
Genomics and post-genomics in parasitology: genome babble or a real opportunity?

K. Gull¹

School of Biological Sciences, University of Manchester, 2.205 Stopford Building, Oxford Road, Manchester M13 9PT, U.K.

Abstract

The genome projects represent one of the most important developments in our knowledge of parasites. However, translation of this knowledge into an understanding of parasite biology and then on to drugs, vaccines and other healthcare developments for the diseases will need some élan and clarity of thought by scientists and funding organizations. Only then will the activity associated with post-genomics be turned from what I have termed 'genome babble' to real opportunities in understanding these parasites.

Introduction

The developed world rightly worries about the twin effects over the last 20 years of the slowing down in the discovery of novel types of antibiotic and the rise of antibiotic resistance. By contrast, many of the major disease threats of developing countries still await the possibility of intervention by modern drugs, antibiotics or vaccines. The World Health Organisation's statistics on these diseases make sobering reading. In the case of malaria, there are estimates of 300 million cases with 2400 million people at risk. Each of the parasitic protozoan diseases, African trypanosomiasis, leishmaniasis and Chagas disease, carry their own frightening statistics as do those caused by parasitic worms: schistosomiasis, lymphatic filariasis and onchocerciasis. Given the healthcare economics of countries affected by these parasitic diseases, drug-discovery programmes targeted against them are clearly not priorities for major pharmaceutical companies. This, and the difficulties with intervention in the parasitic life style of the aetiological agents, have often been rehearsed as major reasons for the absence of modern, effective drugs and vaccines for these diseases. Personally, I suspect that this is only a part of the answer. Probably equally important is the fact that the quality of academic science output in these areas has been patchy over the last 50 years and certainly not generally as strong as that in, for

instance, bacteriology. Again of course, there are many reasons for this, some to do with the science and some to do with the scientists. However, the provision of complete or survey genome information on the parasites should be a great advance in our knowledge and for many of these parasites the information should be with us soon. Interestingly, it brings with it a new set of challenges and clarifies many of the older ones, both for the science and the scientist!

The parasite genome projects: the expected, the likely and the orphan genes

Genome projects are well advanced for *Plasmodium*, *Trypanosoma brucei*, *Leishmania* and *Toxoplasma*, with those for other protozoan parasites and worms being developed rapidly [1]. In the following I will restrict myself to discussion of the protozoan parasites in the main. The genomes of these parasites are large (around 30–40 Mb) and at one time may have been thought of as difficult projects. The present speed of sequence acquisition itself and of informatics assembly and analysis of the product, as shown in the *Drosophila* [2] and human genome projects, now show these to be very tractable targets. Given the nature of genome organization, transcriptional control and processing in the parasitic protozoa we can expect these genomes to be 'gene-dense'. The 12-Mb genome of yeast revealed around 6000 genes on completion of the genome project. We will undoubtedly be looking at more genes in the protozoan parasites. The obvious initial challenge will be to use bioinformatics to map the terrain in terms of genes whose sequences give some clues to their identity or function. Given the increasing sophistication of bioinformatic approaches then one might be hopeful that the parasite projects will benefit from what has gone before and that a relatively deep analysis will emerge initially by this route.

However, it is salutary to remember that around 40–60% of all genes identified in each of the many bacterial genome projects have had to be labelled 'hypothetical protein' on first annotation.

Key words: genome projects, parasite biology, protozoa.
¹e-mail K. Gull@man.ac.uk

Table 1

Orphan genes in parasite genomes: insights into new functions?

Parasitism
Virulence
Survival in environment
Host interactions
Vector interactions
Nuclear and organelle genome plasticity, expression and interaction
Extent, global control and integration of metabolism
Cytoplasmic regionalization and specialization
Cell structures: a new molecular cytology
Signalling and communication: in space and time
The slow growth or no growth condition
Insurance policies and alternative life-cycle pathways

In the yeast genome project this figure of 'orphan' genes without identity or without even *in silico* clues to biochemical or cellular function was 56% of the total; 3480 genes! The figures for *Escherichia coli* were 60% of the genes; 2583 genes! Given the cellular complexity of the parasitic protozoa it is clear that we must look forward to similar figures for the initial analysis of the parasite genomes. This will set up a series of interesting tensions for the molecular parasitologist seeking realistic drug or vaccine targets or, alternatively, wishing to study pathogenicity/life-cycle/virulence determinates. Comparative genomics of the set of parasite genes with 'known' human homologues may provide one useful data set in which differences in parasite proteins flag up opportunities. Identification of 'parasite-specific' genes not present in the human genome provides other opportunities. The difficulty here will be their unknown function and unknown impact on parasite vigour. Getting biologically relevant and meaningful information on these (probably highly numerous) genes will be an even greater task in parasite post-genomics than it has been for model organisms. Before considering how one might approach such a task it is worthwhile thinking about what functions such orphan genes might influence in the parasitic protozoa. In Table 1 I have outlined some thoughts on what we might expect. The orphan gene repertoire should contain sets of genes that have critical functions in parasitism, influencing both survival and virulence in host and vector. Since, the cyto-architecture of protozoan parasites displays many specializations and their genomes (both nuclear and cytoplasmic)

have intriguing properties and expression patterns then we might expect orphan genes to contribute here. Table 1 lists other possibilities including the prediction that orphan genes amongst these protozoan parasites and other microbes will be involved in what I have termed the 'slow growth or no growth' state. For the last 200 years most of our studies of microbes, both pathogenic and free-living, have relied on assays that, in effect, demand growth of the organism. With notable exceptions we have not asked about the no growth or very slow growth condition. Yet it is likely that many microbes in their natural environment spend a great deal of their time in this condition. Parasites are no different and a successful traverse of the life cycle may well involve long periods of survival in this state. Given that microbiology, biochemistry, and cell and molecular biology of microbes have focused for so long on situations where the organism is undergoing rapid growth it appears likely that we have selected rather heavily for processes and genes that function under such conditions. Changing our focus a little to understand the no growth/very slow growth condition and to think of assays that involve or interrogate this condition is likely to be both difficult and rewarding.

Gene to function: the post-genomics arsenal and the problem of phenotype

In moving from gene identification to function for the massive set of genes displayed for a particular parasite genome we are faced with a number of issues. Probably the first of these is to accept that the word function means different things to different people! We will no doubt wish to understand the function of genes at particular levels. In some cases this will initially be at the molecular level of the protein, building then to protein function in assemblies or cellular compartments/organelles. However, others may see an understanding of function encompassing cell types, strains, parasite populations and environmental influences leading through to evolutionary considerations. Thus our level of interrogation of function of genes within the parasite genome will be influenced by many factors, including the art of the possible.

Molecular-genetic analysis of parasites has moved quickly over the past decade and there is a reasonable collection of approaches available for post-genomic analysis of parasite genomes. Naturally, the level of sophistication varies with each parasite. However, gene-knockout, anti-sense,

RNAi (RNA-mediated interference) and conditional-expression approaches are now in place or are rapidly being developed for many of the major protozoan parasites. These technologies will not only be important for academic studies of parasite functions but will be critical to the development of target-validation studies in drug-therapy initiatives. Starting out to design an effective drug for a non-essential gene function is clearly a waste of effort! It is now possible to address these questions directly.

The application of array technologies could be extraordinarily powerful for assessment of global patterns of gene expression, particularly those occurring during stages of parasite life cycles that are refractory to normal experimental interrogation. Gaining an early, yet clear overview of the global expression pattern of the genome within vector and host stages of the parasite will be a major target. Such a view will be key to defining likely drug and vaccine candidates. Again, starting to design an effective drug for a gene product that is not known to be expressed at the particular target parasite stage *in vivo* is a waste of effort. It remains to be seen whether array technologies or proteomic analyses will have the greater impact in revealing global patterns of gene expression. The emphasis may vary with different parasites given their differing reliance on transcriptional, post-transcriptional and translational levels of gene regulation.

What is certain is that these techniques are, and will increasingly be, applied to the study of protozoan parasites. These global analyses bring with them many opportunities. They will find application not only in analysis of life-cycle stages but also of mutant parasites and the parasite's response to experimental insults such as drug treatments, environmental switches, etc. They also bring with them some difficulties. Analyses of such experiments will, more than usually, depend upon rigorous registration of growth conditions, physical and media factors, strains and clones used, parasite cell types, life-cycle stage, etc. Without such registration of experimental conditions we are likely to amass a lot of uninterpretable data!

In particular, the ability to interrogate parasite gene function in vector and host will lead to interesting discussions of what is our working definition of a 'wild-type' parasite strain and what differences might emerge between these and our commonly used laboratory varieties. Post-genomic analyses at the level of the individual gene or

Table 2**Read-outs from the parasite genome projects**

First level

Genome composition and evolution
 Molecular karyotype
 Chromosomal rearrangement
 Genome plasticity
 Synteny
 Ploidy significance
 Multigene families: maintenance and diversity
 Repetitive sequences
 Promoters?
 Processing descriptors
 cis and *trans* splice sites
 Poly(A)⁺ sites
 Centromeres
 Telomeres
 Organelle dependency
 Differential gene expression
 Antigenic variation
 Redundancy

Second level

Insights into:
 New basic biological phenomena
 Parasitism
 Virulence
 Evolution
 Epidemiology
 Complementation of hypothesis-driven research
 Added value and efficiency to all studies of parasites
 Translational science
 Diagnostics
 Drugs
 Vaccines
 Intervention opportunities

the global genome are likely to lead to us re-examining our laboratory *in vitro* models and looking to better, more defined animal host and vector models of disease.

Perhaps the most important challenge for post-genomic studies of protozoan parasites is that of defining phenotype. In fact, we are probably better equipped to make mutants and knockout genes and to analyse gene-expression patterns than we are to analyse complex phenotypes. We will need to invest heavily in defining new approaches to phenotype analysis in the protozoan parasites if we wish to reveal the function of the large number of orphan genes expected in these genomes.

Through genome babble to real opportunities

The excitement of the genome projects has brought with it a massive change in the language of science. While not wishing in any way to deflect this excitement about a post-genomics revolution in molecular parasitology it is probably wise to offer a note of caution. One now often reads articles and other types of molecular parasitology 'literature' that are littered with this language of genomics: from transcriptome to targetome, from MALDI-MS to metabolome and from arrays to data mining! At times this can appear to be something I have termed genome babble (the dictionary definition of babble being; to utter words in an incoherent or indistinct jumble; to talk foolishly, incessantly or irrelevantly).

Given the extra difficulties that analysis of parasite genomes will present and the insight and care that will need to be invested in experimental design, we should perhaps be aware of not overselling the immediate read-outs from these projects into healthcare outcomes.

However, Table 2 illustrates what one might consider to be some of the realistic first-level and second-level read-outs and outcomes from the parasite genome projects. I merely indicate here some of the aspects of parasite biology that are likely to be made clearer by the gain of complete genome sequence. Undoubtedly, the genome projects will influence and enhance what I have termed second-level read-outs. However, ensuring that the basic science read-outs of the first level translate through to some of those in the second level will need a culture change in many academic laboratories and funding agencies. The applied and translational science of healthcare for these protozoan diseases will need industry-standard analyses for assessment of target validation and will require industry-standard approaches to project management. In addition, the interface between science and clinical science will need strengthening in tropical medicine. Finally, much will rest on the success of public/private strategic alliances to extend to concepts of drug and vaccine discovery and development the paradigm programmes that are now in place for selective drug delivery for some tropical diseases, such as onchocerciasis and lymphatic filariasis (<http://www.who.int/tdr/grants/workplans/filariasis4.htm>).

The consequential sociology of genome projects

Finally, it seems worthwhile pointing out to the individual scientist what consequences the parasite genome projects might have for their own behaviour. I call this the consequential sociology of the genome projects and clearly it is not restricted to this field! First I will suggest that the quality of questioning and reasoning in the field will increase rapidly. No longer will one be able to work on, say, one's favourite enzyme without defining why that particular member of the extensive protease or kinase or 'whateverase' family is more interesting than the other 30 in the database. The quality and burden of proof will also rise rapidly. Reviewers of papers will require more extensive analysis and comparison; let-out clauses such as 'this result may be explained by the presence of another similar gene' will disappear as reviewers ask 'which one?' There will be a loss of ownership on areas of study. Laboratories who have traditionally worked in areas of parasite biology will find many other groups publishing novel insights in their area originating from genome analysis. Some of these will come from groups not even associated previously with work in that parasite. The speed of doing science in this area will increase; there will be more consortia operating alongside the hypothesis-driven science of individuals. It may also be that the parasite genome projects may lead to organism restriction, whereby basic scientists choose to work only on those organisms with a completed genome project. Finally, what is sure is that the parasite genome projects will have enormous impact on both the scientists working in this field and on our understanding of the parasites. There are many challenges ahead if the knowledge that flows from the projects is to be translated to effective understanding of the parasites and on to successful healthcare measures for these diseases.

References

- 1 Special edition: Parasite genome projects (2000) *Int. J. Parasitol.* **30**, 327–565
- 2 The genome sequence of *Drosophila melanogaster* (2000) *Science* **287**, 2185–2204

Received 31 July 2000