase II genes up-regulated in procyclic forms					
01_04_15_01	B06		*T. brucei* pyruvate phosphate dikinase, AF048689 [25]	8.9±2.9	
01_04_02_03	B03		*T. brucei* glycosomal malate dehydrogenase, AF079110 [26]	6.7±2.4	
01_17_15_01	H08	AJ438566*	Possible amino acid transporter, *L. major* Q9BHF5 (2.9e−48)	4.8±1.8	PCR
01_10_05_05	D11		Heat shock protein 83, *T. brucei* X14176 [27]	4.7±0.4	PCR
01_11_02_06	E05	AJ438513*	Activator of G-protein signalling, mouse, $P = 7.2e{-}07$. Q9R080, 3 TPR repeats, aminotransferase domain	4.6±1.4	PCR
01_13_09_01	F10		*T. brucei* Heat shock protein 83, X14176 [27]	3.5±0.4	
01_10_12_05	E03	AJ438510*	Hypothetical membrane protein,.*L. major* AL133435	3.3±1.2	Northern

Table 1 (*Continued*)

Id	96w	Acc #	Putative identity	BF/PF	Conf.
01_13_06_21	F08	AJ438528*	Possible amino acid transporter, *L. major* Q9BHG8 (1.3e-08)	3.1±0.8	
01_07_03_22	C09	AJ438501*	*T. brucei* retrotransposon hotspot protein, CAD21755,	3.1±0.8	Northern
01_14_14_20	G04	AJ438549, AJ438550	*T. brucei* F1 ATPase alpha subunit, AY007705	3.1±0.8	PCR
01_11_15_14	E10	AJ438519*	Infective insect stage-specific protein META1, *L. donovani* O43990 [15]	3.1±0.9	
01_10_15_23	E04	AJ438511, AJ438512	Tropomyosins from several species	2.9±0.9	
01_10_10_09	E02	AJ438509*	Hypothetical protein, *L. major* AL139794	2.8±0.9	
01_07_12_01	D02	AJ438503, AJ438504	NADH-cytochrome b5 reductases from several species	2.6±0.9	
01_15_12_22	G09	J438554, AJ438555	Importin beta from several species	2.3±0.8	PCR
Up-regulated in procyclic forms and lacking similarities to genes from other species in databases					
01_15_06_21	G06	AJ438551*	No ORFs found	4.1±1.1	
01_11_09_11	E08	AJ438516, AJ438517	No ORFs found	4.0±1.4	
01_05_06_13	B10	AJ438493, AJ438494	No ORFs found	3.8±1.0	
01_13_03_04	F06	AJ438527*	No ORFs found	3.7±1.9	
01_16_05_10	G11	AJ438557, AJ438558	No ORFs found	3.7±1.4	
01_17_06_22	H05	AJ438563	ORF, no significant matches	3.4±1.2	
01_08_07_22	D04	AJ438506, AJ438507	ORFs? no significant matches	3.1±1.1	
01_04_15_05	B07	AJ438541*	ORF, no significant matches	3.0±1.0	
01_15_10_15	G08	AJ438553*	Match *T. brucei* contig 1.0.4383, 100 nt of *ESAG1* sequence 2 kb away	2.7±1.1	
01_11_04_13	E06	AJ438514*	ORF, no significant matches	2.5±1.1	
01_03_09_23	A11	AJ438536, AJ438537	ORF, no significant matches	2.5±0.8	
01_07_16_20	D03	AJ438505*	No ORFs found	2.4±0.6	Northern

The id is the clone identification number and '96w' is a second identifier. Accession numbers (Acc #) denote deposited raw sequence from single reads. Where the complete sequence of the DNA is (approximately) known, either because the single reads overlapped or because they could be linked by searching the *T. brucei* database, the accession number has an asterisk*. Putative identities are based on searches at the amino acid level, after previous alignment and (where appropriate) correction from the *T. brucei* genome database. 'No ORF' means no apparent open reading frame of more than about 100 amino acids. Since we have relied on single reads, this designation could be clearly be inaccurate because of sequencing errors. BF/PF gives the ratio between Cy5- and Cy3-Signal intensity and the S.D. Positive values indicate overexpression in procyclic stage trypanosomes, negative ratios indicate overexpression in bloodstream stage trypanosomes. The ratio values in the table are means and S.D. of normalised signal intensity ratios derived from six independent hybridisation experiments. 'conf' denotes confirmation by Northern blot or PCR.

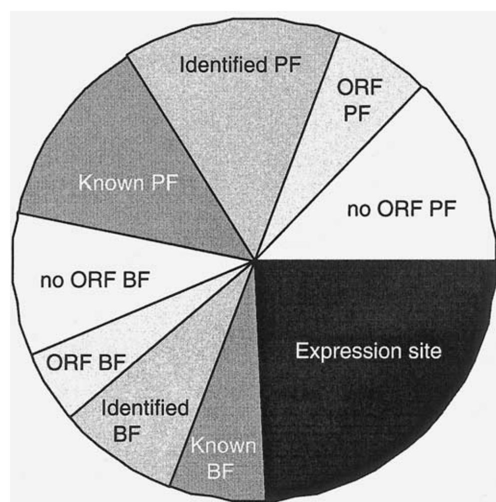

Fig. 2. Broad classification of regulated clones. Clones with 'No ORF' lack open reading frames of 100 amino acids throughout the sequenced length (see also legend to Table 1).

acid transporter, CAAX prenyl protease, axonemal dynein and ATPase. Surprisingly, six clones appeared to lack open reading frames. Procyclic-specific clones again included a number of known regulated genes, several interesting new ones including three membrane proteins with homologues in Leishmania, one of which (META1) had previously been shown to be specific for Leishmania insect stages [15]. Again at least eight procyclic-specific clones lacked recognisable open reading frames.

To confirm our findings we performed semi-quantitative RT-PCRs and Northern blots. Using a dilution series of reverse-transcriptase reactions of bloodstream and procyclic total RNA, the amount of transcript present was quantified. Primers were designed to yield PCR-products of 200–300 base pairs in length. PCR amplification was successfully performed for all selected clones and confirmed the differential transcript levels (Fig. 3). Results confirmed included the bloodstream-

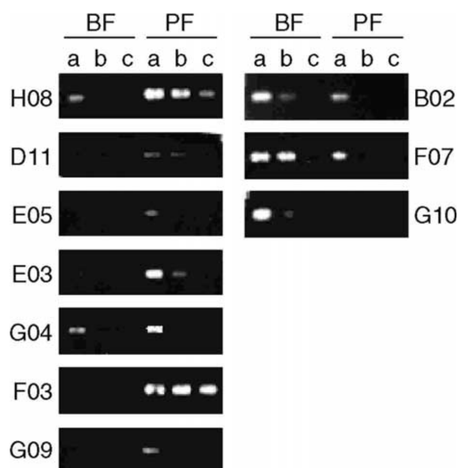

Fig. 3. RT-PCR analysis confirmed the stage specific expression of novel and previously characterised genes. Semi-quantitative PCR analysis was performed on cDNA generated from procyclic (PF) and bloodstream form (BF) trypanosomes. The cDNA template was diluted 1:10 (a), 1:100 (b) and 1:1000 (c). Genes are named by their clone numbers in Table 1.

specific expression of the putative CAAX-prenyl protease, alternative oxidase and GRESAG protein 2.1-1, and the procyclic-specific expression of the possible amino acid transporter, HSP 83, activator of G-protein signalling, hypothetical protein L71710.1, ATPase alpha subunit, putative Importin ß-subunit and *EP*. Some of the results were also confirmed by Northern blotting (see Table 1).

4. Discussion

Within recent years, microarray technology has become a major tool for the investigation of transcriptional variations in gene expression. For comprehensive analyses on all genes of the 35 Mb genome of *T. brucei*, however, the construction of a gene-specific microarray is not yet possible. As an alternative, we generated a microarray from genomic fragments. The 21 024 randomly selected PCR-products correspond to a statistical genome representation of about 86% of the genomic sequence of *T. brucei* [16]. Due to the random nature of the shotgun cloning procedure, there is an unbiased representation of sequences in the genomic library, so genes that are rarely transcribed have the same representation as highly expressed ones.

Sequence annotation is not foolproof. It misses short open reading frames and certainly will not detect many of the unusual but functional non-transcribed RNAs [17,18]. A genomic array—especially if it is built as a tile-path of fragments of known sequence—will obviously always provide more complete genomic representation. The main disadvantage of our current array is that the insert size is rather too large. With an insert size averaging 2 kb, clones can easily include more than one open reading frame and as a result, differential regulation of one transcript might be masked by strong signals from a neighbouring, constitutively-expressed gene.

We found that approximately 2% of the genomic fragments contained sequences that were developmentally regulated, as defined by difference in signal ratio greater than 1.9 after hybridisation with probes representing total RNA from bloodstream and procyclic trypanosomes. 99.5% of the fragments displayed a signal ratio of 1.8 or less. In a previous study, it was found that 57 of 400 cDNA and GST clones in an array were differentially expressed when probed with labelled probes made using poly(A)+RNA [1]. This much higher proportion—14%—is most likely explained by the fact that highly-expressed (and often regulated) mRNAs were likely to be preferentially represented in the cDNA libraries used, and the fact that different RNA fractions were used.

Comparing the ratios seen in our study with those in the literature (and, sometimes, with our own blots) it is quite clear that the amplitude of the regulation seen by Cy3/Cy5 comparison is very considerably suppressed compared with that seen using radioactive probes. For example, regulation of the EP/GPEET loci registered at 2–3-fold whereas the mRNAs are really regulated up to 100-fold. Similarly, the alternative oxidase mRNA is virtually undetectable in procyclic forms but in our study only 2-fold up-regulation was registered. This discrepancy is a regular feature of this type of analysis; it might be caused by competition between the two fluorescent probes (if the RNA is very abundant) as well as by non-linearity in the detection systems.

Considering that we have so far only obtained good sequence from 62 differentially-regulated clones, we have already identified an extraordinarily large proportion of the known regulated genes from *T. brucei.* Comparing our results with other studies, genes which we have so far ‘missed’ include cyt oxidase su IV [19] GP63, RNA helicase, Kinetoplastid MP11, ESAG11, calpain and pyruvate kinase [1]. Some regulated genes that will inevitably escape detection using our approach are members of gene families showing a variety of different expression patterns. Examples of these include the three phophoglycerate kinase mRNAs, which show three different regulation patterns but differ mainly in their 3′-untranslated regions, and the hexose transporter mRNAs, which are either procyclic or bloodstream-specific but differ only within a short segment of the 3′-untranslated region.

One of the most intriguing aspects of our results is the quite large proportion of identified clones (up to 20%) that shows differential regulation at the RNA level but contains neither recognisable open reading frames nor previously-identified non-coding RNA sequences. Some had repetitive segments. These clones might be tem-

plates for previously-unidentified non-coding transcripts, or they might be detected through cross-hybridisation with other, coding RNAs. They could be long 3′-untranslated regions, or they could be intergenic regions that are transcribed as part of polycistronic transcripts and processed because they contain sequences capable of trans splicing. There is so far no evidence that the 'nonsense-mediated decay' system (degradation of mRNAs containing premature termination codons) operates in trypanosomes (see for example [20–22]) and processed but apparently non-coding transcripts have been described [23].

We have found 20 'new' regulated genes from sequencing about 20% of the regulated clones. Thus the total yield of new regulated genes is likely to be around 100. This is a surprisingly accessible number which will lend itself readily to more detailed functional analysis. On the basis of these results, we will be able to extend our analyses to detection of genes induced by specific differentiation stimuli or inhibitor treatments. The abundance of trypanosomatid mRNAs is determined largely by sequences in the 3′-untranslated regions, so the results of these analyses should enable us to detect specific sequences within the 3′-untranslated regions that determine particular differentiation patterns.

Acknowledgements

We are grateful to Achim Stephan for the preparation of poly-L-lysine slides, Marita Schrenk for her excellent technical assistance and Holger Eickhoff at the MPI for Molecular Genetics (Berlin) for access to his robotic picking devices, to Kurt Fellenburg for help with M-CHIPS, and also to Sanjay Shahi, Claudia Hartmann and Drifa Gundulsdottjur-Planck for assistance in trypanosome culture. We thank the students of the Biology practical class A1 (nucleic acids) for performing the Northern analyses and Frank Voncken and Antonio Estevez for supervising them.

References

[1] El-Sayed NM, Hegde P, Quackenbush J, Melville SE, Donelson JE. The African trypanosome genome. Int J Parasitol 2000;30:329–45.

[2] Werner E, Holder AA, Hoheisel JD. Growth and storage of YAC-clones in Hogness freezing medium. Nucleic Acids Res 1997;25:1467–8.

[3] Hillier L, Lennon G, Becker M, Bonaldo MF, Chiapelli B, Chissoe S, Dietrich N, DuBuque T, Favello A, Gish W, Hawkins M, Hultman M, Kucuba T, Lacy M, Le M, Le N, Mardis E, Moore B, Morris M, Parsons J, Prange C, Rifkin L, Rohlfing T, Schellenberg K, Soares MB, Tan F, Thierry-Meg J, Trevaskis E, Underwood K, Wohldman P, Waterston R, Wilson R, Marra M. Generation and analysis of 280 000 human expressed sequence tags. Genome Res 1996;6:807–28.

[4] Diehl F, Grahlmann S, Beier M, Hoheisel JD. Manufacturing DNA microarrays of high spot homogeneity and reduced background signal. Nucleic Acids Res 2001;29:E38.

[5] Hirumi H, Hirumi K. Continuous cultivation of *Trypanosoma brucei* bloodstream forms in a medium containing a low concentration of serum protein without feeder cell layers. J Parasitol 1989;75:985–9.

[6] Overath P, Czichos J, Haas C. The effect of citrate/*cis*-aconitate on oxidative metabolism during transformation of *Trypanosoma brucei*. Eur J Biochem 1986;160:175–82.

[7] Welford SM, Gregg J, Chen E, Garrison D, Sorensen PH, Denny CT, Nelson SF. Detection of differentially expressed genes in primary tumor tissues using representational differences analysis coupled to microarray hybridisation. Nucleic Acids Res 1998;26:3059–65.

[8] Beißbarth T, Fellenberg K, Brors B, Arribas-Prat R, Boer J, Hauser NC, Scheideler M, Hoheisel JD, Schütz G, Poustka A, Vingron M. Comparison and quality measures of DNA array hybridisation data. Bioinformatics 2000;16:1014–22.

[9] Fellenberg K, Hauser NC, Brors B, Neutzner A, Hoheisel JD, Vingron M. Correspondence analysis applied to microarray data. Proc Natl Acad Sci USA 2001;98:10781–6.

[10] Fellenberg K, Hauser NC, Brors B, Hoheisel JD, Vingron M. Microarray data warehouse allowing for the statistical analysis of experiment annotations, Bioinformatics 2002;19:423–33.

[11] Carrington M, Bülow R, Reinke H, Overath P. Sequence and expression of the glycosyl-phosphatidylinositol-specific phospholipase C of *Trypanosoma brucei*. Mol Biochem Parasitol 1989;33:289–96.

[12] Ziegelbauer K, Multhaup G, Overath P. Molecular characterisation of two invariant surface glycoproteins specific for the bloodstream stage of *Trypanosoma brucei*. J Biol Chem 1992;267:10797–803.

[13] Berberof M, Pays A, Lips S, Tebabi P, Pays E. Characterization of a transcription terminator of the procyclin PARP A unit of *Trypanosoma brucei*. Mol Cell Biol 1996;16:914–24.

[14] Chaudhuri M, Hill GC. Cloning, sequencing, and functional activity of the *Trypanosoma brucei* brucei alternative oxidase. Mol Biochem Parasitol 1996;83:125–9.

[15] Uliana SR, Goyal N, Freymuller E, Smith DF. Leishmania: overexpression and comparative structural analysis of the stage-regulated meta 1 gene. Exp Parasitol 1999;92:183.

[16] Clarke L, Carbon J. A colony bank containing synthetic ColE1 hybrid plasmids representative of the entire *E. coli* genome. Cell 1976;9:91–9.

[17] Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG. Life with 6000 genes. Science 1996;274:546–67.

[18] Kumar A, Harrison PM, Cheung K-H, Lan N, Echols N, Bertone P, Miller P, Gerstein MB, Snyder M. An integrated approach for finding overlooked genes in yeast. Nat Biotechnol 2002;20:58–63.

[19] Matthews KR, Gull K. Identification of stage-regulated and differentiation-enriched transcripts during transformation of the African trypanosome from its bloodstream to procyclic form. Mol Biochem Parasitol 1998;95:81–95.

[20] Pays E, Coquelet H, Pays A, Tebabi P, Steinert M. *Trypanosoma brucei*: posttranscriptional control of the variable surface glycoprotein gene expression site. Mol Cell Biol 1989;9:4018–21.

[21] Vassella E, Braun R, Roditi I. Control of polyadenylation and alternative splicing of transcripts from adjacent genes in a procyclin expression site: a dual role for polypyrimidine tracts in trypanosomes? Nucleic Acids Res 1994;22:1359–64.

[22] Liniger M, Bodenmuller K, Pays E, Gallati S, Roditi I. Overlapping sense and antisense transcription units in *Trypanosoma brucei*. Mol Microbiol 2001;40:869–78.
[23] Scholler JK, Aline RF, Jr, Stuart KD. Variant specific transcripts from the co-transposed segments of variant surface glycoprotein genes in *Trypanosoma brucei*. Mol Biochem Parasitol 1988;29:89–104.
[24] Porcel BM, Åslund L, Pettersson U, Andersson B. *Trypanosoma cruzi*: a putative vacuolar ATP synthase subunit and a CAAX prenyl protease-encoding gene, as examples of gene identification in genome projects, Exp Parasitol 2000; 95:176–186.
[25] Bringaud F, Baltz D, Baltz T. Functional and molecular characterisation of a glycosomal PPi-dependent enzyme in trypanosomatids: pyruvate, phosphate dikinase. Proc Natl Acad Sci USA 1998;95:7963–8.
[26] Anderson SA, Carter V, Hagen CB, Parsons M. Molecular cloning of the glycosomal malate dehydrogenase of *Trypanosoma brucei*. Mol Biochem Parasitol 1998;96:185–9.
[27] Van der Ploeg LHT, Giannini SH, Cantor CS. Heat shock genes: regulatory role for differentiation in parasitic protozoa. Science 1985;228:1443–5.