Genomics and the biology of parasites

David A. Johnston,1* Mark L. Blaxter,2 Wim M. Degraeve,3 Jeremy Foster,4 Alasdair C. Ivens,5 and Sara E. Melville6

Summary
Despite the advances of modern medicine, the threat of chronic illness, disfigurement, or death that can result from parasitic infection still affects the majority of the world population, retarding economic development. For most parasitic diseases, current therapeutics often leave much to be desired in terms of administration regime, toxicity, or effectiveness and potential vaccines are a long way from market. Our best prospects for identifying new targets for drug, vaccine, and diagnostics development and for dissecting the biological basis of drug resistance, antigenic diversity, infectivity and pathology lie in parasite genome analysis, and international mapping and gene discovery initiatives are under way for a variety of protozoan and helminth parasites. These are far from ideal experimental organisms, and the influence of biological and genomic characteristics on experimental approaches is discussed, progress is reviewed and future prospects are examined. BioEssays 1999;21:131–147.

Prologue
The need for parasite genome analysis
Despite the medical and healthcare revolution of the late 20th century, billions of people still suffer from one or more tropical parasitic disease and the constant drain imposed by chronic sickness, loss of productive labour and premature death, imposes a multibillion dollar restriction on the economic development of the Third World.(1,2)

Amongst the most severe parasitic diseases, targeted for special attention by the World Health Organization (WHO) are malaria, leishmaniasis, Chagas disease (American trypanosomiasis), and African sleeping sickness (African trypanosomiasis), caused by protozoan parasites, and schistosomiasis (Bilharzia) and cutaneous and lymphatic filariasis, caused by metazoan helminth parasites (Table 1). These diseases are intimately linked with poverty and persist through the complex intertwining of socioeconomic factors (inadequate healthcare, poor housing and sanitation, political inertia, difficult communications, etc.) with biological factors (repeat infection, drug resistance, malnutrition, immunosuppressive effects of HIV/AIDS, emergence of disease vectors resistant to chemical control, etc.). Economic development itself plays a role through environmental change that favours the spread of, and human contact with, vector species, through human migration (with immunologically naive people moving into endemic areas and infected people spreading disease to new localities), and through uncontrolled (shanty-town) urbanisation.

Current control measures rely heavily on pharmacologic intervention to alleviate symptoms and reduce parasite transmission, coupled to vector control. However, widespread...
<table>
<thead>
<tr>
<th>Disease and causative organism</th>
<th>Vector/intermediate host</th>
<th>People infected (millions)</th>
<th>Disease cases/year (millions)</th>
<th>Deaths/year (millions)</th>
<th>At risk (millions)</th>
<th>Distribution</th>
<th>Main control problems</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Protozoan parasites</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African sleeping sickness</td>
<td>Tsetse fly</td>
<td>Unknown</td>
<td>0.25–0.3</td>
<td>0.225–0.27 I invariably fatal if untreated (90% of cases)</td>
<td>55</td>
<td>36 countries of sub-Saharan Africa</td>
<td>• Insufficient medical surveillance • Premature cessation of control programmes • Drug toxicity</td>
</tr>
<tr>
<td><em>Trypanosoma brucei</em> subssp. (kinetoplastid protozoan)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Chagas Disease (American trypanosomiasis) | Triatome bugs | 16–18 | 6–10 million suffer chronic disease | Unknown, but infection is fatal in 10–40% of chronic cases | 90–100 (25% of S. and Cent. America) | S. USA, Cent. and S. America | • Zoonotic infections • Blood transfusions |
| *Trypanosoma cruzi* (kinetoplastid protozoan) |

| Leishmaniasis | Sandflies | 12–15 | 2 | Unknown | 350–370 | 88 countries of N. Africa, Middle East, and S. America | • Increase in vector populations as malaria vector control programs are wound down • Immunosuppressive effects of HIV • Prolonged drug treatment/ hospitalisation required |
| *Leishmania* sp. (13 species) (kinetoplastid protozoan) |

| Malaria | Mosquitoes | Unknown | 300–500 | 1.5–2.7 | 2,500 (40% global population) | 98 countries, mainly in sub-Saharan Africa, also Arabia, Latin America, S. E. Asia | • Drug resistance in parasite • Insecticide resistance in vector • Migration and travel |
| *Plasmodium* sp. (4 species) (apicomplexan protozoan) |

| Metazoan helminth parasites | Freshwater snails | 250 | 20 | 0.3–0.5 | 600 | 76 countries of Africa, S. and Cent. America, Caribbean, India, China and S. E. Asia | • Water development programmes • Zoonotic infections • Drug resistance • Migration |
| *Schistosomiasis* (Bilharzia) *Schistosoma* sp. (5 species) (trematode platyhelminth) |

| Cutaneous and lymphatic filariasis | Mosquitoes and other arthropods (black-flies, mites, copepods) | 150 | 43 | Unknown, but small | 1,100 | 73 countries of sub-Saharan Africa, S. and Cent. America, Caribbean, India, China, S. E. Asia, and the Pacific | • Potentially eradicable |
| 7 species in 5 genera including *Brugia*, *Onchocerca* and *Wuchereria* (filarial nematodes) |
(and improper) use of the few treatments that were previously cheap, effective, and easy to administer has resulted in the appearance and spread of resistant parasites and vectors. Even where drug resistance is not yet a serious problem, available treatments may require long-term drug administration and/or hospitalisation, involve regimens that have not been standardised or use compounds that are subcurative, invoke severe (sometimes fatal) side effects or may not even be officially licensed. In the longer term, vaccines will provide a major impact but factors such as antigenic switching, mutation, diversity, autoimmune stimulation and poor antigenicity have limited progress and none are currently available for these six diseases.

The development of new drugs that are both safe and effective and the identification of new vaccine antigens are urgent needs. However, eukaryotic parasites are far from ideal experimental organisms, and the mass screening techniques available for prokaryote pathogens are rarely applicable. Therefore, rational design strategies, based on a sound understanding of how parasites develop, survive, and reproduce in their different hosts, of parasite-host and parasite-immune system interactions and of the factors that determine behaviour, pathogenicity, drug resistance and antigenic variation represent the only practical approach to identifying new therapies. Sequencing those parts of the genome that encode this information is an urgent need, and the past few years have seen the initiation of Parasite Genome Initiatives to facilitate the acquisition, analysis, and distribution of such data.

In this article, representatives from several of the Parasite Genome Initiatives discuss the rationale behind current approaches, report on progress, and predict future developments as information accumulates and research shifts from genomic data generation to postgenomic analysis. Genome projects break the mould of conventional scientific practice by being collaborative efforts, performed with a service function in mind and with the wish to disseminate data as widely and rapidly as possible (much of which is inappropriate for formal publication). Thus, extensive reference will be made to information sources on the Internet. It is not intended to provide either a “blow-by-blow” account of the activities under way within each of the Parasite Genome Initiatives or detailed descriptions of genome analysis techniques. For these, the reader is referred to other articles in this issue, to the general literature, and to the review-based publications and Worldwide Web (WWW) resources generated by the Initiatives (Table 2).4–17

### Overview of the parasite genome projects

Of the six Parasite Genome Projects discussed here, the Malaria Genome Project was initiated with funding from the Wellcome Trust and has since attracted some US$ 25,000,000 of governmental and charitable support, potentially enough to sequence the entire genome. The Schistosoma, Brugia malayi (as a model filarial nematode), Trypanosoma brucei, T. cruzi and Leishmania Initiatives fall within the WHO’s Special Programme for Research and Training in Tropical Disease. WHO’s resources are extremely limited and its funding is largely “pump-priming” to facilitate development of resources, strategies and collaborations to support applications to national or international agencies for larger scale projects. This “pump-priming” is beginning to bear fruit with the Leishmania and T. brucei Genome Initiatives recently attracting multimillion dollar support.

Briefly stated, the main aims of all the initiatives are: (1) to increase knowledge of parasite molecular biology, especially with respect to mechanisms of drug resistance, antigenic variation, and genetic diversity; (2) to identify genes with key cellular functions that could represent new drug targets and to identify antigens with diagnostic and/or vaccine potential; (3) where practical, to develop a physical map of the parasite’s

---

**TABLE 2. Internet Information Sources for Parasite Genome Projects**

<table>
<thead>
<tr>
<th>Parasite Genome Initiative WWW sites</th>
<th>Malaria Genome Network WWW sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>P. falciparum gene sequence tag project: <a href="http://parasite.arf.ufl.edu/malaria.html">http://parasite.arf.ufl.edu/malaria.html</a></td>
<td></td>
</tr>
<tr>
<td>P. falciparum genome sequencing project: <a href="http://www.sanger.ac.uk/Projects/P_falciparum/">http://www.sanger.ac.uk/Projects/P_falciparum/</a></td>
<td></td>
</tr>
</tbody>
</table>

Leishmania Genome Network WWW site: http://www.ebi.ac.uk/parasites/leish.html

Trypanosoma brucei Genome Network WWW site: http://parsun1.path.cam.ac.uk

Trypanosoma cruzi Genome Network WWW site: http://www.ddbm.fiocruz.br/genome/tcruzi/tcruzi.html

Schistosoma Genome Network WWW site: http://www.nhm.ac.uk/hosted_sites/schisto/

Filarial Genome Network WWW sites

UK: http://helios.bio.ed.ac.uk/mbx/fgn/filgen1.html

US Mirror: http://math.smith.edu/~sawl/fgn/filgen.html

---

**WHO Coordinating sites**

WHO Parasite Genome WWW site: http://www.ebi.ac.uk/parasites/parasite-genome.html (includes a BLAST-server for individual parasite datasets)

WHO Parasite Genome FTP site: ftp.ebi.ac.uk/pub/databases/parasites/

WHO-TDR Parasite Genome Committee: http://www.who.ch/tdr/ workplan/genome.htm

Parasite Genome Mailing List Archive: http://www.mailbase.ac.uk/lists-p-t/parasite-genome/

**Parasite Genome Databases**

In addition to being available for local installation (follow links in the above sites for information), the WHO-TDR Parasite Genome Databases are available, on-line, to registered users of the Human Genome Mapping Project Resource Centre.4–17
A further aim of the WHO Initiatives is to equip scientists from disease-endemic countries with expertise in genomics. The Developing World contains 70% of the world population and 24% of the scientific community, yet receives only 5% of global research funding and contributes to 3% of indexed research publications. Most of the large-scale biotechnology initiatives, including the model organism genome projects, were conceived as domestic initiatives by developed countries and the WHO Initiatives represent one of the few opportunities available to Third World countries to collaborate in long-term, far-reaching projects and to acquire the expertise necessary to exploit advances in biotechnology for national development (Fig. 1). The Malaria Genome Project has been, to date, an exclusively “First World” project involving the United States, United Kingdom, and Australia although the importance of involving scientists from endemic countries is recognised.

Problems
Through advances in laboratory technology, automation and informatics, it is now possible to rapidly acquire genomic data for any organism considered of sufficient importance to justify the effort. In an ideal world, genome analysis would encompass karyotype analysis, chromosome mapping, physical mapping, genetic mapping, gene discovery, and genomic sequencing approaches allied to ongoing informatics and functional analysis. Eukaryotes such as the yeast Saccharomyces cerevisiae and the free-living nematode Caenorhabditis elegans have been selected for detailed genome analysis precisely because they represent ideal “model” organisms. They are easy and cheap to maintain in the laboratory, they show rapid growth/development, they can be obtained in very large numbers, they are amenable to diverse experimental manipulations and analyses and hundreds of strains have been identified and characterised in detail, allowing a vast amount of information on biochemistry, behaviour, development, genetics, etc., to be accumulated to underpin genomic analyses. Moreover, their genomes are organised in a “conventional” manner. Parasites, however, belong in the “real world,” they often require protracted animal passage for maintenance of all, or part, of the life cycle, many cannot be cultured in vitro, restricting experimental manipulations, certain stages of the life cycle may be available in extremely limited quantities (for example, restricting the amounts of material available for cDNA library construction), and their genomes may display unique features that complicate analysis. Being human pathogens, their study is also frequently subject to strict controls. Thus, what is desirable to analyse and what is practical may be radically different. In this section, we discuss some of the problems that parasite genome analysis faces and how this influences the approaches being used.

Choice of organism to target for analysis
The diseases discussed in this article are, in fact, spectra of pathologies caused by different strains, subspecies, species or, in the case of filariasis, genera of parasite that may be separated by many millions of years of evolutionary history, resulting in dramatic differences in distribution, vector and definitive host specificity, behaviour and pathology. The choice of organism to study, thus, is critical. Selecting the single most pathogenic, human-infective form generates a unified data set of immediate clinical relevance but may reveal little about the causes of biomedically significant variation. Moreover, for parasites that require animal passage, one that can parasitize small laboratory mammals has a much wider potential for study and experimental manipulation than one that requires a primate host, even if it is not, in itself, a human pathogen. Such considerations are becoming increasingly important as the postgenomic era approaches because small mammal models are widely available and permit the development/analysis of transgenic parasite systems. On a purely practical level, adopting a single species may require that the life cycle (including that of the hosts) or culture system is transferred between laboratories, not always a simple task. There is also the question of strain selection; a single reference strain produces a unified data set and allows selection for useful biological characteristics but may require establishment of life cycles/culture systems, construction of new DNA and cDNA libraries, and extensive work to define its biological features and karyotype. The choice of experimen-
tal system is inevitably a compromise, and the different Parasite Genome Initiatives have reached different decisions based on the characteristics of the individual organisms (Table 3).

### Genetic analysis

For model organisms such as *S. cerevisiae* and *C. elegans*, genetic mapping (the identification and positioning on the chromosomes of duplications, deletions, and marker loci that control phenotypically identifiable genetic traits by analysis of recombination frequencies from genetic crosses) forms an integral component of genome analysis. With the notable exception of the protozoan *Toxoplasma* (not discussed in this review), complexities of the life cycle mean that parasites are rarely amenable to true genetic analysis. No sexual cycle has been observed in either *Leishmania* or *T. cruzi*, precluding any type of genetic study. In *Plasmodium*, experimental crosses are technically difficult and only two have been reported, but a crude genetic linkage map is available. Genetic analysis of *T. brucei* is possible but restricted by difficulties in obtaining large numbers of hybrid offspring from tsetse fly infections. However, isolation and cloning of large...
numbers of hybrids and selection of polymorphic markers (e.g., for drug resistance, human infectivity, transmissibility) is ongoing and genetic analysis is an integral part of the *T. brucei* genome project.\(^{(29)}\) Some experimental crosses have also been performed in schistosomes, but again technical difficulties preclude widespread application,\(^{(30)}\) none have yet been performed in *Brugia*.

**Influence of genome characteristics**

The characteristics of individual parasite genomes (Table 4) impose severe limitations on the types of analysis that are possible. An immediate distinction can be drawn between the protozoan parasites, which have (relatively) small genomes and chromosomes that do not condense during cell division, and the helminths, which possess (relatively) large genomes with correspondingly large, typically eukaryote chromosomes that condense during cell division. This distinction has profound effects on the strategies that can be adopted.

**Genome organization in the protozoan parasites**

The smaller overall size of the protozoan genomes (27–50 megabases [MB]) means that complete analysis, including full genomic sequencing, is potentially feasible. For prokaryotes with small genomes (<3 MB), the full genome sequence can be assembled by computer from random “shotgun” sequences. However, computational systems capable of larger shotgun assembly projects such as those of the parasitic protozoa are in their infancy and so a physical map (the defined minimal set of overlapping clones that spans the entire genome) is a prerequisite for genomic sequencing. The physical map, in turn, needs to be referenced to the karyotype (chromosome set) and so establishing the chromosome compliment, assigning reference marker loci to individual chromosomes and determining the extent and causes of genomic variation are vital first steps. Because the chromosomes of the protozoan parasites do not condense during cell division, they cannot be visualized by light microscopy and karyotype definition/analysis must be performed on chromosomes separated by pulse field gel electrophoresis (PFGE) techniques. This process has proved no easy task.

The simplest organization is found in *P. falciparum* for which the entire chromosome compliment can be resolved on a single PFGE run.\(^{(31)}\) Homologous chromosomes of different isolates can vary in size by up to 20%,\(^{(32)}\) with polymorphism resulting from various phenomena, including chromosome breakage in subtelomeric regions followed by addition of telomere repeats, crossover events, interchromosomal transpositions and amplifications of regions of the genome under selective pressure;\(^{(32,33)}\) specific amplifications, linked to drug resistance, have also been identified in central regions.\(^{(34)}\) Despite this variation, both linkage and synteny (gene order) appear to be well conserved between isolates and between species,\(^{(35,36)}\) suggesting that the physical map being developed for the reference strain will be broadly applicable. A further problem with genome analysis in *P. falciparum* is that, due to its extremely high A:T content, fragments of *P. falciparum* DNA > 5 KB are generally unstable in *Escherichia coli*;\(^{(37)}\) therefore, conventional bacterial mapping vectors such as cosmids and bacterial artificial chromosomes (BACs) cannot be used for library construction.

PFGE of *Leishmania* chromosomes reveals approximately 25 visible bands that vary in staining intensity (suggest-

<table>
<thead>
<tr>
<th>Parameter</th>
<th><em>T. brucei</em></th>
<th><em>T. cruzi</em></th>
<th><em>L. major</em></th>
<th><em>P. falciparum</em></th>
<th><em>Schistosoma</em></th>
<th><em>B. malayi</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ploidy</td>
<td>diploid</td>
<td>diploid</td>
<td>diploid</td>
<td>haploid in man</td>
<td>diploid</td>
<td>diploid</td>
</tr>
<tr>
<td>Haploid genome size (MB)</td>
<td>37–40</td>
<td>43–50</td>
<td>35–34</td>
<td>27–30</td>
<td>270</td>
<td>100</td>
</tr>
<tr>
<td>Chromosome number</td>
<td>At least 11 pairs of megabase house-keeping chromosomes</td>
<td>30–40, exact number still to be determined</td>
<td>n = 36</td>
<td>n = 14</td>
<td>n = 8 ZW, sex determination</td>
<td>n = 5 XY, sex determination</td>
</tr>
<tr>
<td>Chromosome size and type</td>
<td>1.15–5.2 MB, non-condensing</td>
<td>0.45–4.0 MB, non-condensing</td>
<td>0.3–3.0 MB, non-condensing</td>
<td>0.6–3.4 MB, non-condensing</td>
<td>condensing</td>
<td>condensing</td>
</tr>
<tr>
<td>A:T content</td>
<td>50% overall, 40–50% in coding regions, 50–60% in intergenic regions</td>
<td>49%</td>
<td>37%</td>
<td>82% overall, up to 90–95% in intergenic regions</td>
<td>66%</td>
<td>75%</td>
</tr>
<tr>
<td>Composition</td>
<td>12% highly repetitive, 20% middle repetitive</td>
<td>20–40% repetitive</td>
<td>30% repetitive</td>
<td>approx. 40% coding</td>
<td>60% repetitive, 30% single copy</td>
<td>17% repetitive, no methylated DNA</td>
</tr>
<tr>
<td>Expressed genes</td>
<td>10,000</td>
<td>10,000</td>
<td>8,000–10,000</td>
<td>5,000–7,500</td>
<td>15,000–20,000</td>
<td>16,000</td>
</tr>
</tbody>
</table>
ing that they contain varying numbers of unresolved chromosomes) and karyotypes of different strains and species show dramatic polymorphisms that have led to confusion about chromosome number. However, physical linkage patterns revealed by hybridising probe sequences to PFGE-separated chromosomes from a variety of species and strains have confirmed that 36 chromosome pairs exist. Variation in size between homologous chromosomes and extensive comigration of nonhomologous chromosomes results in shuffling of chromosome identification numbers with respect to chromosome size. This makes chromosome identification difficult and the side-by-side analysis of two strains of L. infantum and one of L. major under three sets of PFGE conditions is required to unambiguously resolve all 36 chromosome pairs and so confirm a map location. Despite this polymorphism, linkage groups appear well-conserved between species, suggesting that chromosome structure and gene order are maintained and the karyotype is now being transferred to the genome project reference strain. Variation between homologous chromosomes results from amplification/deletion in subtelomeric regions, indicating that the central core region of the chromosomes (which contains the coding and structural sequences) is more stable. The karyotype of cloned strains appears to be very stable, possibly as Leishmania lacks a sexual cycle and utilises polycistronic transcription, which may select against genome rearrangement.

T. brucei possesses three classes of chromosome, i.e., mini-, intermediate, and large, that serve different functions. The large (megabase) chromosomes, which contain the housekeeping genes, are the main focus for genomic study. Polymorphism amongst these chromosomes is more extreme than that in Leishmania with overall DNA content varying by up to 30% and chromosome homologues varying by up to 400%. For the megabase chromosomes, chromosome number has been determined by hybridising probes to PFGE-separated chromosomes and confirmed by comparative mapping of karyotypically polymorphic stocks under various PFGE conditions, allowing a chromosome numbering system to be proposed. Despite the huge size polymorphisms, no loss of synteny has been observed, indicating that variation is not due to large translocations between chromosomes. No gene probes hybridise exclusively to the mini- or intermediate chromosomes, suggesting that these may contain only variant surface glycoprotein genes (VSGs) involved in antigen switching for immune avoidance. VSG expression sites and related sequences. Huge arrays of transcriptionally silent VSGs are also found on several megabase chromosomes and extensive genome rearrangements occur to bring them into transcriptionally active telomeric sites. Despite these rearrangements, and the size polymorphisms, the overall structure of homologous chromosomes appears to be highly conserved. Unlike P. falciparum and Leishmania, size-polymorphic regions in T. brucei chromosomes can occur either subtelomERICally or within the central, gene-rich area. Retrotransposon-like elements have been described from both VSG-related and other regions, and there is speculation that they contribute to genome rearrangement.

The karyotype of T. cruzi is as polymorphic as that of T. brucei; 20 to 42 bands that vary in staining intensity are seen in different strains under a variety of PFGE conditions. Chromosome homologues vary by up to 50% within and between isolates but physical linkage and overall chromosomal organization appear to be conserved. Densitometric analysis that uses telomeric probes reveals approximately 64 chromosome equivalents per cell, but the definitive karyotype remains to be determined and probe hybridisation to PFGE-separated chromosomes is being used to identify further linkage groups. The genome of reference strain CL-Brener appears to be very stable, possibly because T. cruzi is an intracellular parasite and so does not possess the minichromosomes, VSGs, and associated gross genomic rearrangements that T. brucei requires for immune evasion. However, genomic rearrangements have been detected, variation in gene copy number/gene family composition can occur and breaking of physical linkage groups has been observed. Moreover, there are telomere-associated genes, including some for surface antigens, and exchange occurs between them. Retrotransposons have also been detected.

Despite such problems, the fact that the protozoan chromosomes can be (imperfectly) separated on, and isolated from, PFGE gels means that chromosome-by-chromosome approaches are possible, dividing the genome into more manageable portions and simplifying the task of physical map generation.

Genome organization in the helminth parasites

Because the chromosomes of Schistosoma and Brugia condense during cell division, they are visible by light microscopy. Moreover, differences in chromosome size, shape, centromere position and banding pattern allow the individual chromosomes to be identified, facilitating karyotype analysis and allowing probes to be mapped onto chromosome spreads by fluorescent in situ hybridisation (FISH) (although currently only for large fragment probes). To date, approximately 100 yeast artificial chromosome (YAC) clones have been mapped to the schistosome karyotype by FISH to anchor and orientate the future physical map and primed in situ hybridisation (PRINS) techniques are being optimised to increase the sensitivity of chromosome mapping. Such techniques are also applicable to Brugia but have not yet been extensively used.

Gross chromosomal morphology is conserved in schistosomes, although there is slight variation in centromere position and C-band pattern between strains, and greater variation between species. There are reports of major genome...
Brugia is known to carry a Wolbachia-like endosymbiont immune avoidance. Such findings are disputed by other workers, and the situation remains to be finally clarified although retrotransposons are known to be present. In addition, natural, hybridisation of schistosome species is a well-documented phenomenon, and there is preliminary evidence for gene transfer between species. What effect such plasticity (if it truly exists) will have on genome analysis remains to be determined. In contrast, the Brugia genome appears to be stable, chromatin diminution (the fragmentation and elimination of germ-line-specific chromatin), which is common in nematodes and can make the production of representative DNA libraries difficult, does not occur and transposable elements have not been reported. However, Brugia is known to carry a Wolbachia-like endosymbiont within its cells and the presence of this “foreign” genome may complicate analyses, necessitating a parallel, endosymbiont Genome Initiative. Furthermore, as with Plasmodium, the relatively high A:T content of Brugia DNA, may cause problems in library construction/analysis.

The downside of possessing large chromosomes is that even the smallest of the helminth chromosomes are too large to be separated by PFGE; therefore, individual chromosomes cannot be isolated. This means that chromosome-specific analyses cannot be undertaken and that physical mapping strategies are restricted to chromosome walking techniques (which are relatively inefficient) or to global shotgun approaches (a massive undertaking on genomes of this size). The resources required for full-scale analysis of the helminth genomes would be comparable to, or greater than, those required by the C. elegans project, and it is unlikely that full genomic sequencing will ever be initiated (the schistosome genome also contains so much repetitive DNA that this is unlikely to be a worthwhile endeavour). Therefore, under current circumstances, the helminth Initiatives must be selective, targeting specific regions of the genome for detailed analysis and their initial priority has been to target coding regions through gene discovery.

**Progress**

The Parasite Genome Initiatives are ongoing efforts and, ultimately will encompass as many different kinds of analysis as are possible to perform in the individual systems. However, in the initial phases, practical considerations have required the different Initiatives to focus on different aspects of genome analysis. Because full genomic sequencing is a stated goal of the protozoan Initiatives, priority has been given to karyotype definition (see above) and physical mapping as a prelude to genomic sequencing, whereas the helminth Initiatives have focused on gene discovery and development of resources for future mapping efforts.

**Physical mapping**

A physical map, that enables sequences to be ordered by reference to their parent clone, is a prerequisite for assembly of the protozoan genome sequencing projects and even for the parasitic helminths, where genome size is likely to preclude full-scale sequencing, will provide major benefits (1) by identifying the minimal clone set and regions of repetitive DNA, it minimises sequencing efforts; (2) it expedites the cloning of genes identified through EST sequencing and facilitates isolation of syntenous (similarly positioned) homologues from related species; (3) it permits analysis of gene structure/order, transcription units, regulatory elements, and polymorphic regions; and (4) it allows interspecific comparison of local gene order, physical linkage, and long range synteny. Many techniques are available for physical mapping and are appropriate for different circumstances.

Although P. falciparum DNA is generally unstable in traditional (bacterial) mapping vectors, it is stable in YACs and these have been used in a polymerase chain reaction (PCR)-based mapping strategy. Nested pools of YAC clones were prepared as PCR templates, and PCR primers specific for known P. falciparum sequences (from characterised genes, microsatellites, and genome sequence tags (GSTs, short regions of sequence generated from randomly selected DNA clones) used to amplify sequences from them. A positive PCR reaction tags any YAC pool that contains the specific sequence under test, and the nested pools positively identify the individual YAC(s) containing that sequence without needing to individually screen each clone for each sequence. These YACs became anchor points for map generation. End sequence was generated from the anchor YACs and used to design PCR primers for the next screening round. Consecutive cycles of sequencing and PCR screening assembled a series of contigs that eventually covered the whole chromosome, generating the map and, at the same time, assigning defined markers to it. Other chromosomes were mapped by the probe saturation approach by using GSTs from chromosome fragment libraries as markers. Comparative restriction mapping of YAC contigs and genomic DNA confirmed correspondence between map and genome. Complete maps are available for chromosomes 2, 3, 4, 8, and 12, but not all are from the reference strain and so work is ongoing to transfer them to it. Excepting chromosomes 10 and 11, all the other chromosome maps are effectively complete, with 85% of the genome assembled, and the density of marker sequences is now being increased by hybridisation of expressed sequence tags (ESTs, see Gene Discovery section below) and GSTs to YAC clones spotted onto high density filters. Low-resolution restriction maps are also available for most chromosomes, again not all derived from the reference
strain but comparison between strains suggests correspondence.\(^{(6)}\)

For *Leishmania*, a cosmid fingerprinting strategy has been used for physical mapping. Individual cosmid clones were cut with restriction enzymes (enzymes that cut DNA at particular sequences), and the fragments were end labelled, separated on polyacrylamide gels, and visualized by autoradiography. Fragment patterns were matched by computer to identify overlapping cosmids and assemble them into contigs. Probes generated from clones at the ends of contigs were then hybridised to cosmid filters to link contigs, and contigs were assigned to chromosomes by hybridisation to PFGE-separated chromosomes. From a library of 9,216 cosmid clones, 9,004 fingerprints have been assembled into 39 contigs, covering >91% of the genome.\(^{(77)}\) ESTs are currently being placed on the physical map to provide transcriptional loci to facilitate open reading frame (ORF) analysis of future genomic sequence data.

For *T. brucei*, the cloned reference stock TREU 927 has a smaller genome than most other stocks and its homologous chromosomes are of more even size.\(^{(43)}\) This is fortuitous for mapping because both homologues are represented in the genomic libraries which are constructed from diploid stages of the life cycle. P1 (bacteriophage cloning vectors capable of carrying inserts of 85–100 KB) and cosmid libraries are available and have been used to generate a complete contig/restriction map of chromosome 1\(^{(178)}\) and large contigs anchored to the other 10 megabase chromosomes by EST markers (Melville, unpublished). Comparison of the map with size-variable homologues in other cloned stocks indicates that polymorphisms affect the entire chromosome, although dispersed repeat sequences are confined to one specific region.\(^{(10)}\) Map generation is by hybridisation of ESTs, GSTs, and clone end probes onto gridded genomic libraries, and a BAC library is under construction to facilitate mapping. Concurrently, ESTs are being mapped onto the karyotype to anchor and orientate the map contigs.

For *T. cruzi*, cosmid, BAC, and YAC libraries are available\(^{(52,53,79)}\) and have been screened from chromosomes isolated by PFGE to produce chromosome-specific sublibraries.\(^{(80)}\) Mapping strategies have been adapted from those used in the yeast genome project, with predominately chromosome-specific cDNA clones (identified by hybridisation), rather than cosmids, being used as probes to screen the cosmid library gridded onto filters.\(^{(53,81)}\) This approach combines map production with gene mapping and minimises problems that would result from repetitive DNA sequences present in cosmid probes. Once contigs are assembled, restriction mapping identifies the overlapping regions to minimise future sequencing efforts and provide markers to confirm the accuracy of sequencing data. The vectors used for cosmid library construction have been specifically designed to facilitate both this restriction analysis and future transfection studies.\(^{(53,82)}\) Contig assembly is ongoing as a prelude to genomic sequencing and chromosomes 3 and 4 have already been mapped from cosmids.\(^{(83)}\) In addition, 850 KB of chromosome XVI has been assembled from YACs.\(^{(52)}\)

For the metazoan parasites *Brugia* and *Schistosoma*, genome size means that only low-resolution maps are possible with current resources and efforts have concentrated on generating the resources for future studies.

For *Brugia*, two complimentary mapping strategies will be used: (1) probing BAC library grids with 5,000 ESTs to provide sequenced marker loci; (2) sample-without-replacement screening of BAC libraries with BAC end probes to assemble contigs (in this approach, BAC clones are selected at random and probes generated from the ends of their inserts, these probes are then used to screen the BAC library gridded onto filters and identify overlapping clones, nonhybridising clones are used for each successive round of screening until all of the clones have been identified).\(^{(84)}\) The endosymbiont genome is also targeted for mapping.\(^{(85)}\)

For *Schistosoma*, a YAC library exists,\(^{(60)}\) a BAC library is currently under construction, and a pilot project is under way to produce a low-resolution map for chromosome 3 (which has the greatest density of YAC markers assigned to it). A chromosome walking strategy that uses the YACs as anchor points and assembles contigs outward from them until the whole chromosome is covered will be used.\(^{(86)}\) The majority of the schistosome genome is composed of repetitive DNA, and it is not yet known how this will affect mapping strategies, although it is likely that end probes, rather than whole inserts, will need to be used as probes to minimise false positive signals due to dispersed repetitive sequences.

Although the helminth chromosomes are too large to be isolated by PFGE, preventing production of chromosome-specific sublibraries, it has been possible to physically scrape individual schistosome chromosomes from chromosome spreads by using a micromanipulator.\(^{(59)}\) It is, theoretically, possible to generate libraries from such material\(^{(87)}\) and this approach may allow more focused mapping and sequencing studies to be performed in the future.

**Gene discovery**

With the identification of new drug and vaccine targets and the elucidation of mechanisms of drug resistance, antigenic variation, and genetic diversity being priorities for all of the Parasite Genome Initiatives, gene discovery programmes form an integral part of their activities. Much of the initial work has exploited the conventional EST approach,\(^{(88)}\) in which short (100–500 base) sequences are generated from the ends of randomly picked cDNA clones to provide “tags” for that clone. These are then compared with the public sequence databases to try to identify the gene they were derived from, allowing the rapid production of a gene catalogue for the organism. Examination of cDNA libraries gener-
ated from different stages of the life cycle detects developmentally regulated genes and generating/exploiting these resources has been an integral part of the Initiatives. (89–95)

Due to the limited availability of in vitro culture systems, parasites are usually recovered from laboratory hosts, with the inherent risk of contamination of both cDNA and genomic DNA libraries with host material. All trypanosome and Leishmania mRNAs, and the majority of nematode mRNAs, undergo post-translational modification through trans-splicing of a spliced leader (SL) sequence onto the 5’ end of transcripts. (96,97) As the SL sequence is parasite specific, using it for second-strand cDNA synthesis produces libraries that are free from contamination with host cDNA (they are also, by definition, intact at their 5’ end). Such libraries have been extensively used. (92–95) A schistosome SL has also been identified (98) but is restricted to a small proportion of cDNA transcripts (estimated at <10%, R. Davis, personal communication). Its significance remains to be determined, and schistosome SL libraries have not yet been examined.

In conventional cDNA libraries, differences in gene transcription levels and cloning bias mean that some transcripts are very abundant, others very rare, (99) and so redundancy in the data set, due to the same genes being “discovered” over and over again, can rapidly reduce the efficiency of EST projects. Several approaches are being explored to overcome such problems: (1) examining a large number of cDNA libraries, because the bias of each will differ; (2) library normalisation, to reduce the compositional bias (already in use in the T. cruzi initiative and shortly to be used in Schistosoma); (3) the use of arbitrary-primed reverse transcriptase (RT)-PCR to generate mini-libraries, which also allows library construction from the limiting amounts of cDNA available from many stages of parasite life cycles; (4) prescreening of libraries, to remove abundant transcripts; (9,85) and (5) random genomic DNA sequencing (see below).

Expressed sequence tagging was devised for efficient gene discovery in organisms with massive genomes in which the coding regions represent only a small fraction of the total and so random genomic sequencing would not be an efficient way to identify new genes. Such constraints are not universally applicable. The T. brucei genome is very gene dense; only one intron has ever been found, (102) and intergenic regions are very short. (103) Thus, for T. brucei, random sequencing of clones from genomic DNA libraries is as efficient a means of gene discovery as random sequencing of cDNA clones. (103) It has further advantages in that genes expressed at any stage of the life cycle can be identified (including those where cDNA library construction is impossible), the library is effectively “normalised,” so redundancy is minimised and, because intergenic regions are short, sequencing both ends of an insert can detect different ORFs. The P. falciparum genome is similarly gene dense (up to 40% coding), (91) and here another technique is being used. For as yet ill-defined reasons, mung bean nuclease specifically cuts P. falciparum DNA before and after structural genes and within some introns. The resulting noncoding regions do not clone efficiently; (104,105) therefore, genomic libraries constructed from mung bean nuclease-restricted P. falciparum DNA are efficient gene discovery resources (105) with similar benefits to those in T. brucei. Moreover, manipulation of digestion conditions allows preferential release of either whole genes or exons, facilitating isolation of sequences for further analysis. (106) Mung bean nuclease shows a similar “specificity” on trypanosome and Leishmania DNA (107,108) but has not yet been exploited for gene discovery in these systems.

Irrespective of strategy used, parasite gene discovery has been hugely successful and has resulted in a logarithmic increase in the catalogue of known parasite sequences (Fig. 2); parasites being amongst the highest placed organisms in the EST league table. (109) For example, if gene discovery in Schistosoma were to have continued at the pre-EST rate, it would have taken some 700 years to obtain the full gene catalogue, whereas 5 years of the EST initiative has tagged 15%–20% of the genes in its genome. All the EST sequences and their parent clones are in the public domain to serve as a resource for the research community. Hundreds of clones

Figure 2. The impact of parasite gene discovery programmes (data to mid-September of 1998). Current totals are available on the WWW. (109)
have been distributed for ongoing detailed characterisation studies, and many have been used as probes to identify homologous genes in other species. They are also being used as defined probes in the physical and chromosome mapping projects.

**Full genome sequencing**

As discussed above, features of the *P. falciparum* and *T. brucei* genomes allow efficient gene discovery from genomic DNA clones. However, the entire genome sequence of a parasite represents the complete inventory of virulence factors, immunogens, drug targets, etc., expressed by that organism and, thus, is the ultimate resource for the research community. Determining full genomic sequence for even a small genome requires very significant investment in labour, reagents, and facilities. Funding has been secured to sequence the entire *P. falciparum* genome, with the workload shared by major sequencing facilities in the United States (TIGR and Stanford) and United Kingdom (Sanger Centre). This work is proceeding, chromosome-by-chromosome, with shotgun sequencing of individual chromosomes purified by PFGE. The Sanger and Stanford groups also use shotgun libraries of YACs previously localised to chromosomes by the physical mapping initiative to identify contigs from the same region and aid gap closure. To date, over 176,000 sequencing runs have been performed; pilot projects on chromosomes 2 and 3 are nearly complete with chromosomes 1, 4, and 14 well under way, and chromosomes 9, 10, 11, 12, and 13 started. With large-scale genome sequencing projects costing some US$ 500,000 per MB, complete sequencing of the genomes of all the other parasitic protozoa seems unlikely in the short term. However, medium-scale funding has been secured for several projects. For *Leishmania*, where the minimal tiled cosmid set from the physical map allows a chromosome-by-chromosome approach, funding is available to sequence 45% of the genome (15 MB) by the year 2000, with the goal of obtaining the complete genome sequence by 2002. Chromosome 1 is complete, with chromosomes 2, 4, 5, and 6 in progress. Genomic sequencing means that EST initiatives for *Leishmania* have now been halted. For *T. brucei*, two projects are under way, one to completely sequence chromosome 1, the other a large-scale, random genome sequence survey contributing to both gene discovery and physical mapping as a prelude to sequencing whole chromosomes. These endeavours will produce dividends in cataloguing new genes, especially those expressed at very low levels and so unlikely to be identified through ESTs (and will be vital when redundancy makes EST generation inefficient), in identifying regulatory elements and in elucidating genome structure and organization. The *T. cruzi* initiative is still seeking major funding, but a pilot project has virtually completed chromosome 3 and chromosomes 2 and 4 are under way. Despite a larger genome, genomic sequencing may allow efficient gene discovery in *T. cruzi* as gene-rich areas, with ORFs every 3 KB, have been identified. In addition, experiments indicate some synteny between *Leishmania* and *T. cruzi*; therefore, map and genomic sequence data from the former may allow targeted analysis in the latter.

For the metazoan parasites *Brugia* and *Schistosoma*, which have massive genomes (and in *Schistosoma*, very high levels of repetitive DNA), full genome sequencing is probably neither practical, nor necessary, even if the resources were available. Instead, targeted areas of the genome will be selected for sequencing. For *Brugia*, the imminent completion of the *C. elegans* genomic sequence will provide an invaluable "roadmap" to aid such studies.

**Informatics**

The Parasite Genome Projects are generating large amounts of data of diverse forms; sequences, homology search results, sequence clusters, library details, bibliographic, laboratory and personnel information, clone hybridisation results, network policy documents, etc., and much of these data are of wide interest as it pertains to groups of organisms that are traditionally under-researched. The Internet provides the ideal mechanism for rapid dissemination of these data; all of the parasite genome networks maintain WWW servers providing up to date information for their projects and an email discussion list facilitates communication. In addition, a dedicated site at the European Bioinformatics Institute provides general information and analytical services, including a parasite-specific BLAST server (Table 2).

For all of the Parasite Genome Initiatives, cluster analysis of the sequence data sets is an important ongoing activity. By grouping sequences that are derived from the same gene, these analyses reveal redundancy in the datasets and identify the actual number of different genes tagged to date (as opposed to sequences generated), thus monitoring the efficiency of the ongoing work, for example, in *Brugia*, with some 16,000 sequences available, approximately one in three new sequences tags a previously unknown gene, whereas in *Schistosoma* (8,000 sequences available) it is two in three. Clustering also creates consensus sequences that may facilitate gene identification and identifies highly expressed genes of unknown function, that may be important candidates for early study. To facilitate data analysis, standardised clone nomenclature systems have been devised for individual Initiatives (details on the Network WWW sites, see Table 2) and, for *Brugia*, a standardised gene nomenclature has been proposed.

Data generation is one thing but these data need to be made available to the research community in a form that is easy both to access and to navigate. To this end, as well as their WWW sites, all the Parasite Genome Initiatives curate integrated, hypertext databases of genomic and related data. These are all based on ACeDB, the database engine devised...
for the C. elegans Genome Initiative,(119) which is specifically designed for the storage, analysis, and retrieval of genome data, with graphical displays to facilitate interpretation of specific data types such as sequences and maps (Fig. 3).

All the parasite genome databases are available for local installation(120,121) and the WHO parasite databases are also mounted at the Human Genome Mapping Project Resource Centre (access requires registration).(122) WWW interfaces to the databases are under development to facilitate Internet access and a CD-ROM version is being produced for distribution to laboratories in developing countries where Internet access is insufficient to permit downloading.

Prospects
Despite the limitations of the biological systems involved, the Parasite Genome Initiatives have already proved to be extremely effective mechanisms for generating a wealth of fundamental information about parasites. The Initiatives are ongoing and future work will both enhance existing datasets and generate new ones as mapping and genome sequencing efforts develop. In addition, the coming years are likely to see huge improvements in technology that will transform molecular biology from an analytical science into a discipline of bioinformatics, structure/function analysis and in vivo manipulation. In this final section, we examine how the Parasite Genome Initiatives might develop and where their data may lead us.

Parasite Genome Initiatives exist for a purpose, i.e., to identify new drug, vaccine, and diagnostics targets; to determine the molecular basis of parasite function; and to provide resources to facilitate analysis. These resources are in the public domain and “post genomic” investigations that exploit them for functional analysis are already being initiated. Database homology searches have identified hundreds of ESTs/GSTs as being of potential interest, including potential immuno-modulatory proteins, signal transduction molecules, cell cycle regulators, key metabolic enzymes, etc., and these have been distributed for detailed characterisation studies.

For vaccine development, as well as work to identify and characterise new candidate antigens, attention is increasing being focused on the use of novel delivery systems such as multiantigen cocktails, DNA vaccines, or live viral vectors, or on the modulation of parasite-encoded immunomodulatory proteins.(123–130)

The majority of newly tagged genes from the gene discovery programmes (55%–85%) and of the ORFs identified by genomic sequencing have no recognisable homology with characterised genes on the public databases(9,89–95,103,105) and a major challenge for the future will be to develop effective, user-friendly bioinformatics systems capable of assigning function to them or of identifying potential functional motifs within them. Furthermore, with the advent of major genome sequencing efforts comes the need to identify the location of genes within the sequence. Gene prediction
algorithms already exist for Plasmodium and C. elegans but will need to be trained to work for other species.

Early identification of gene function is useful but by no means essential, and the mass screening of pooled, uncharacterised ESTs for their protein and/or DNA vaccine potential is also being considered. Similarly, the development of standardised phenotypic assays, suitable for mass screening of cultured protozoan parasites with novel drug compounds is being addressed.

ESTs derived from different life cycle stages already provide basic information on the ontogeny of gene expression (with the caveat of post-transcriptional control), but the absence of a gene from a particular EST dataset does not necessarily mean that the gene is not expressed at that stage, it may just not have been picked from the relevant library. As the datasets grow, analysis of gene expression patterns will be achieved far more effectively through the use of DNA chip technology, where the known gene set is gridded and screened with cDNA from different developmental stages, from parasites with different biological properties, and from parasites subjected to different environmental stresses (drug challenge, cytokine stimulation, temperature shock, etc.) to identify genes that are induced, repressed, or developmentally regulated. Production of gene arrays is high on the agenda of all of the Parasite Genome Initiatives, although the expressed gene content of these organisms means that, at least in the short term, such arrays are likely to be generated “in-house,” by gridding the known gene set onto filters, rather than by using state of the art, but costly, commercial technologies that synthesise gene sequences directly on glass slides through a combination of oligonucleotide synthesis and photolithography techniques.

Comparative genomics has enormous potential to identify gene homologues in related organisms, to provide information about functional diversity (e.g., across the filarial nematodes) or to assist in the development of experimental models (e.g., by using data from P. falciparum as a template for rapid analysis of rodent malarias in which parasite gene knockout/transfection can more readily be studied). Sequences that match only sequences from related organisms or have no database match may also represent specific phylogenetic or pathologic adaptations and, thus, be possible targets for intervention. Some of the most exciting comparative analysis are between Brugia and C. elegans. The latter, a free-living nematode, has been studied in exquisite detail; a near complete physical map and full genome sequence will be available by 1999. the full cell lineage is known, anatomy and neural connections have been reconstructed from serial sections, and thousands of morphologic, developmental, and behavioural mutants have been genetically mapped. Techniques are available to analyse gene expression and function in C. elegans and many may be transferable to Brugia.

Not surprisingly, a large proportion of Brugia ESTs have a C. elegans gene as their closest homologue. Frequently, the latter has been fully sequenced, facilitating functional identification of the Brugia version. The evolutionary distance between the two nematodes is such that shared sequence motifs are likely to have functional significance and could represent targets for broad spectrum nematocides. Comparison in the opposite direction is also of value with Brugia ESTs, confirming the existence of hypothetical C. elegans genes that have been predicted from genomic sequence. Synteny between the two species is also being investigated as a possible route to targeted gene discovery; if a potentially interesting gene in C. elegans is flanked by two housekeeping genes, and gene order is conserved between the two species, the housekeeping genes can be used as markers for rapid isolation of the Brugia homologue. Similarly, for the trypanosomases and Leishmania, it may be possible to map genes whose location is known in one species onto the physical map of the other species.

Perhaps the most exciting prospects are for gene analysis by in vivo genetic manipulation (knockouts, transgenics, etc.), both for mass screening assays to identify functionally important genes and for analysis of individual genes. Transfection systems already exist for a number of parasitic protozoa and recent improvements in reporter gene technology will facilitate gene expression studies on these organisms. Schistosome transfection systems are under development, and for Brugia it may even be possible to use its endosymbiont as a vehicle for gene delivery. With the exception of Plasmodium, in which the human infective stage is haploid, all the parasites are diploid and this complicates the production of “knockout” lines as both copies of the targeted gene must be inactivated. For Leishmania, chromosome fragmentation techniques aimed at generating a comprehensive panel of strains in which known regions of the genome are rendered haploid, and other knockout strategies, are scheduled for investigation. Functional analysis in well-defined heterologous systems will also prove vital. It is already possible to express genes from parasitic nematodes in C. elegans and to rescue S. cerevisiae mutants with malaria genes. An international collaboration is currently aiming to produce a panel of knockouts for every single S. cerevisiae gene (some 6,000) in each of four genetic backgrounds which would provide a major resource for functional analysis. Adaptation of such technologies to those parasitic protozoa that can be cultured in vitro may ultimately be possible (T. brucei, T. cruzi, and some Leishmania species (but not L. major) can be grown in axenic culture, allowing the production of large numbers of parasites for screening assays, culture of Plasmodium requires addition of mammalian erythrocytes). Expression of malaria genes in Toxoplasma may also be possible.
Proteomics, the study of gene expression through the databasing and preliminary characterisation of its protein products, is a major new research field that complements more traditional aspects of genomics.\(^{(136)}\) Two-dimensional gel electrophoresis is used to separate proteins from extracts of cells, tissues, organs, or whole organisms and to produce a database of their isoelectric point, molecular weight, and relative abundance. Individual protein spots (as little as 1 picomole) are then excised from the gel and analysed by N-terminal peptide sequencing or automated “time of flight” mass spectroscopy. Comparative analysis of protein profiles obtained from different sources (life cycle stages, experimental conditions, etc.) directly reveals changes in protein expression levels. The amino acid sequence tag data from these analyses is then compared with the DNA or protein sequence databases to identify known proteins. For unknown proteins, amino acid sequence is used to design oligonucleotide probes to isolate the genomic or cDNA clones corresponding to that protein. Without doubt, the Parasite Genome Initiatives will shortly commence proteomic studies, and the integration of proteomic data with existing genomic databases will be an important task. The potential is already clear; comparison of a peptide sequence from the filarial nematode \textit{Dirofilaria immitis} with the \textit{Brugia} EST database has been used to identify \textit{Brugia} cDNA clones for the homologous gene, and these then were used to isolate the \textit{Dirofilaria} equivalent (Blaxter, unpublished).

On the purely technical side, optical restriction mapping techniques that use light microscopy to size and position chromosome fragments may prove of benefit in the parasitic protozoa, in analysing size polymorphism in homologous chromosomes and identifying markers for distinguishing con- somes and identifying markers for distinguishing con-migrating, nonhomologous chromosomes.\(^{(137)}\)

In the excitement of product-oriented analysis, we should not forget that parasites are fascinating organisms in their own right, providing systems for the study of differentiation, development, gene expression, and evolution. For example, platyhelminthes are traditionally regarded as the earliest diverging bilateral metazoans, and schistosomes are unique amongst them in possessing separate sexes, this being determined through heteromorphic sex chromosomes. Were resources available to map and sequence these chromosomes and compare them with the chromosomes of hermaphroditic relatives and to the sex chromosomes of other organisms, fascinating and fundamental biological questions could be addressed.

To date, much of the output of the Parasite Genome Initiatives has been generated by individual researchers working together in international collaborations. This approach has done much to stimulate communication within the research community and to provide opportunities for the training of scientists from disease endemic countries. As the Initiatives develop, there is likely to be a shift toward the major sequencing centres where the huge economies of scale necessary for cost-effective genome sequencing are possible, and toward more insular research where postgenomic studies with commercial potential are concerned. The challenge will be to keep the dialogue going and to ensure that the need to fund functional genomics does not jeopardise progress in the genomic studies from which they derive.

Concluding remarks

Enshrined in the mandate of the WHO is the right of each and every person, including the world’s underprivileged, to good health and the successful control of tropical diseases is essential to achieve this goal. In the past few years, collaborations between institutes in the Developed and Developing Worlds have clearly demonstrated the potential contribution that Parasite Genome Initiatives can make to this goal by identifying new targets for the development of diagnostic tests, drugs, and vaccines, and by providing the biological and informatics resources for their rapid characterisation and analysis. Moreover, progress has been achieved on a remark-ably small budget. Parasitology is a historically underserved research discipline (per fatality, malaria gets 2% of the research funding that HIV/Aids receives), significant analysis of any genome is a massive, labour intensive and costly exercise, and there is much competition for limited funds. The Parasite Genome community has successfully established collaborative links and proved the feasibility of its approaches. What is needed now is the will, on the part of funding agencies, to dedicate the funds to allow full-scale genomic and postgenomic analyses to occur.\(^{(138)}\)

Acknowledgments:

We thank colleagues in the Parasite Genome Networks for sharing data. This paper is dedicated to the memory of Prof. Mette Strand, coordinator of the WHO \textit{Schistosoma} Genome Network from 1994 to 1997.

References

2. \url{http://www.malaria.org/MAC.HTM}
Review articles

29. http://www.gla.ac.uk/Acad/WUMPwumptai.htm
52. http://www.dbtbr.fiocruz.br/genome/cruzir/meet97.html
Review articles


61. http://www.nhm.ac.uk/hosted_sites/schisto/mapping/FISHmaps.html


82. el-Sayed NM, Alarcon CM, Beck JC, Sheffield VC, Dowson JE. cDNA expressed sequence tags of Trypanosoma brucei rhodesiensis provide new insights into the biology of the parasite. Mol Biochem Parasitol 1995;73:75–90.


110. http://www.sanger.ac.uk/Projects/P_falciparum/


113. http://www.ebi.ac.uk/parasites/ftp.ebi.ac.uk/pub/databases/parasites/


