Accession number: txt001dcn; **Original accession date:** 5 May 1998;
Revision version accession date: 10 June 1998, **Archived previous version**: yes (txt001dcn5May98)

er molecular medicine

The Malaria Genome Sequencing Project

Daniel J. Carucci, Malcolm J. Gardner, Herve Tettelin, Leda M. Cummings, Hamilton O. Smith, Mark D. Adams, Stephen L. Hoffman and J. Craig Venter

An international consortium of genome centres, advanced development teams and funding agencies has begun the task of sequencing the genome of the parasite *Plasmodium falciparum*, the most important cause of human malaria. Sequencing is proceeding chromosome by chromosome, and the annotated sequence of chromosome 2 is nearly finished. With the continual release of sequence data as they are generated, malaria researchers have access to a steady stream of genomic sequences and will soon have the complete annotation of all of the estimated 5000–7000 *P. falciparum* genes. The task will then be how to best apply these data to the development of new anti-malarial drugs, vaccines and diagnostic tests. This review provides a brief overview of the Malaria Genome Sequencing Project and suggests potential directions for future malaria research.

In 1977, Fred Sanger and his colleagues heralded in the field known as genomics, having shown that it was possible to determine the entire genetic sequence of the virus phi-X (ϕX174) (Ref. 1). In that same year the completed genome of another virus, simian virus 40 (SV40), was also reported (Refs 2, 3). Sequencing progressed rapidly as genomes an order of magnitude larger, such as the bacteriophages T7 and lambda, were completed (Refs 4, 5). Within 15 years of the first viral genome being completed, the genomes of plant chloroplasts and the Epstein–Barr virus (EBV) were finished (Refs 6, 7), and many individual gene sequences were deposited into the public domain (Ref. 8). These first projects, though tedious to complete and requiring a great deal of manual effort, first suggested that 'unlocking the key to the genome' was possible; they also showed that it was technically feasible to determine the entire genome sequence of an organism and thus gain access to the description of its fundamental biology. Automated sequencing apparatus and

Daniel J. Carucci[1] (Corresponding author)
[1]Malaria Program, Naval Medical Research Institute, 12300 Washington Avenue, Rockville, MD 20852, USA, Tel: +1 301 295 6989; Fax: +1 301 295 6171; e-mail: caruccid@nmripo.nmri.nnmc.navy.mil

Malcolm J. Gardner[2], Herve Tettelin[2], Leda M. Cummings[2], Hamilton O. Smith[2], Mark D. Adams[2], Stephen L. Hoffman[1] and J. Craig Venter[2]
[2]The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA

https://doi.org/10.1017/S146239949800012X Published online by Cambridge University Press

Accession number: txt001dcn; **Original accession date:** 5 May 1998;
Revision version accession date: 10 June 1998, **Archived previous version**: yes (txt001dcn5May98)

improvements in sequencing chemistries progressed through the late 1980s (Ref. 9), and soon sequencing laboratories scaled up production, refined computational hardware and software, and developed coordinated methods to produce and analyse large quantities of DNA rapidly and efficiently (Refs 10, 11, 12, 13). Stimulated by the success of previous sequencing projects and the tantalising potential benefits of large-scale sequencing, scientists in the mid-1980s considered the enormous task of determining the sequence of the entire human genome. At 3 billion (3×10^9) base pairs (bp), the Human Genome Project represented the largest genome project ever undertaken. Although most of the initial focus of the Human Genome Project has been the production of genome 'maps', attention is now turning towards sequencing (Ref. 14). Indeed, even before human genome maps were completed, large-scale efforts were directed towards sequencing individual human genes (Refs 10, 11, 12).

Efforts have not been solely centred on the Human Genome Project; the determination of the genetic sequences of microbial organisms, especially those of human pathogens, has the potential to revolutionise the development of new drugs and vaccines. A milestone in genomics was reached in 1995 when the first complete bacterial genome was reported (Ref. 15). The genome of the free-living bacteria *Haemophilus influenzae*, at 1.8 million bp, was the largest completed genome, and the first example of a 'whole-genome sequencing strategy' being applied to a microbial organism. On the heels of this news, Barry Bloom in a leader article in *Nature* considered that 'the power and cost-effectiveness of modern genome sequencing technology mean that the complete genome sequences of 25 of the major bacterial and parasitic pathogens could be available within five years' (Ref. 16). At that time, he thought that 'for about 100 million dollars, we could buy the sequence of every virulence determinant, every protein antigen and every drug target' (Ref. 16). He predicted that this up-front investment in genome sequencing would produce information that would be available to scientists forever, and that 'we could then think about a new, post-genomic era of microbe biology' (Ref. 16).

Genome sequencing is progressing at an extraordinary rate; there are now 13 published microbial genomes and over 60 additional microbial genomes being sequenced (The current status of microbial sequencing can be found at http://www.tigr.org/tdb/mdb/mdb.html). Few people in 1977 could have imagined the advances in genome sequencing that have occurred since the 5386 base pairs of phi-X (ΦX174) were first published. Today, the entire genome of that first sequenced virus could be completed by a handful of people in a genome centre in an afternoon.

Sequencing strategies

The approach taken towards the sequencing of the genome of an organism depends on a variety of factors, including the availability of sequencing reagents, the size of the genome of the organism and other characteristics of the genome. As the size of a given genome project increases, methods are employed to partition the entire genome into smaller subunits. This usually means constructing a DNA library, by dividing the genomic DNA into smaller fragments and cloning them into a vector, such as a cosmid. Cosmids can accept inserts as large as 40 kilobase pairs (kb) and are arranged in order along the genome, creating a physical map of the genome. Once the minimum number of overlapping cosmids has been determined, sequencing begins on each identified cosmid. By sequencing cosmid-sized fragments of the genome, data handling becomes more manageable and assembly of the final sequence is facilitated. However, as the genome size increases beyond several megabase pairs (Mb) (consider, for example, the human genome, which comprises 3×10^9 bp), an additional, upper layer of organisation is needed. The reason for this becomes clear when one considers subcloning the entire human genome using cosmids; a ten-fold representation of the genome (every region in the DNA library should be present at least ten times) would require 750,000 cosmid clones. By today's standards, this would be an unmanageable number of clones to characterise and arrange in order along the genome. For such large organisms, a large-insert library must be created, often using bacterial artificial chromosomes (BACs) (Ref. 17) or yeast artificial chromosomes (YACs; Ref. 18 and reviewed in Ref. 19), which can accept DNA fragments that were several hundred kb in size. A low-resolution physical map (with relatively few clones widely separated) is created using BACs or YACs: these are then subcloned into cosmids to produce a high-resolution map. The

https://doi.org/10.1017/S146239949800012X Published online by Cambridge University Press

Accession number: txt001dcn; **Original accession date:** 5 May 1998;
Revision version accession date: 10 June 1998, **Archived previous version**: yes (txt001dcn5May98)

production of physical maps requires a tremendous amount of up-front effort in development, characterisation and construction; for example, the generation of a physical map for the Human Genome Project has required nearly 10 years of effort and millions of dollars. In addition, large-insert clones often contain insert DNA that is rearranged (brought together in a different position) or chimeric (two previously separate sections are brought together) and, thus, are of no use for sequencing. Furthermore, DNA in YACs is often contaminated with yeast chromosomal DNA during the purification process. Therefore, new strategies to approach large-genome sequencing are being considered that do not rely on previous mapping data. These include 'random-shotgun sequencing', using small (1–2kb) fragments of sheared DNA, of complete genomes and, for larger genomes, a BAC-end sequencing strategy (Ref. 20). In BAC-end sequencing the minimum number of overlapping clones necessary to cover a region of the genome is identified and sequenced. Only the end of each clone is sequenced; the information is collected and the gaps are then filled by further full-length sequencing. As sequencing technology and computer assembly algorithms improve, it will become possible to sequence larger genomes by the 'shotgun method'. These sequenced-based methods have the potential to expedite genome sequencing and further reduce the overall costs involved.

Sequencing microbial genomes

For smaller genomes, such as those of bacteria that range in size from 600 kb to 5 Mb, shotgun sequencing of the whole genome is becoming routine. The first bacterial genome, *H. influenzae,* was completed almost entirely by 'shotgun sequencing' (Ref. 15); that is, the entire genome (1.8 Mb) was sheared into small (1–3 kb) fragments and randomly cloned into the sequencing plasmid. The development of a coordinated sequencing effort, an integrated database management system, and improved sequence-assembly software meant that the entire genome of *H. influenzae* could be completed with ~ 24,000 successful sequencing reactions, all within approximately one year. Even as the *H. influenzae* sequence publication 'went to press', sequencing of two additional microbial genomes was already nearing completion (Refs 21, 22). Of the 13 microbial genomes that have been sequenced to date, seven were completed using whole-genome shotgun sequencing.

The Malaria Genome Sequencing Project

In May 1996, at a meeting sponsored by the US National Institutes of Health (NIH) and the Burroughs Wellcome Fund, scientists and funding agencies met to discuss the possibility of sequencing one or more plasmodium genomes, the parasites responsible for human and animal malaria. The result of the meeting was the establishment of an international consortium, comprising genome centres, advanced development teams and funding agencies, whose goal was to sequence and annotate the entire genome of *P. falciparum*, the parasite responsible for nearly all of the deaths due to malaria in humans (Ref. 23). A pilot project at The Institute for Genomic Research (TIGR, Rockville, MD, USA) and the Naval Medical Research Institute (NMRI, Rockville, MD, USA) was funded by the NIH and the US Department of Defense (DoD) to develop sequencing strategies for the Malaria Genome Sequencing Project. This has resulted in the complete 1-Mb sequence of *P. falciparum* chromosome 2 (manuscript, in preparation). Sequencing of the other chromosomes is proceeding at three genome centres: TIGR/NMRI, the Sanger Centre (Hinxton, UK) and Stanford University (Stanford, CA, USA), with funding from the Burroughs Wellcome Fund, the NIH, the Wellcome Trust, and the DoD. Early efforts have met with success and thus the consortium is pushing forward with the intent of completing the entire genome of *P. falciparum* by 2002–2003.

Clinical implications/applications

Why sequence the malaria genome?

The world malaria situation is worsening. The World Health Organization (WHO) estimates that one quarter of the population of the world lives in malarious areas and that 300–500 million cases of malaria occur annually (Ref. 24). Although more than 2.6 million people die every year from this disease, few people in the developed world realise the enormous economic, political and social burden that malaria places on those living with this disease (Ref. 24). In addition, increasing numbers of people will be exposed to malaria as the effects of global warming are manifest and as the mosquito vector of malaria encroaches into the non-malarious world (Refs 25, 26). With increasing air travel in a 'shrinking world', in the

https://doi.org/10.1017/S146239949800012X Published online by Cambridge University Press

future, more people that previously were not generally exposed to malaria will be placed at risk from this disease. Unfortunately, drug resistance in *P. falciparum* to chloroquine, one of the best anti-malarial drugs ever developed, is widespread and is found in most of the malarious world. Other species of *Plasmodium*, particularly *P. vivax*, are also beginning to develop patterns of chloroquine resistance (Ref. 27). Success in developing new anti-malarial drugs has been short-lived because plasmodium parasites continue to develop resistance to broad classes of anti-malarial drugs; in fact, most of the drugs used for anti-malarial prevention and therapy such as mefloquine are no longer effective in parts of the malarious world. Moreover, despite numerous clinical trials of malaria vaccines, there is, as yet, no licenced malaria vaccine (Ref. 28). It is clear that novel strategies are urgently needed to combat the menacing problem of malaria.

The difficulty of the situation that faces malaria researchers can be best appreciated when one examines the complexities of the parasite. The malaria parasite is an extraordinarily complex microorganism, which has evolved over the past millennia in a hostile immune environment; there is no apparent symbiosis between the malaria parasite and its human host. The plasmodium parasite possesses a complex multistage life cycle (Fig. 1, fig001dcn) in both a vertebrate (such as human) and an invertebrate (such as an *Anopheles* sp. mosquito) host. It exists (1) free in the circulatory system; (2) inside liver cells (hepatocytes), which are capable of presenting parasite antigens in association with major histocompatibility complex (MHC) molecules; and (3) inside red blood cells, which in humans do not have an MHC-restricted antigen presentation pathway. The parasite is exposed to both humoral (soluble) and cellular immune mechanisms. It has also developed complex drug-resistance mechanisms, which span a broad range of compounds. The design of new anti-malarial drugs and vaccines must, therefore, consider both immune evasion (avoidance of the immune system) and drug-resistance mechanisms. Any new approach must also be directed against multiple stages of the parasite, altogether a momentous undertaking. In many ways, malaria researchers are woefully under-equipped to deal with this complex parasite. Because *in vitro* cultivation of most malaria parasites is routinely possible only for the blood stages, experimental access to the other stages of the parasite life cycle and their respective antigens is limited. Animal models do exist for malaria; however, none reproduces accurately the pathology that is seen in humans. Although the transfection of genes into malaria parasites has been developed recently, it is being used in only a few laboratories and is not yet routine. Finally, of the estimated 5000–7000 genes in *Plasmodium* spp., only a few hundred are known; these represent little more than a brief 'snapshot' of all of the genes used by the parasite. Clearly, more information is needed to develop novel anti-malarial strategies. The advances in genome sequencing over the past two decades now make it possible to consider 'unlocking the malaria genome'. For malaria researchers, access to the malaria genome will undoubtedly provide tools to assist in the discovery of novel targets for the development of malaria vaccines and anti-malaria drugs. It should yield targets for improved diagnostic tests and provide a better understanding of the development of both drug resistance and immune evasion. These should, almost certainly, result in better control of this parasite, and potentially the eradication of malaria in humans.

Genetics and molecular biology

The plasmodium genome

The genome of *Plasmodium* spp. is ~30 Mb and is distributed among 14 chromosomes, which range in size from 650 kb to 3.5 Mb each (Table 1, tab001dcn). Figure 2 (fig002dcn) shows a comparison of the sizes of some other genomes. *Plasmodium falciparum* and several other *Plasmodium* spp. are unusual in that their genomes have an extraordinary bias towards two nucleotides: adenine (A) and thymine (T). In regions that code for proteins, the A-T bias is greater than 76%, whereas in intergenic regions (regions between genes) and in introns (regions within genes that are removed before final transcription), the A-T content can approach 100% (Refs 29, 30). This extreme A-T bias is thought to be responsible for the observed difficulty in cloning and maintaining large segments (greater than several kb) of *P. falciparum* DNA in *Escherichia coli* (Ref. 31). This instability has been problematic because there are, as yet, no bacterial libraries available that can accept large inserts of *P. falciparum* DNA. The development of YACs (Ref. 18) has been applied with success to

https://doi.org/10.1017/S146239949800012X Published online by Cambridge University Press

Accession number: txt001dcn; **Original accession date:** 5 May 1998;
Revision version accession date: 10 June 1998, **Archived previous version**: yes (txt001dcn5May98)

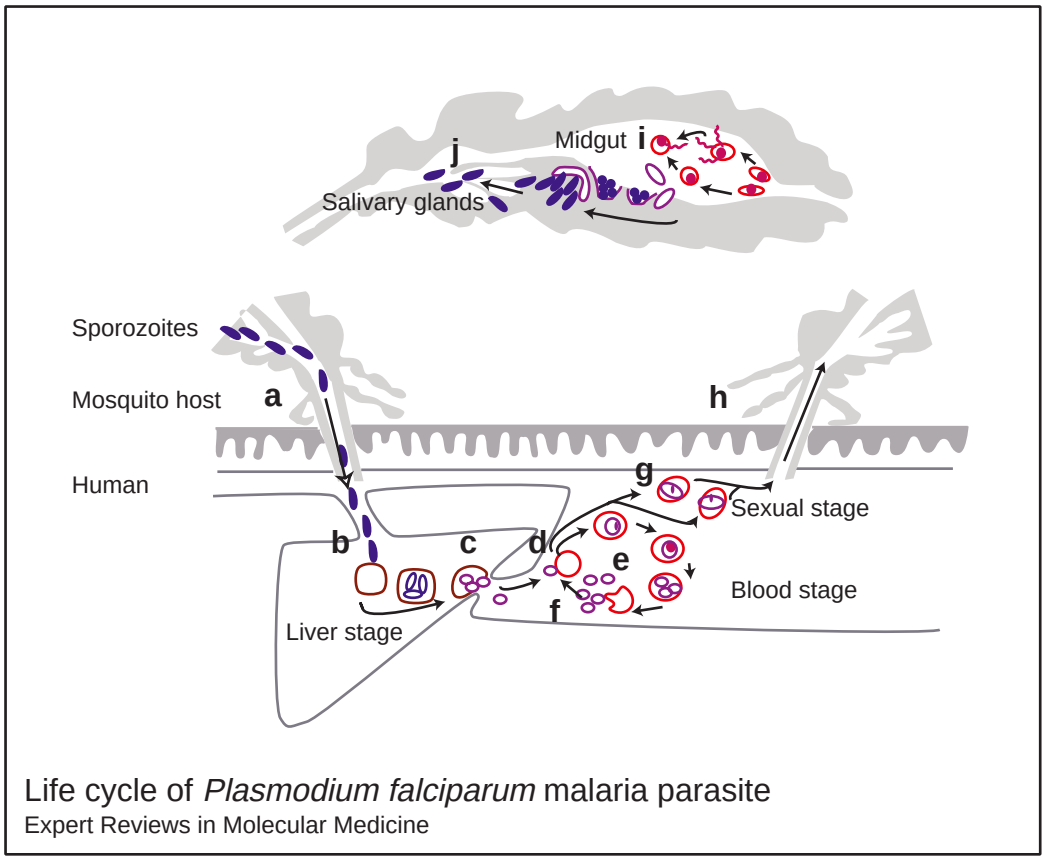


Figure 1. The life cycle of *Plasmodium falciparum* malaria parasite. Malaria is caused by infection with an obligate, intracellular protozoan parasite of the genus *Plasmodium*. Of the four species that infect humans (*Plasmodium falciparum*, *Plasmodium vivax, Plasmodium ovale* and *Plasmodium malariae*), it is *P. falciparum* that is responsible for virtually all deaths. The life cycle of *Plasmodium* spp.is complex and somewhat specific to the parasite species. (a) *P. falciparum* infection in humans begins when an infected *Anopheles* sp. mosquito takes a blood meal and injects infective sporozoites into the peripheral circulation. (b) Within minutes, these sporozoites invade hepatocytes in the liver and, over approximately one week, undergo asexual multiplication, producing tens of thousands of merozoite forms of the parasite. (c) When the infected hepatocyte ruptures, merozoites are released into the peripheral circulation. (d) The merozoites invade red blood cells (rbcs) and (e) complete another round of multiplication within 48–72 h, with the production of 16–20 additional merozoites per rbc, which devour the rbc haemoglobin in the process. (f) The released merozoites invade additional rbcs and carry on the cycle. It is the synchronous release of merozoites that is thought to be responsible for the periodic fevers associated with malaria. (g) Some invading merozoites do not divide, but differentiate into male (microgametocyte) and female (macrogametocyte) sexual forms. (h) These sexual forms are taken from the bloodstream by a feeding *Anopheles* sp. mosquito and (i) fertilise in the mosquito midgut to form zygotes. These zygotes further differentiate into motile forms, called ookinetes, migrate through the mosquito gut wall and divide within oocysts on the external gut wall to form thousands of sporozoites. (j) The infective sporozoites are released into the mosquito haemocoele and move to the salivary gland, where they await injection into another human host, thus completing the life cycle **(fig001dcn)**.

P. falciparum (Refs 32, 33) and most recently to *P. vivax* (Ref. 34), presumably owing to the similar nucleotide composition of *Plasmodium* spp. and the yeast *Saccharomyces cerevisiae*. This strategy is being used extensively by malaria researchers (Ref. 35).

Sequencing the *P. falciparum* genome

In designing sequencing strategies for the *P. falciparum* genome project, the consortium focused first on several technical hurdles. The first one was the concern that the high A-T bias and observed inability to produce representative large-insert genomic libraries in *E. coli* would

https://doi.org/10.1017/S146239949800012X Published online by Cambridge University Press

Accession number: txt001dcn; **Original accession date:** 5 May 1998;
Revision version accession date: 10 June 1998, **Archived previous version**: yes (txt001dcn5May98)

Table 1. Sequencing of *Plasmodium falciparum* chromosomes (tab001dcn)

Chromosome	Size (in Mb)	Sequence centre	Status/comment
1	0.65	Sanger	Partially sequenced
2	1.0	TIGR/NMRI	Sequence completed
3	1.2	Sanger	Nearing completion of sequence
4	1.4	Sanger	Partially sequenced
5	1.6	Sanger	
6	1.6	Sanger	
7	1.7	Sanger	
8	1.7	Sanger	
9	1.8	Sanger	
10	2.1	TIGR/NMRI	
11	2.3	TIGR/NMRI	
12	2.4	Stanford	Partially sequenced
13	3.2	Sanger	
14	3.4	TIGR/NMRI	

Abbreviations used and links to further sequence information: Mb = megabase pairs; Sanger = the Sanger Centre (Hinxton, UK), http://www.sanger.ac.uk/Projects/P_falciparum; TIGR = The Institute for Genomic Research (Rockville, MD, USA), http://www.tigr.org/tdb/mdb/pfdb/pfdb.html; NMRI = Naval Medical Research Institute (Rockville, MD, USA), http://www.nmri.nnmc.navy.mil; Stanford = Stanford University (Stanford, CA, USA), http://sequence-www.stanford.edu/group/malaria/index.html

exclude the possibility of using large-insert bacterial libraries to sequence the whole genome using the BAC-end sequencing approach (Ref. 20). The second hurdle was that although YAC libraries of *P. falciparum* were available (Refs 32, 33) YACs were not considered to be good substrates for sequencing; also, bacterial subclone libraries derived from YACs were notorious for their contamination with yeast chromosomal DNA. Finally, the large size of the malaria parasite genome and the intention to divide the sequencing efforts among several laboratories meant that a whole-genome shotgun approach was not practical because sequencing efforts could not be easily partitioned.

Sequencing strategies for *P. falciparum*

The consortium agreed on an approach based on the fact that most of the 14 plasmodium chromosomes can be separated by pulsed-field gradient gel electrophoresis (PFGGE). In fact, using this commonly used molecular biology technique (Ref. 36), over 80% of the genome from *P. falciparum* can be separated as individual chromosomes; the remaining 20% that consists of five co-migrating chromosomes cannot be separated in this way. A decision was made, therefore, to approach the sequencing of the malaria genome one chromosome at a time. The plan was to separate those *P. falciparum* chromosomes that do not co-migrate by PFGGE and treat each one as if it were an individual 1–3-Mb microbial sequencing project. Mapping data that were already available from the ongoing *P. falciparum* Genome Mapping Project, sponsored by the Wellcome Trust (Ref. 35), would provide important information for the process of closing the gaps. Individual chromosomes that were to be sequenced were assigned to three genome centres: TIGR (in conjunction with the NMRI), the Sanger Centre and Stanford University (Table 1, tab001dcn). Random-shotgun libraries that were specific for *P. falciparum* chromosomes were prepared from chromosomes purified using PFGGE and the clones in the libraries were sequenced. In addition, some sequencing centres have used some YAC-based sequencing. At the time of writing, the sequencing of the 1-Mb chromosome 2 (at TIGR/NMRI) and the 1.2-Mb chromosome 3 (at the Sanger Centre) is nearly complete, and significant progress has been made on chromosome 12 (at Stanford University) and on several other chromosomes.

Research in progress and outstanding research questions

The successful sequencing of the first two of the 14 chromosomes of *P. falciparum* has proven the feasibility of completing the entire genome of this parasite. Malaria researchers and genome centres

https://doi.org/10.1017/S146239949800012X Published online by Cambridge University Press

Accession number: txt001dcn; **Original accession date:** 5 May 1998;
Revision version accession date: 10 June 1998, **Archived previous version**: yes (txt001dcn5May98)

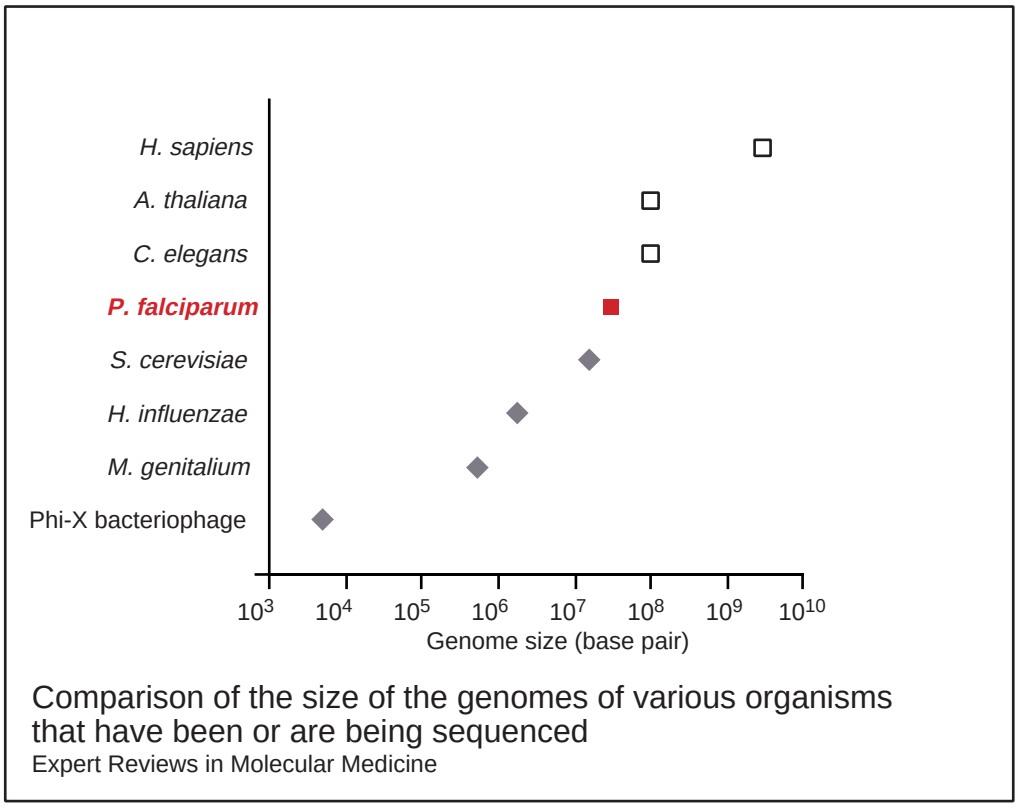


Figure 2. Comparison of the size of the genomes of various organisms that have or are being sequenced. The diamond symbols denote organisms that have been sequenced: phi-X bacteriophage = ΦX174 (a virus that infects bacteria); *M. genitalium* = *Mycoplasma genitalium* (a bacterium that infects the urogenital tract of humans; it has no cell wall and is the smallest known cellular organism, that can reproduce independently); *H. influenzae* = *Haemophilus influenzae* (a bacterium that infects humans, giving rise to one form of meningitis); *S. cerevisiae* = *Saccharomyces cerevisiae* (budding yeast). The square symbols denote organisms that are being sequenced: *P. falciparum* = *Plasmodium falciparum* (a malaria parasite that infects humans); *C. elegans* = *Caenorhabditis elegans* (a nematode that is used for developmental biology studies); *A. thaliana* = *Arabidopsis thaliana* (thale cress, a model system for plant biologists); *H. sapiens* = *Homo sapiens* (humans) **(fig002dcn)**.

must now develop the necessary strategies to use these genomic data. The enormous amount of information generated by this project will need to be translated to facilitate the development of vaccines, new drugs and experimental reagents to study the complex biochemical pathways of this parasite and the mechanisms of drug resistance. Research will no longer be restricted to single-gene approaches; it will soon be possible to consider the entire genetic complement of hundreds or thousands of genes at a time. This information alone will not be enough; integrated information–database management will need to be established to allow malaria researchers throughout the world to access and manipulate the enormous amounts of data generated from the project: from individual gene sequences, to protein predictions, to expression data. These data will need to be linked to other genome databases so that they too can be exploited. Indeed, as advances in genomics continue, these databases need to be sufficiently flexible to be able to incorporate new genomic data derived from novel technologies. Insight into the very 'core' of the malaria parasite will provide the best, if not the only, remaining means by which the devastating impact that malaria has on humans can be reduced and might explain, at least in part, the human interaction with this parasite.

Acknowledgements and disclaimer

The opinions and assertions herein are those of the authors and are not to be construed as official or as reflecting the views of the US Navy or naval service at large. The work was supported by the Office for Research on

https://doi.org/10.1017/S146239949800012X Published online by Cambridge University Press

Minority Health of the National Institutes of Health and by the Naval Medical Research and Development Command work units STO F 6.3a63002AA0101HFX, STOF6.161102 AA0101BFX, STOF 6.262787A00101EFX and STEP C611102A0101BCX.

References

1 Sanger, F. et al. (1977) Nucleotide sequence of bacteriophage phi X174 DNA. Nature 265, 687–695
2 Reddy, V.B. et al. (1978) The genome of simian virus 40. Science 200, 494–502
3 Fiers, W. et al. (1978) Complete nucleotide sequence of SV40 DNA. Nature 273, 113–120
4 Dunn, J.J. and Studier, F.W. (1983) Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. J. Mol. Biol. 166, 477–535
5 Sanger, F. et al. (1982) Nucleotide sequence of bacteriophage lambda DNA. J. Mol. Biol. 162, 729–773
6 Hiratsuka, J. et al. (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol. Gen. Genet. 217, 185–194
7 Baer, R. et al. (1984) DNA sequence and expression of the B95-8 Epstein–Barr virus genome. Nature 310, 207–211
8 Watson, J.D. (1990) The human genome project: past, present and future. Science 248, 44–49
9 Smith, L.M. et al. (1986) Fluorescence detection in automated DNA sequence analysis. Nature 321, 674–679
10 Adams, M.D. et al. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252, 1651–1656
11 Adams, M.D., Kerlavage, A.R. and Venter, J.C. (1993) 3,400 new expressed sequence tags identify diversity of transcripts in human brain. Nat. Genet. 4, 256–267
12 Adams, M.D. et al. (1993) Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. Nat. Genet. 4, 373–380
13 Adams, M.D. et al. (1994) A model for high-throughput automated DNA sequencing and analysis core facilities. Nature 368, 474–475
14 Marshall, E. (1995) Emphasis turns from mapping to large-scale sequencing. Science 268, 1270–1271
15 Fleischmann, R.D. et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269, 496–512
16 Bloom, B.R. (1995) Genome sequences. A microbial minimalist. Nature 378, 236
17 Shizuya, H. et al. (1992) Cloning and stable maintenance of 300-kilobase-pair fragment of human DNA in *Escherichia coli* using an F-factor-based vector. Proc. Natl. Acad. Sci. USA 89, 8794–8797
18 Burke, D.T., Carle, G.F. and Olson, M.V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 236, 806–812
19 Monaco, A.P. and Larin, Z. (1994) YACs, BACs, PACs and MACs: artificial chromosomes as research tools. Trends Biotechnol. 12, 280–286
20 Venter, J.C., Smith, H.O. and Hood, L. (1996) A new strategy for genome sequencing. Nature 381, 364–366
21 Fraser, C.M. et al. (1995) The minimal gene complement of *Mycoplasma genitalium.* Science 270, 397–403
22 Bult, C.J. et al. (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. Science 273, 1058–1073
23 Hoffman, S.L. et al. (1997) Funding for the malaria genome sequencing project. Nature 387, 647
24 Anon. (1994) World Malaria Situation in 1992. Weekly Epidemiol. Rec. 69, 309–314
25 Patz, J.A. et al. (1996) Global climate change and emerging infectious diseases. JAMA 275, 217–223
26 Sharp, D. (1996) Malaria range set to spread in a warmer world. Lancet 347, 1612
27 Baird, J.K. et al. (1997) Diagnosis of resistance to chloroquine by *Plasmodium vivax*: timing of recurrence and whole blood chloroquine levels. Am. J. Trop. Med. Hyg. 56, 621–626
28 Hoffman, S.L. and Miller, L.H. (1996) Perspectives on malaria vaccine development, in Malaria Vaccine Development: A Multi-Immune Response Approach (Hoffman, S.L., ed.), ASM Press, Washington, DC, pp. 1–17
29 Weber, J.L. (1988) Molecular biology of malaria parasites. Exp. Parasitol. 66, 143–170
30 Pollack, Y.T. et al. (1982) The genome of *Plasmodium falciparum* I. DNA base composition. Nucl. Acids Res. 10, 539–546
31 Goman, M. et al. (1982) The establishment of genomic DNA libraries for the human malaria parasite *Plasmodium falciparum* and identification of individual clones by hybridisation. Mol. Biochem. Parasitol. 5, 391–400

https://doi.org/10.1017/S146239949800012X Published online by Cambridge University Press

32 Triglia, T. and Kemp, D.J. (1991) Large fragments of *Plasmodium falciparum* DNA can be stable when cloned in yeast artificial chromosomes. Mol. Biochem. Parasitol. 44, 207–212
33 de Bruin, D., Lanzer, M. and Ravetch, J.V. (1992) Characterization of yeast artificial chromosomes from *Plasmodium falciparum*: construction of a stable, representative library and cloning of telomeric DNA fragments. Genomics 14, 332–339
34 Camargo, A.A. et al. (1997) Construction and characterization of a *Plasmodium vivax* genomic library in yeast artificial chromosomes. Genomics 42, 467–473
35 Dame, J.B. et al. (1996) Current status of the *Plasmodium falciparum* genome project. Mol. Biochem. Parasitol. 79, 1–12
36 Schwartz, D.C. and Cantor, C.R. (1984) Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell 37, 67–75

Further reading, other resources and other contacts

The Malaria Project genome centres:

The Sanger Centre (Hinxton, UK).
http://www.sanger.ac.uk/Projects/P_falciparum

The Institute for Genomic Research (Rockville, MD, USA).
http://www.tigr.org/tdb/mdb/pfdb/pfdb.html

Naval Medical Research Institute (Rockville, MD, USA).
http://www.nmri.nnmc.navy.mil

Stanford University (Stanford, CA, USA).
http://sequence-www.stanford.edu/group/malaria/index.html

The Malaria Foundation (a global network against malaria) website.
http://www.malaria.org

The Malaria Database is an information resource for malaria researchers, and is maintained at The Walter and Eliza Hall Institute of Medical Research (Melbourne, Australia).
http://www.wehi.edu.au/MalDB-www/who.html

The (US) National Center for Biotechnology Information (NCBI) and the (US) National Institutes of Health (NIH) National Institute of Allergy and Infectious Disease (NIAID) maintain a website on malaria genetics and genomics (new URL, access shouldn't be restricted).
http://www.ncbi.nlm.nih.gov/Malaria

Tables
Table 1. Sequencing of *Plasmodium falciparum* chromosomes (tab001dcn).

Schematic figures
Figure 1. Life cycle of *Plasmodium falciparum* malaria parasite (fig001dcn).
Figure 2. Comparison of the size of the genomes of various organisms that have or are being sequenced (fig002dcn).

https://doi.org/10.1017/S146239949800012X Published online by Cambridge University Press