

Investigating amoebic pathogenesis using *Entamoeba histolytica* DNA microarrays

UPINDER SINGH^{*,#,†} and PREETAM SHAH^{*}

Departments of Internal Medicine^{*} and Microbiology[#], Stanford University, Stanford, CA 94305, USA

[†]Corresponding author (Fax, 650-724-3892; Email, usingh@stanford.edu)

Entamoeba histolytica, a protozoan parasite, causes diarrhea and liver abscesses resulting in 50 million cases of infection worldwide annually. Elucidation of parasite virulence determinants has recently been investigated using genetic approaches. We have undertaken a genomics approach to identify novel virulence determinants in the parasite. A DNA microarray of *E. histolytica* is being developed based on sequenced genomic clones from the genome sequencing efforts of The Institute of Genomic Research (TIGR) and the Sanger Center. Hybridization of the slides with samples labelled differentially using fluorescent dyes allows the characterization of transcriptional profiles of genes under the biological conditions tested. Additionally, a genome-wide comparison of *E. histolytica* and *E. dispar* can be undertaken. The development of an *E. histolytica* microarray will be outlined and its uses in identifying novel virulence determinants and characterizing amoebic biology will be discussed.

[Singh U and Shah P 2002 Investigating amoebic pathogenesis using *Entamoeba histolytica* DNA microarrays; *J. Biosci. (Suppl. 3)* 27 595–601]

1. Introduction

In recent years, genome sequencing of many protozoan parasites has been undertaken and sequencing efforts are currently underway for a number of parasites, including the enteric pathogen *Entamoeba histolytica*. The availability of sequence data information has revolutionized the study of parasite biology and facilitated the development of powerful post-genomics tools such as DNA microarrays. This review summarizes the approach of DNA microarrays to study parasite biology with an emphasis on its applications in the study of *E. histolytica*.

The protozoan parasite *E. histolytica* causes invasive diarrhea and liver disease (Petri 2002). Approximately 50 million people have invasive disease annually resulting in 100,000 deaths per year (WHO/PAHO/UNESCO report 1997). Several virulence factors have been identified that are important for parasite pathogenesis, however the majority of molecular mechanisms that the parasite

utilizes as it establishes invasive disease remains unclear. A number of essential biological issues have yet to be addressed in the study of parasite biology. For example, a genetically closely related but avirulent species *Entamoeba dispar* has been identified (Diamond and Clark 1993). However the genetic factors that make *E. histolytica* virulent and *E. dispar* avirulent, although an area of intense study, remain to be elucidated (Huston and Petri 1999). Additionally, the majority of virulence determinants that enable *E. histolytica* to cause invasive colonic and hepatic disease have not been identified and a vital aspect of the parasite's life cycle (developmental biology of trophozoite to cyst conversion) is not well understood (Stanley 2001; Petri 2002).

With the pending whole genome sequence of *E. histolytica* and the advent of the genomics tools, researchers can make significant strides in dissecting amoebic biology. Techniques such as DNA microarrays that allow rapid, quantitative and reproducible analysis on a genome-

Keywords. DNA microarrays; *Entamoeba dispar*; *Entamoeba histolytica*; genomic; pathogenesis

Abbreviations used: ORF, Open reading frame; PCR, polymerase chain reaction

wide scale are being developed and promise to further accelerate studies of amoebic pathogenesis. We believe that these microarrays will have a multitude of uses and will prove useful to a number of scientists in the *Entamoeba* scientific community. Microarray studies promise to unveil novel molecular aspects of amoebic pathogenicity and provide a number of genetic targets for improved diagnostic and treatment strategies.

2. DNA microarrays

A nucleic acid array or DNA microarray is used to monitor genome-wide analysis of DNA or RNA. This tool has been used to study genomic abundance or RNA abundance in a variety of parasites including *Plasmodium* spp (Hayward *et al* 2000; Ben Mamoun *et al* 2001; Rathod *et al* 2002), *Toxoplasma gondii* (Singh *et al* 2002), and *Leishmania* spp. The method uses nucleic acid samples (genomic DNA, cDNA, or oligonucleotides) that are mechanically spotted on silica coated glass slides, each of which can hold upwards of 30,000 spots. The DNA spotted on the slides (targets) is hybridized with genomic DNA or cDNA (probes) from differentially fluorescently labelled biological samples. The relative hybridization of each probe to the target gives an estimation of genomic DNA or message abundance for each clone, similar to Southern or Northern blot analysis. However the exceptional power of the microarray approach lies in the fact that each array with thousands of targets represents thousands of hybridizations which can be analysed in a quantitative manner. Microarray studies have been used in a number of diverse systems to study cancer biology (Han *et al* 2002; Nielsen *et al* 2002), drug resistance (Wilson *et al* 1999), gene expression changes in response to various stimuli (DeRisi *et al* 2000; Khodursky *et al* 2000; Dunman *et al* 2001), differences in DNA copy number (Pollack *et al* 1999), cell cycle (Futcher 2000), and transcriptional control (Weinmann *et al* 2002).

Microarrays are typically constructed based on sequence information obtained from genome sequencing projects and their most common use is in transcript profiling. Various applications of parasite biology including stage-specific developmental regulation, drug responses and resistance, host response to infection, and virulence have been studied using microarray technology. Among the protozoan parasites, this technology is currently being applied to study *Plasmodium falciparum* (Hayward *et al* 2000; Ben Mamoun *et al* 2001; Degraeve *et al* 2001; Rathod *et al* 2002), *Toxoplasma gondii* (Blader *et al* 2001; Singh *et al* 2002), *Leishmania* spp, *Cryptosporidium parvum* (Deng and Abrahamsen 2001; Straub *et al* 2002), *Trypanosoma cruzi* (de Avalos *et al* 2002) and *E. histolytica*.

3. Methods

3.1 *Entamoeba histolytica* microarray design

The Institute of Genomic Research (TIGR) and the Sanger Center are currently sequencing the *E. histolytica* genome; we are constructing a genomic DNA microarray of 10,000 sequenced clones from a random sheared genomic library of strain HM1 : IMSS (average insert size = 2 kb). The *E. histolytica* genome has short intergenic regions and relatively few and short introns making the use of genomic library for microarray development a feasible approach (Bhattacharya *et al* 2000; Wilihoeft *et al* 2001). A similar approach using genomic arrays from the parasite *Leishmania major* has been successfully used to identify developmentally regulated genes. For the *E. histolytica* microarray, the sequenced clones are polymerase chain reaction (PCR) amplified directly from bacterial stocks using primers (M13 forward and M13 reverse) that are part of the multiple cloning sites of the p-HOS vector (figure 1). The PCR products are electrophoresed on agarose gels to verify the presence of an insert, ethanol precipitated, re-suspended in water, dried down and re-suspended in a 3X SSC buffer. These products are then printed on poly-lysine coated glass slides. Our preliminary analysis shows that the use of genomic DNA microarrays is feasible for *E. histolytica* in studying genomic abundance as well as transcriptionally regulated genes. We are currently utilizing these arrays to identify species-specific differences between *E. histolytica* and *E. dispar*. Additionally, we are using these arrays to identify virulence determinants of *E. histolytica* that are involved in the host-pathogen interaction.

3.2 *Experimental design and data analysis*

The general design of microarray experiments is based on competitive hybridization of two fluorescently labelled probes to the targets spotted on the array. Experiments are often designated as "type I" or "type II". In type I, two experimental samples are labelled, one with a fluorophore dye that fluoresces green and the other with a dye that fluoresces red. The signal of the samples relative to each other is then quantitated. In type II experiments the first sample always represents a "reference" sample and the second sample is the experimental sample. The reference sample can consist of sheared genomic DNA, RNA from a mixture of steady-state conditions or, as in the case of the *E. histolytica* arrays, a piece of the vector that is common to all the PCR spots on the array. We have chosen type II experiments for our DNA microarrays as they allow normalization of the signal from the experimental sample to the signal from the reference

sample for each array. This approach has some advantages, notably the ability to perform inferred comparisons amongst experiments. Thus even though two experimental samples were never compared “head to head” the sample/reference signal for each clone can be compared and an inferred comparison can be made. An additional advantage of the reference that we have utilized (PCR product from the multi-cloning site) is that it labels the spotted PCR products in a linear manner to the amount of DNA spotted on the array. This allows estimation of relative probe abundance under a given condition, for example message abundance under log phase trophozoite conditions.

Following hybridization of the probe to the array, the arrays are scanned, labelled and analysed. A number of analytical tools and approaches are now freely available on the web for data analysis including:

Scanalyze	http://rana.lbl.gov/EisenSoftware.htm
Significance analysis of microarrays	http://www-stat.stanford.edu/~tibs/SAM/
Treeview	http://rana.lbl.gov/EisenSoftware.htm
Cluster	http://rana.lbl.gov/EisenSoftware.htm

Other websites with excellent information on array generation, use and data analysis include:

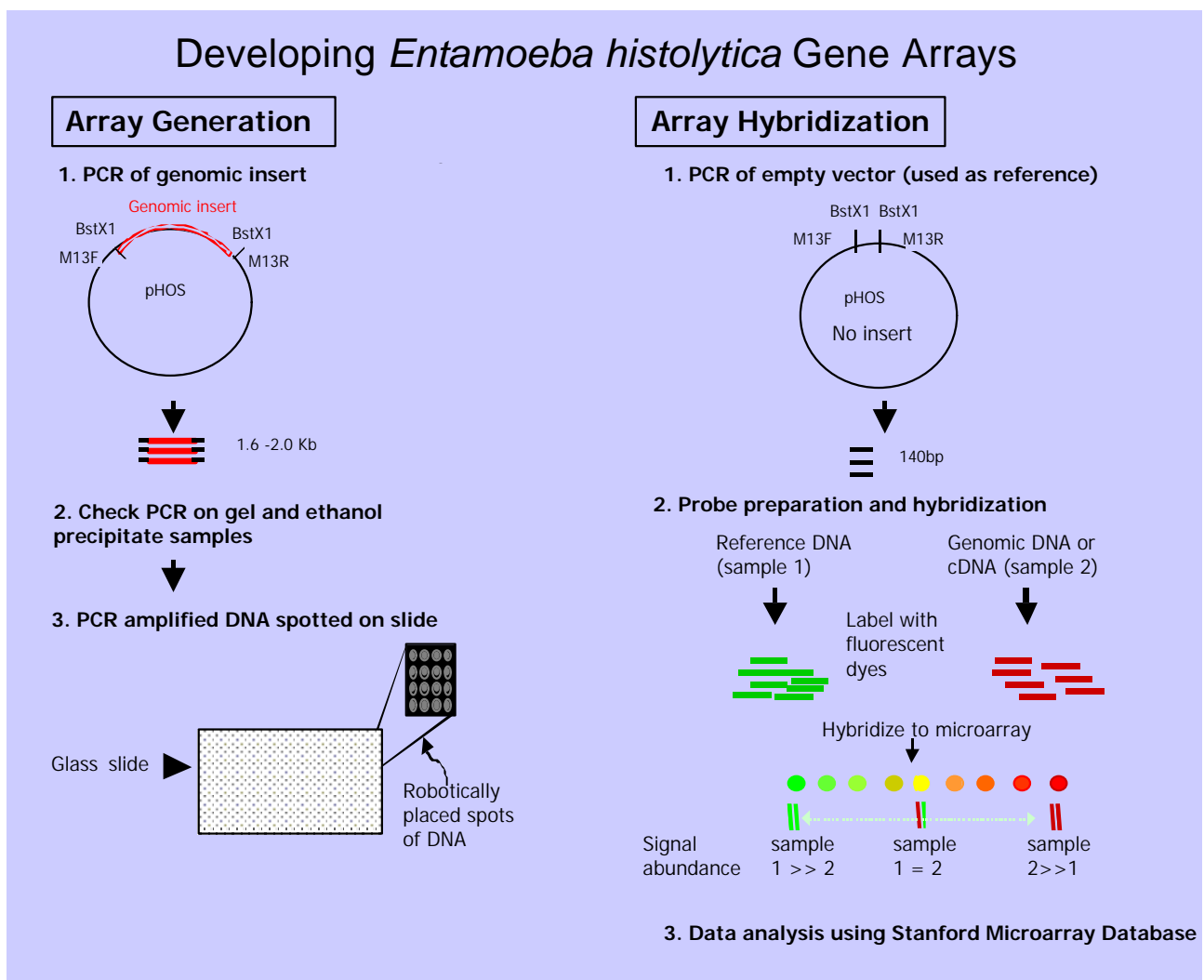


Figure 1. A schematic of *E. histolytica* array generation and hybridization. Amplification of the genomic insert is performed using primers in the multi-cloning site of the vector. The PCR products are then checked on an agarose gel, cleaned and spotted on glass slides. Array hybridization is performed with two samples which are differentially fluorescently labelled and signal abundance is quantitatively measured.

Stanford Microarray Database <http://genomewww5.stanford.edu/MicroArray/SMD/>
 J DeRisi lab <http://derisilab.ucsf.edu/microarray/index.html>
 P Brown lab <http://cmgm.stanford.edu/pbrown/>

Due to the extreme sensitivity of microarray technology, the biological experiments must be carefully planned and controlled. Apparently subtle differences in the culture conditions, growth phase of the parasites, the biologic phenomenon being tested, and method of RNA isolation and labelling can result in significant variability in microarray data. For each condition to be tested we perform a minimum of two biological experiments. Two independent microarray labellings and hybridizations are generated from each RNA sample; thus a minimum of four arrays are generated for each condition tested. Experimental variability is assessed and only those DNA clones

that give statistically reproducible results are functionally pursued. Confirmation of microarray data has typically been done with another method that assesses RNA abundance (Northern blot analysis or Real Time PCR) for a limited number of clones on the microarray. With the general use and acceptance of microarray data as being reliable and quantitatively accurate, it remains to be seen whether the scientific community will continue to require such confirmatory data in the future.

4. Results and discussion

4.1 Current studies with the *Entamoeba histolytica* array

4.1a Estimation of *E. histolytica* message abundance: We have used these arrays to estimate relative abundance of amoebic genes under log phase trophozoite conditions.

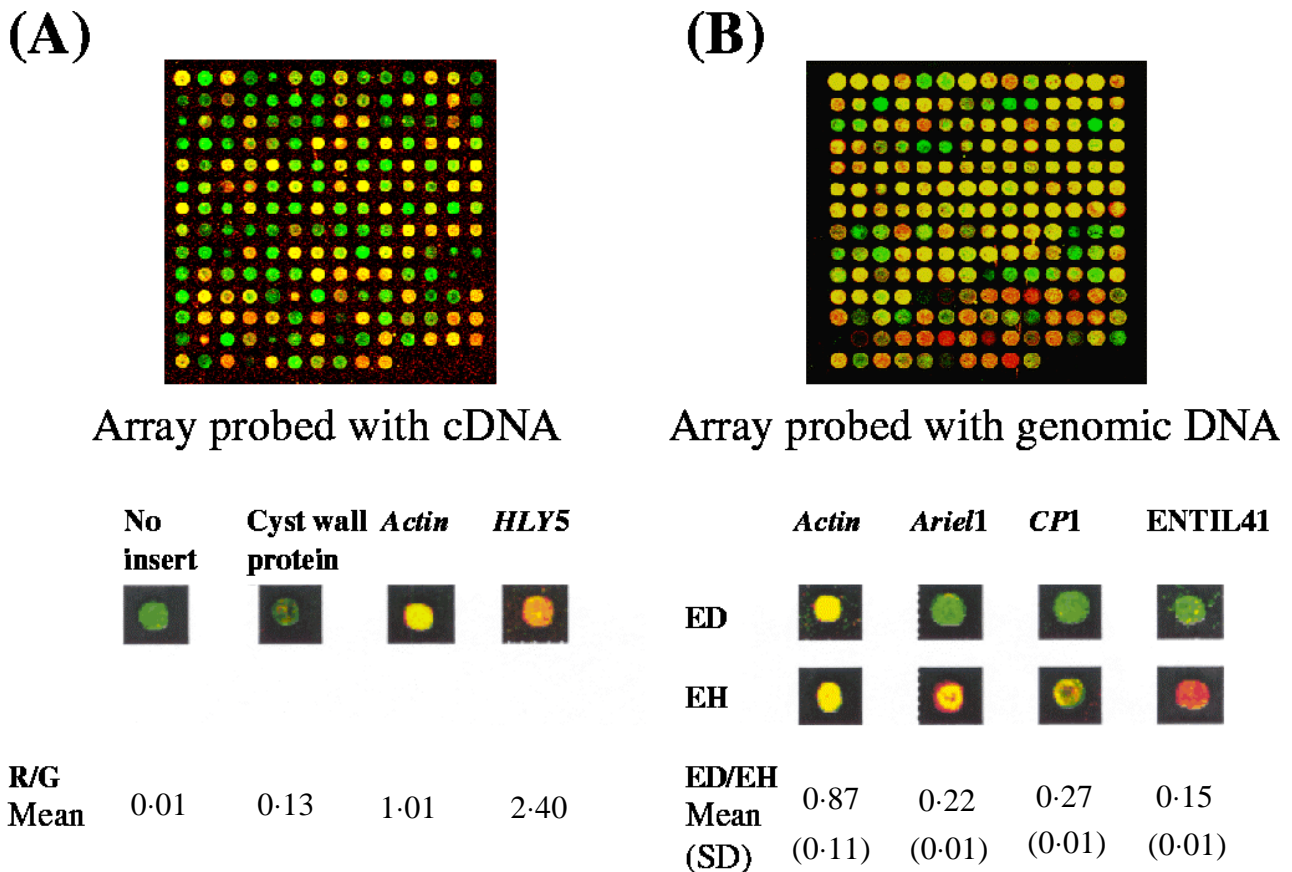


Figure 2. A representative quadrant of the *E. histolytica* microarray probed with (A) cDNA and (B) genomic DNA reveals good spot quality and good signal to noise ratio. (A) Message abundance of genes is estimated by the red/green (R/G) ratio with clones that have low abundance (cyst wall protein) under trophozoite conditions with low signal and clones with higher message abundance (HLY5) with higher signal. (B) An example of the genomic abundance of clones from *E. dispar* versus *E. histolytica* as assessed by hybridization to the microarray. Clones with equal genomic abundance (actin) have *E. dispar*/*E. histolytica* (ED/EH) ratios of ~ 1.0. Clones such as CP1 and Ariel1 which have previously been shown to be absent in *E. dispar* have low ED/EH ratios (≤ 0.5).

Preliminary data shows that our arrays are accurately able to assess message abundance (figure 2A) and additionally, have the ability to distinguish between members of gene families that have $\leq 80\%$ identity at the nucleotide level (data not shown). Therefore transcriptional changes that might arise in parasites under a variety of conditions (figure 3) should be able to be assessed using this methodology.

4.1b Estimation of genomic differences between *E. histolytica* and *E. dispar*: *Entamoeba histolytica* and the closely related non-pathogenic species *E. dispar* are indistinguishable morphologically, have similar cell biology and identical developmental stages, but have vastly different disease pathogenesis. *E. dispar* colonizes humans but never causes invasive disease and does not require treatment (WHO/PAHO/UNESCO report 1997). There have been several efforts to delineate the biological differences between the two species in order to identify the virulence determinants that make *E. histolytica* pathogenic compared to *E. dispar*. Such efforts however have not met with much success. To date only a handful of genes have been identified as being present in *E. histolytica* but absent in *E. dispar*. Since the genomic sequences of the

two parasites are so similar (orthologous coding and non-coding regions are 6.5% and 13.1% different respectively) microarray technology can be utilized to identify genes that are absent in *E. dispar*. We have begun to utilize the *E. histolytica* microarray to identify genetic differences between the two species. Preliminary results indicate that our arrays will be able to identify genomic clones that have a low genomic abundance or missing in *E. dispar* compared to *E. histolytica* (figure 2B). Thus genes such as *Ariell* and *CP1*, which have previously been shown to be absent in *E. dispar*, would have been accurately identified as being absent using microarray technology.

4.1c Identification of virulence determinants: The arrays can be used to study the progression of invasive colonic disease. As the parasite colonizes in the host large intestine, it first interacts with enteric bacteria, then invades the host colonic epithelium, adapts itself to varying oxygen tensions and ingests erythrocytes. We plan to use these arrays to dissect these aspects of amoebic pathogenesis and identify virulence determinants that are involved in causing colonic disease. Additionally, a number of diverse biological questions can be addressed using DNA microarray technology (figure 3).

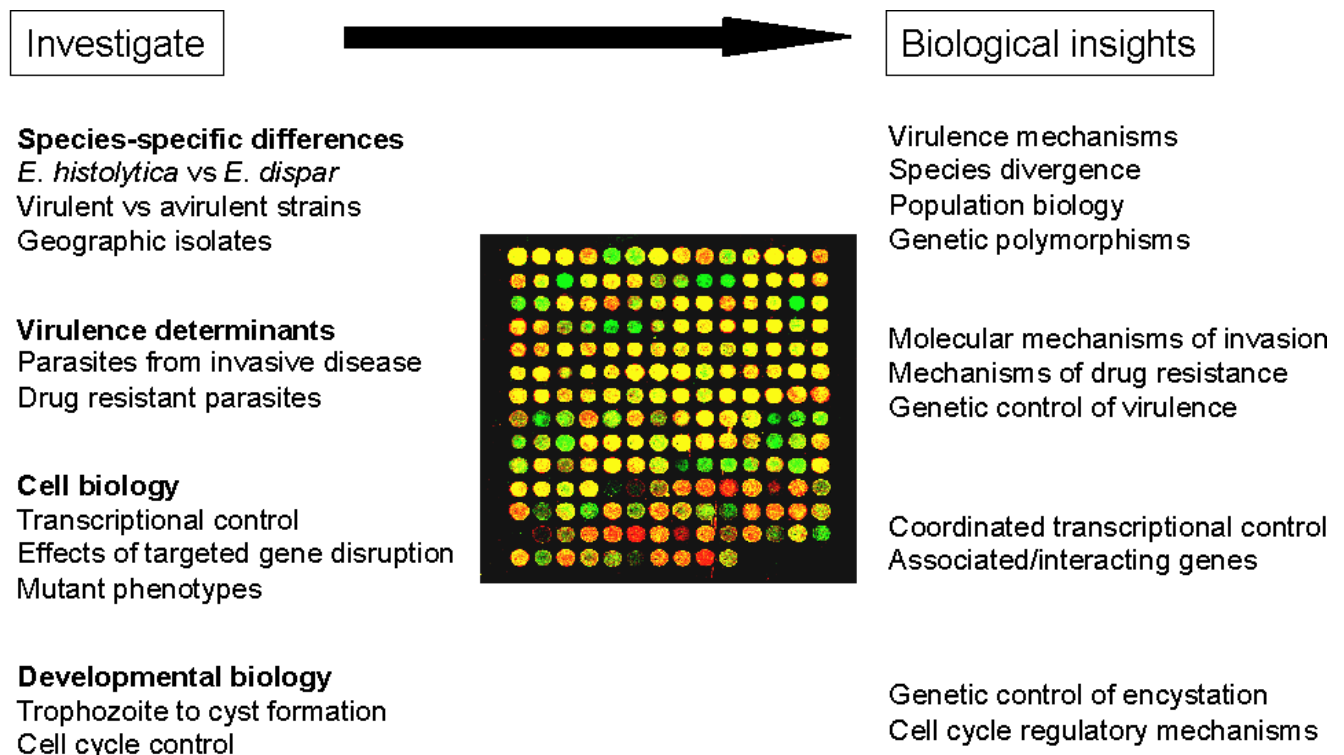


Figure 3. A schematic representation of the numerous biological questions and the possible insights in amoebic biology that can be gained using *E. histolytica* microarrays.

4.1d Limitations of microarray approach: While microarrays are an exceptionally powerful tool to study parasite biology on a genome-wide scale, a few limitations do exist. Most notably, studies of message abundance using RNA will identify genes as being differentially regulated only if they are controlled at the mRNA level. Therefore genes whose abundance is controlled using other means (such as alternative splicing, translational control or post-translation modifications) will not be identified. Additionally, these microarrays will not distinguish between probes that are greater than 80% identical at the nucleotide level. Thus genes that have $\geq 80\%$ nucleotide identity will cross-hybridize on the arrays. Since our arrays are generated from genomic fragments, some clones may have more than one open reading frame (ORF). In such clones if the expression pattern of the two ORFs is divergent (i.e. one has increased expression while the other has decreased expression under a given condition) the signals from the two clones will neutralize each other. However if both clones have similar expression profiles (both increase or both decrease under a given stimulus) then the signal will be accurately gauged. Our initial assessment of the number of clones on our array that have more than one ORF is 5%. Since our library is redundant we will be likely to have other clones on the array that represent the same genomic locus but have only one open reading frame.

4.1e Future challenges for amoeba research: Some challenges still remain for the study of amoebic biology. Biological samples that are obtained in small quantities (cysts from patients and parasites from *in vivo* animal models of liver or colonic disease) represent technical hurdles since obtaining 6 μg of total amoebic RNA (the amount of RNA we are currently using in our hybridizations) from these conditions remains challenging. However, newer techniques such as T7 amplification and global single cell reverse transcription polymerase chain reaction (GSC RT-PCR) have been developed. These methodologies allow generation of cDNA from limiting amounts of mRNA in a linear manner and are thus proportional to the initial mRNA abundance (Brail *et al* 1999; Pabon *et al* 2001). Additionally, some cell biological (*in vitro* encystation) and genetic tools (insertion into the genome and targeted disruption) still need to be developed for *E. histolytica* in order make best use of the genomics and post-genomics era in the study of amoebic biology.

This is an exciting time for amoebic research with a multitude of tools being developed based on the genome sequence. Analysis of amoebic pathogenesis with these genomic tools will undoubtedly facilitate the identification of genes that may be important virulence determinants or targets for immunologic or chemical therapy.

Acknowledgements

This work was supported by a Burroughs Wellcome collaboration award to US and an NIH sponsored divisional grant (T 32 AI07502) to PS.

References

- Ben Mamoun C, Gluzman I Y, Hott C, MacMillan S K, Amarakone A S, Anderson D L, Carlton J M, Dame J B, Chakrabarti D, Martin R K, Brownstein B H and Goldberg D E 2001 Co-ordinated programme of gene expression during asexual intraerythrocytic development of the human malaria parasite *Plasmodium falciparum* revealed by microarray analysis; *Mol. Microbiol.* **39** 26–36
- Bhattacharya A, Satish S, Bagchi A and Bhattacharya S 2000 The genome of *Entamoeba histolytica*; *Int. J. Parasitol.* **30** 401–410
- Blader I J, Manger I D and Boothroyd J C 2001 Microarray analysis reveals previously unknown changes in *Toxoplasma gondii*-infected human cells; *J. Biol. Chem.* **276** 24223–24231
- Brail L H, Jang A, Billia F, Iscove N N, Klamut H J and Hill R P 1999 Gene expression in individual cells analysis using global single cell reverse transcription polymerase chain reaction (GSC RT-PCR); *Mutat. Res.* **406** 45–54
- de Avalos S V, Blader I J, Fisher M, Boothroyd J C and Burleigh B A 2002 Immediate/early response to *Trypanosoma cruzi* infection involves minimal modulation of host cell transcription; *J. Biol. Chem.* **277** 639–644
- Degrave W M, Melville S, Ivens A and Aslett M 2001 Parasite genome initiatives; *Int. J. Parasitol.* **31** 532–536
- Deng M and Abrahamsen M S 2001 Microarray-based expression analysis of human epithelial cell response to *Cryptosporidium parvum* infection; *J. Eukaryot. Microbiol.* 42S–43S
- DeRisi J, van den Hazel B, Marc P, Balzi E, Brown P, Jacq C and Goffeau A 2000 Genome microarray analysis of transcriptional activation in multidrug resistance yeast mutants; *FEBS Lett.* **470** 156–160
- Diamond L S and Clark C G 1993 A redescription of *Entamoeba histolytica* Schaudinn 1903 (Emended Walker 1911) separating it from *Entamoeba dispar* Brumpt 1925; *J. Eukaryot. Microbiol.* **40** 340–344
- Dunman P M, Murphy E, Haney S, Palacios D, Tucker-Kellogg G, Wu S, Brown E L, Zagursky R J, Shlaes D and Projan S J 2001 Transcription profiling-based identification of *Staphylococcus aureus* genes regulated by the agr and/or sarA loci; *J. Bacteriol.* **183** 7341–7353
- Futcher B 2000 Microarrays and cell cycle transcription in yeast; *Curr. Opin. Cell. Biol.* **12** 710–715
- Han H, Bearss D J, Browne L W, Calaluce R, Nagle R B and Von Hoff D D 2002 Identification of differentially expressed genes in pancreatic cancer cells using cDNA microarray; *Cancer Res.* **62** 2890–2896
- Hayward R E, Derisi J L, Alfadhli S, Kaslow D C, Brown P O and Rathod P K 2000 Shotgun DNA microarrays and stage-specific gene expression in *Plasmodium falciparum* malaria; *Mol. Microbiol.* **35** 6–14
- Huston C D and Petri W A 1999 Amebiasis: Clinical Implications of the Recognition of *Entamoeba dispar*; *Curr. Infect. Dis. Rep.* **1** 441–447
- Khodursky A B, Peter B J, Cozzarelli N R, Botstein D, Brown P O and Yanofsky C 2000 DNA microarray analysis of gene expression in response to physiological and genetic changes

- that affect tryptophan metabolism in *Escherichia coli*; *Proc. Natl. Acad. Sci. USA* **97** 12170–12175
- Nielsen T O, West R B, Linn S C, Alter O, Knowling M A, O'Connell J X, Zhu S, Fero M, Sherlock G, Pollack J R, Brown P O, Botstein D and van de Rijn M 2002 Molecular characterisation of soft tissue tumours: a gene expression study; *Lancet* **359** 1301–1307
- Pabon C, Modrusan Z, Ruvolo M V, Coleman I M, Daniel S, Yue H and Arnold L J Jr 2001 Optimized T7 amplification system for microarray analysis; *Biotechniques* **31** 874–879
- Petri W A 2002 *Entamoeba histolytica*: Clinical Update and Vaccine Prospects; *Curr. Infect. Dis. Rep.* **4** 124–129
- Pollack J R, Perou C M, Alizadeh A A, Eisen M B, Pergamenschikov A, Williams C F, Jeffrey S S, Botstein D and Brown P O 1999 Genome-wide analysis of DNA copy-number changes using cDNA microarrays; *Nat. Genet.* **23** 41–46 (java/Propub/genetics/ng0999_0941.fulltext java/Propub/genetics/ng0999_0941.abstract)
- Rathod P K, Ganesan K, Hayward R E, Bozdech Z and DeRisi J L 2002 DNA microarrays for malaria; *Trends Parasitol.* **18** 39–45
- Singh U, Brewer J L and Boothroyd J C 2002 Genetic analysis of tachyzoite to bradyzoite differentiation mutants in *Toxoplasma gondii* reveals a hierarchy of gene induction; *Mol. Microbiol.* **44** 721–733
- Stanley S L 2001 Pathophysiology of amoebiasis; *Trends Parasitol.* **17** 280–285
- Straub T M, Daly D S, Wunshel S, Rochelle P A, DeLeon R and Chandler D P 2002 Genotyping *Cryptosporidium parvum* with an hsp70 Single-Nucleotide Polymorphism Microarray; *Appl. Environ. Microbiol.* **68** 1817–1826
- Weinmann A S, Yan P S, Oberley M J, Huang T H and Farnham P J 2002 Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis; *Genes Dev.* **16** 235–244
- WHO/PAHO/UNESCO report 1997 A consultation with experts on amoebiasis. Mexico City, Mexico 28–29 January 1997; *Epidemiol. Bull.* **18** 13–14
- Wilihoeft U, Campos-Gongora E, Touzni S, Bruchhaus I and Tannich E 2001 Introns of *Entamoeba histolytica* and *Entamoeba dispar*; *Protist* **152** 149–156
- Wilson M, DeRisi J, Kristensen H H, Imboden P, Rane S, Brown P O and Schoolnik G K 1999 Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization; *Proc. Natl. Acad. Sci. USA* **96** 12833–12838