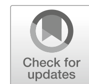

Parasitic Genomics

Sumeeta Khurana and Parakriti Gupta

Learning Objectives

1. To know the tools used for genomic studies and the targets.
2. To have an understanding about the application of genetic studies in parasitology.

Introduction

Parasites are unique among all the microorganisms as most have complex life cycles involving one or more hosts, are difficult to culture in the laboratory, and lack suitable experimental models. The advent of novel approaches to genetic study and manipulation has promoted the research on parasites. International parasitic genome networks are now established and have resulted in an exponential increase in genomic data for parasites. All this data is stored in databases and can be accessed online and used for structural and functional analyses. Availability of genomic data has altered the way infectious diseases are studied. However, this mammoth big data serves no purpose unless it is interpreted as gene annotation and error correction are still exhaustive and critical. Prediction of the gene function is still a major challenge. Lately, *metagenomics* is making possible analysis of the relationship of complex microbial communities, especially those which cannot be cultured.

Beginning of Parasitic Genomic Era

Decrypting of the whole genome of most parasites has been successful owing to their smaller genome size of approximately 10–270 megabases (Mb) (Table 1). However, parasitic genomes vary in size, nucleotide composition, content, polymorphism, and repetitive sequences, all of which affect the feasibility and application of sequencing strategies. The groundbreaking landmark was witnessed in 2002, when the genomic sequence of protozoon *Plasmodium falciparum* was published as a product of international collaborative determinations. Subsequently, genomes of *Trypanosoma cruzi*, *Trypanosoma brucei*, and *Leishmania major* and now many more parasites have been decoded. Helminths have a much larger genome compared to protozoa but very small when compared to mammals. However, they contain almost the same number of genes as humans. However, in nematodes, genes are gained and lost frequently, and there is horizontal gene transfer from bacteria, fungi, amoeba, or endosymbionts. The free-

S. Khurana (✉) · P. Gupta
Department of Medical Parasitology, Post Graduate Institute of Medical Education and Research, Chandigarh, India

© Springer Nature Singapore Pte Ltd. 2022

S. C. Parija, A. Chaudhury (eds.), *Textbook of Parasitic Zoonoses*, Microbial Zoonoses,
https://doi.org/10.1007/978-981-16-7204-0_4

Table 1 Genomic sizes of common parasites of humans and animals

Species	Host	Genome size (in Mb)
Giardia duodenalis	Humans	12.6
Entamoeba histolytica	Humans	24
Plasmodium falciparum	Humans	22.8
Ancylostoma caninum	Dogs	344
Ascaris lumbricoides	Humans	230
Brugia malayi	Humans	96
Onchocerca volvulus	Humans	150
Trichinella spiralis	Humans, pigs	63
Trichinella muris	Mouse	96
Echinococcus multilocularis, *Echinococcus granulosus*	Humans, rodents	150
Taenia solium	Humans	270
Schistosoma mansoni	Humans	390
Schistosoma japonicum	Humans	400

living nematode, *Caenorhabditis elegans*, is the first and one of the best-studied metazoan parasites whose genome has been fully sequenced. In recent years, the draft genomic sequences of many other parasites have been made available.

The molecular techniques available for decrypting the genome of parasites include Sanger's sequencing (Fig. 1), microsatellite markers, microarray, Luminex (multianalyte profiling), random amplification of polymorphic DNA or arbitrarily primed PCR, restriction fragment length polymorphism, amplified fragment length polymorphism, whole-genome sequencing, RNA interference, and bioinformatics approaches (Table 2). The relatively new next-generation sequencing technology has enabled analysis of genomic sequence with comparative genomics, functional genomics, transcriptomics, metabolomics, proteomics, and epigenetics.

The data obtained from sequencing studies is interpreted in three categories: (1) complete or nearly complete genomic sequences as *contigs* of series of overlapping DNA sequences, (2) genome-survey sequence tags (GSS)

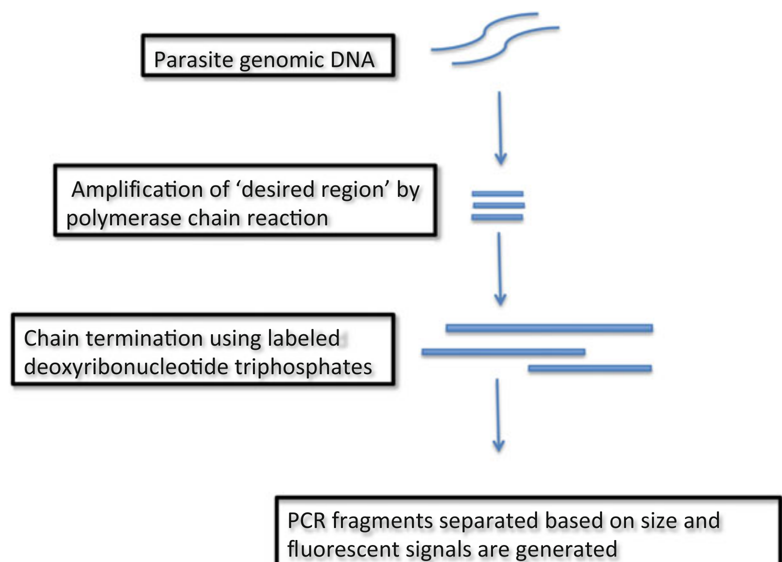

Fig. 1 Flowchart depicting Sanger sequencing in parasites

Table 2 Molecular techniques available for decrypting genome of parasites

Molecular techniques	Examples of parasites
Sanger sequencing	*Leishmania*, *Plasmodium*, *Echinococcus*, and *Trypanosoma*
Microsatellite markers	*Plasmodium*, *Ascaris*
Microarray	*Plasmodium*, *Toxoplasma*, and *Trypanosoma*
Multianalyte profiling	*Plasmodium*, *Cryptosporidium*, and *Leishmania*
Random amplification of polymorphic DNA	*Plasmodium*, *Leishmania*, *Echinococcus*, and *Trypanosoma*
Restriction fragment length polymorphism	*Cryptosporidium*
Amplified fragment length polymorphism	*Cryptosporidium* and *Leishmania*
RNA interference	*Plasmodium*, *Giardia*, and *Entamoeba*
Whole-genome sequencing	*Plasmodium*
Bioinformatic approaches	*Plasmodium*

generated after skimming of the genomic sequences, and (3) expressed sequence tags (ESTs) generated from the respective mRNA expressed in different stages of the parasitic life cycle. All these data are stored in databases and can be accessed online at their own sites, but most of the data is available via *GenBank* and by web-based *BLASTSearch*. These data can be used for structural and functional analyses. The Institute for Genomic Research (TIGR), founded in 1992, located in Maryland, USA, is involved in the sequencing of the genomes of prokaryotic and eukaryotic organisms and post-sequence analysis. One of the main objectives for generating the sequence data is to identify genes related to parasite evolution, development, metabolism, pathogenicity, immune evasion, diagnostic markers, etc.

Application in Parasitology

Genomics has manifold applications in parasitology (Table 3), some of which are discussed here.

Parasitic Diagnosis

Conventionally, diagnosis of parasitic infections has relied upon microscopic demonstration of different life stages of parasites. However, these techniques are often insensitive to warrant exploration of newer, more sensitive modalities. Factors responsible for failure of conventional techniques include lower parasite numbers in comparison to bacteria and viruses and presence of morphologically similar nonpathogens. Parasites are usually not cultured routinely due to lack of expertise and facility to culture, longer turnaround time, and parasite fastidiousness to grow. Nucleic acid amplification technologies, especially polymerase chain reaction (PCR) and its modifications, are increasingly being used to facilitate parasitic diagnosis, especially in the developed centers. For example, amebiasis is routinely diagnosed in fecal samples or liver pus by PCR as it can differentiate *Entamoeba histolytica* from the nonpathogenic morphologically identical *Entamoeba dispar* with excellent sensitivity and specificity. Similarly, kinetoplast DNA, 18S, and ITS regions are used as targets for identification of *Leishmania* species.

Unlike direct pathogen identification, serology is frequently employed for diagnosis of infectious diseases. However, it has a number of limitations. Firstly, serological response, especially specific antibodies, can only be detected in serum after a lapse of a few days to weeks of infection. Sometimes distinction between current and past infection is not possible as antibodies can persist for a period of months to years. Moreover, sensitivity of serological tests varies with the organism and host immune system and may be unreliable in immunocompromised conditions. One such example of appropriate use of technology is diagnosis of toxoplasmosis during pregnancy which was conventionally diagnosed by serology. PCR on amniotic fluid has become the new gold standard for detection of maternally transmitted toxoplasma infection to the fetus.

Table 3 Applications of genomic studies in parasites

Entity	Examples of parasites
Parasitic diagnosis	*Entamoeba* spp.
Parasite discovery	*Plasmodium knowlesi*
Host–parasite interactions	*Brugia malayi* and *Wolbachia*
Host susceptibility studies	*Trypanosoma*, *Schistosoma*, and *Plasmodium*
Molecular mimicry and antigenic variation	*Plasmodium* and *Trypanosoma*
CRISPR-Cas studies	*Anopheles*, *Plasmodium*, *Trypanosoma*, and *Leishmania*
Epidemiological tracking	*Giardia*
Drug discovery and resistance	*Plasmodium*
Vaccine development	*Toxoplasma*, *Echinococcus*

Pathogen Discovery

In addition to diagnosis, many new parasites infecting humans have been discovered recently and are the result of application of genomic technologies like whole-genome sequencing, metagenomics, etc. For example, *Plasmodium knowlesi*, which was initially misidentified as *Plasmodium malariae*, could be discovered only by sequencing and has now been established as the fifth human malaria species originating from monkeys. In addition, many new parasites are being discovered owing to genomic technologies, e.g., *Bertiella*, *Taenia asiatica*, etc.

Host–Parasite Interactions

Apart from studying the genomic characterization of parasites on their own, extensive research has also been carried out in scrutinizing host–parasite interactions. A landmark study described the prototypic endosymbiotic interaction between *Brugia malayi* and *Wolbachia* after whole-genome sequencing of the *B. malayi* genome was deciphered. *Wolbachia* as a bacterial endosymbiont is required for filarial multiplication, and lipopolysaccharide present in the bacterial cell wall also acts as one of the most potent virulence factors for the parasite, inducing a profound inflammatory reaction. This discovery was subsequently sustained by a highly effective therapeutic response to tetracycline that acts on *Wolbachia* and not *B. malayi*, which resulted in dramatic decline in filarial burden and is currently under research as a potential vaccine candidate. *Plasmodium* pathogenesis is increasingly being explored utilizing genomics, proteomics, and transcriptomics. For example, several Toll-like receptors, e.g., TLR-2 and TIR-domain carrying adaptor molecule 2, have been associated with difference in clinical severity based on genomic studies.

Host Susceptibility Studies

Parasites successfully subsist in the host, which can be attributed to very specific genetic adaptations. Thus, detection of such genes which are important for this adaptation can be utilized to understand disease and propose a cure. Genomic studies are the key to establishing the parasite–host–microbe relationship by analyzing the helminth-induced changes in human gut using transcriptional repeats following parasitic infection. This can be further exploited to delineate the role of gut microbiota in parasitic infections and in selecting novel targets for limiting infections, novel drugs, and vaccines. Furthermore, host susceptibility, severity, and mortality of infection are also exposed by these extensive genetic studies.

It is a long-established fact that patients with sickle cell anemia and thalassemia and those with absence of glycoprotein receptors are resistant to malarial infection. Similarly, other host cell receptors like CD234 have been associated with susceptibility to infection. CD234 is essential for *Plasmodium vivax* entry into red blood cells, and

FY mutation deleting this CD234 provides protection from malaria. Individuals with ENU mutation in ankyrin-1, a significant membrane protein, with erythropoietic protoporphyria with decreased levels of ferrochelatase are noted to be resistant to cerebral malaria. In contrast, patients with mdr1a mutation in P-glycoprotein, CD36 deficiency, apobec3b deficiency, and metallobetalactamase-2 deficiency are more susceptible to cerebral malaria. Some mutations making the host susceptible or resistant to infections have also been noted in other parasites e.g, *nramp* 1 mutation and *lpl* gene mutation in *Leishmania*, single nucleotide polymorphisms in STAT-3 in *Entamoeba*, and *apo2–1* mutation in *Trypanosoma* increase susceptibility of host to infection.

Molecular Mimicry and Antigenic Variation

Genomic approaches have been utilized by researchers to identify the phenomenon of molecular mimicry in *Plasmodium* species, exhibited by members of the KIR family. Many genes have a high level of identity with the molecular domains of CD99, an immunoregulatory protein present on the membrane of lymphocytes and T cells. Furthermore, transcriptomic studies have shown the sharing of *P. falciparum* genes to short sequence motifs upstream of ATGs. Site-directed mutagenesis validated their role in promoter activity, especially AP2 in the ookinete stage. However, transcription of all the genes including the antigenic variation is further regulated at chromosomal level. Since the antigenic variation is not present in all parasites, researchers conducted extensive studies to disclose the basis and correlation between the antigenic variation of *Plasmodium* spp. and *Trypanosoma* spp. Epigenetic and chromatin immunoprecipitation studies unearthed the role of histone enzymes, histone methylase, and deacetylase in switching and antigenic variation.

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Associated (CRISPR-Cas) in Parasites

Though the parasitic genome was decoded quite late compared to other microbes, CRISPR-Cas technology has been exploited in *Plasmodium*, *Leishmania*, *Trypanosoma*, platyhelminths, as well as vectors like *Anopheles* as a genome-editing tool. In *Plasmodium*, gene expression can be modulated even in the absence of genome editing via usage of CRISPR interference or activation (CRISPRi/a). The role of CRISPR-Cas in *Anopheles* is reasonably promising in the field of *gene drive*, which refers to the genetic systems in which a specific trait is selectively transmitted between the populations in a selfish manner defying the normal Mendelian inheritance. The main goal of this gene drive is to negatively control mosquito fertility by use of genetic modifications. CRISPR-Cas is exploited in identification of genes involved in infectious process; subsequently, gene libraries are formulated and can be exploited in the generation of immunogenic or non-virulent or nonpathogenic parasites serving as candidates for drugs or vaccines.

Epidemiological Tracking

Newer whole-genome sequencing is now a major tool to discriminate between closely related strains and track real-time evolution of disease-associated clonal isolates and plays an important role in epidemiological investigations. The commonly employed methods used to discriminate the strain relatedness in such scenarios include RFLP, AFLP, RAPD, and PFGE.

Drug Discovery

Primarily through genomic technology, novel promising targets are identified, and appropriate expansion of these functional genes is

accomplished, followed by assay for high-throughput sequencing. The genome can provide information about the biochemical pathways likely to be involved for a proposed drug or treatment. Thus drug discovery has been able to be expedited tremendously compared to conventional approaches.

Drug Resistance

One of the strongest examples in this scenario is the artemisinin resistance in malarial parasite *P. falciparum*, in which kelch 13 mutations revealed by genomic data could help in framing a tiered approach, focusing on the areas with emergence of resistance. Malaria GEN is one such approach to containing the spread of resistance by enhancing the global research based on sequencing of a large number of samples for chloroquine and pyrimethamine resistance lineages and *Anopheles* vector. Another added advantage of this strategy is exacting the molecular basis of resistance using expression profiling. Screening using microarray and serial analysis of gene expression is of immense importance in localizing and predicting the role of a particular gene product at the right time for the right patient in the right place as an immunological target. Resistance studies have also been conducted to analyze the selection pressure and spread of resistance genes as in *Plasmodium* and *Leishmania*.

Vaccine Development

Very few vaccines are available for parasitic infections like toxoplasma in sheep, echinococcosis in animals, and malaria in humans. Completed genomes will provide the source material for vaccine development. Genomic *chinks in the armor* of parasites screened by expression of libraries with immune sera help to identify the candidate antigens, inciting an immune response and the cryptic ones that fail to elicit immune response and evade the immune system. Other strategies for vaccine development include mRNA-based techniques, differential display, and serial analysis of gene expression. Expressed sequence tags (ESTs) are identified in clusters followed by generation of their consensus sequences that enables the rapid assembly of data. Candidate antigens that can activate Toll-like receptors are also being identified.

Conclusion

Thinking beyond the role of genomic analysis, functional manipulations and editing is the need of the hour to analyze the information obtained after sequencing and to scrutinize the resulting hypothesis. Though maintenance of parasites is challenging, transgenesis and CRISPR-Cas experiments have already been achieved in some parasites. The choice of an appropriate parasitic model being edited, utilizing older and newer molecular tools, can help us comprehend the basic biology of these parasites, simultaneously deciphering the ability to manipulate and control the human illnesses caused by them and exploiting an interactive and collaborative team of molecular parasitologist, epidemiologist, infectious disease physician, and computational and data analyst.

Case Study

The application of genomic studies has been elegantly brought out in one report regarding the population genetic analysis of Guinea worms (*Dracunculus medinensis*) collected from Chad, Africa. There was an apparent reemergence of human Guinea worm disease in Chad after an almost 10-year absence of reported cases. It also coincided with a recent finding of high prevalence of Guinea worm infection in dogs in Chad. To determine whether worms from human and non-human hosts were, in fact, the same species, the worms were collected from both human and non-human hosts. Genetic variation was measured in these worms using sequence variation of mitochondrial DNA genes and repeat number polymorphism at 23 nuclear microsatellite loci. It was found that Guinea worms collected from

nonhuman hosts were *D. medinensis* and that the same population of worms infects both humans and dogs in Chad. This genetic data and the epidemiological evidence suggest that transmission in the Chadian context is currently being maintained by canine hosts.

1. Name some other zoonotic parasites for which similar kinds of study have been conducted.
2. List the common techniques by which strain relatedness among parasites is determined.
3. Name the parasites for which genetic studies have been done to determine drug resistance.
4. Enumerate the important target genes which have been identified and used as molecular markers for identification and quantification of *Leishmania* in clinical samples.

Research Questions

1. What are the fundamental cellular mechanisms involved in host–parasite interactions?
2. What is the molecular basis of response to anti-parasitic drugs and vaccines?
3. What is the genetic basis of emergence and reemergence of parasitic infections?

Further Readings

Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, et al. Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. Nature. 2008;455(7214):757–63. PubMed [Internet]. [Cited 2020 Mar 22]. Available from: https://pubmed.ncbi.nlm.nih.gov/18843361/

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature. 2002;419(6906):498–511. PubMed [Internet]. [Accessed 2020 Mar 22]. Available from: https://pubmed.ncbi.nlm.nih.gov/12368864/

Tarleton RL, Kissinger J. Parasite genomics: current status and future prospects. Curr Opin Immunol. 2001;13(4): 395–402. PubMed [Internet]. [Accessed 2020 Mar 5]. Available from: https://pubmed.ncbi.nlm.nih.gov/11498294-parasite-genomics-current-status-and-future-prospects/

Tavares RG, Staggemeier R, Borges ALP, Rodrigues MT, Castelan LA, Vasconcelos J, et al. Molecular techniques for the study and diagnosis of parasite infection. J Venom Anim Toxins Incl Trop Dis. 2011;17(3):239–48.

Thiele EA, Eberhard ML, Cotton JA, Durrant C, Berg J, Hamm K, et al. Population genetic analysis of Chadian Guinea worms reveals that human and non-human hosts share common parasite populations. PLoS Negl Trop Dis. 2018;12(10):e0006747.

Winzeler EA. Advances in parasite genomics: from sequences to regulatory networks. PLoS Pathog. 2009;5(10):e1000649.