



OPEN ACCESS

EDITED AND REVIEWED BY

Alex Loukas,
James Cook University, Australia

*CORRESPONDENCE

Makedonka Mitreva
mmitreva@wustl.edu

SPECIALTY SECTION

This article was submitted to
Parasitology Omics,
a section of the journal
Frontiers in Parasitology

RECEIVED 15 July 2022

ACCEPTED 20 July 2022

PUBLISHED 05 August 2022

CITATION

Mitreva M (2022) Parasite OMICS, the
grand challenges ahead.
Front. Parasitol. 1:995302.
doi: 10.3389/fpara.2022.995302

COPYRIGHT

© 2022 Mitreva. This is an open-access
article distributed under the terms of
the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution
or reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Parasite OMICS, the grand challenges ahead

Makedonka Mitreva*

Department of Medicine, Washington University School of Medicine, St. Louis, MO, United States

KEYWORDS

parasites, genomics, transcriptomics, proteomics, metabolomics, omics, applied genomics

Background and introduction

Great progress has been made in parasitology as a result of the advancements made in both omics data generation and omics data interrogation approaches. The OMICS journey for many parasites started over 2 decades ago by generating and analyzing so called “expressed sequence tags” that profiled the most abundantly expressed transcripts in an organism at the time of sampling (Clifton and Mitreva, 2009). Improvements of these transcriptional profiles were technology-driven, starting with advancements in sample preparation and transitioning from traditional to more advanced sequencing platforms (from Sanger to Roche/454, and eventually the Illumina, PacBio and Oxford Nanopore platforms). With the new sequencing platforms, the amount of data that could be generated increases while the cost rapidly decreases, so the parasitology community continually adapts to the newest platforms and optimizes the methodology. As a result, some model species such as *Caenorhabditis elegans* have been resequenced repeatedly as each new platform became available, while new genomes for important parasites started being produced and improved. Parasites with smaller and simpler genomes were benefiting from these technologies earliest. The first eukaryotic genome to be published was the human malaria parasite *Plasmodium falciparum* in 2002, reporting a 23 Mb nuclear genome that encodes about 5,300 genes (Gardner et al., 2002). In 2005, the genomes of the three protozoan kinetoplastids were published: *Leishmania major* [33 Mb genome assembly containing 8,272 protein coding genes (Ivens et al., 2005)], *Trypanosoma cruzi* [67 Mb and ~12,000 genes (El-Sayed et al., 2005a)] and *T. brucei* [26 Mb and 9,068 genes (Berriman et al., 2005)], followed by a genome-wide comparison of these three members of the family *Trypanosomatidae* (El-Sayed et al., 2005b). By 2008, a 100 years after *Toxoplasma gondii* was initially described in Tunisia by Nicolle and Manceaux (1908) (Kim and Weiss, 2008), a total of 12 *Toxoplasma* genomes were available including the *T. gondii* ME49 assembly (63 Mb and 8,032 genes). The first genome of a more complex multicellular eukaryotic parasite was that of the roundworm *Brugia malayi* [90 Mb and 11,500 genes (Ghedini et al., 2007)], published in 2007, 9 years later than the *C. elegans* genome (*C. elegans* Sequencing Consortium, 1998), and in 2009 the first flatworm *Schistosoma mansoni* [363 Mb, 11,809 genes (Berriman et al., 2009)] was published. In the next 7 years a total of 11 parasitic nematode genomes were published, when a “50 helminth genome initiative” was launched, resulting in the biggest data generation (45 new genomes) and comparative genome analysis published 5 years later (International Helminth Genomes Consortium, 2019). Genome size varied greatly from 42 to 700 Mb within nematodes, and from 104 to 1,259 Mb within platyhelminths,

and this comparative study of the two major parasitic helminth phyla included 1.4 million genes and 81 helminth species.

As more genomes became available, it became clear that variations in parasite biology, physiology, mode of parasitism and tissue tropism were reflected in their genomes, as their varying size and complexity. The initially published genomes and transcriptomes revealed novel insights into the many parasites that occupy human, animal or plant hosts. However, there are many OMICS related challenges ahead of us that will require our immediate focus, including (but not limited to): (a) improving available genomes that are fragmented and/or inadequately annotated, and sequencing underrepresented species; (b) expanding on the types of OMICS data for many parasite species, and (c) applying the acquired OMICS-driven knowledge toward translational studies. These challenges are described below in more detail and are critically needed to better understand the complex biology of the parasites and advance their diagnosis, prevention, and control.

Improving omics resources

Despite the progress in parasite genome production and improvement, the draft nature of many parasite genomes poses challenges for analysis and interpretation of the results, especially since post-genomic applications frequently require comparative genomics on a gene and single nucleotide level. Performing these analyses on draft genomes is inadequate due to gene fragmentations, gene model errors resulting from misassembly of allelic sequences as separate loci, the collapse of recently duplicated and diverged sequences into a single locus, and the large numbers of unordered contigs within scaffolds. The resulting incomplete or incorrect gene models for parasite species adversely affects many important analyses including (i) parasite drug target identification, since known or putative protein targets may not be properly annotated and therefore a drug's effect will not be predicted), (ii) homology modeling, for which accurate protein structure and active residue locations are critical for modeling drug interactions correctly, and (iii) the identification and classification of parasite-host interacting proteins, which are often difficult to annotate due to their diversification within and across species. In addition to the problems associated with gene model errors, highly fragmented genome assemblies negatively affect the analysis of selection signatures and identification of quantitative trait loci in experimental crosses and natural populations. Therefore, it is of pivotal importance that we continue to improve parasite assemblies and annotations using newly available technologies. Some examples of improvements include using HiFi PacBio long read sequencing, Oxford Nano Pore (ONT UL sequencing) and scaffolding with HiC linked reads, resulting in more contiguous assemblies. Direct reannotation of existing assemblies by single molecule PacBio mRNA sequencing can also be used to improve

eukaryotic parasites genome annotation (Magrini et al., 2018) along with using multiple functional omics data and the newest available computational tools (Logan et al., 2020). While such improvements may not be as critically needed for some parasites that have smaller and/or simpler genomes, the genomic complexity of some species makes proper genome assembly and annotation very challenging; for example, the liver fluke *Fasciola hepatica* genome is 1.2 Gb, with 65% of the genome being repetitive. One of the big challenges ahead is that once the genome is published it is difficult to justify to funding agencies effort needed for improvement, so advocating for the importance of iterative improvements is essential. In addition to genome improvement, expanding the genomic resources for more parasitic species is equally important, as the genomes of some parasites of socio-economic, veterinary, or agricultural importance have yet to be sequenced.

Expanding the omics repertoire

In addition to their genomes, advanced resources with the potential for major practical applications have been produced for many major parasites. Transcriptional profiles have been produced for whole parasites across their life cycles, in addition to parasite tissue-specific transcriptional profiles and transcriptional profiles of host responses to infection. In many cases, after these two data types (DNA and mRNA based) have become available, the sequencing data has been used to guide proteomics approaches including protein arrays and targeted and untargeted M/S proteomics to characterize peptides and catalog molecules involved in host-parasite interactions (such as excretory/secretory products of parasites) and to identify proteins bound by host antibodies following infection. Metabolomics approaches have also been used to understand gene pathway responses at a small molecule level, which can be used to validate the disruption of pathways thought to be targeted by drugs, or to identify changes in metabolism in the parasite under different conditions. These small biological molecules collectively known as the “metabolome” have a complex interplay in host-parasite interactions, and include molecules derived from/utilized by both the host and parasite, and is an emerging field of study (Whitman et al., 2021). mRNA expression profiling has recently started advancing from “bulk” RNAseq (whole-organism or whole-tissue) sequencing to single-cell RNA-seq (scRNAseq). While scRNAseq datasets are somewhat widely available for some model organisms, the field remains in its infancy for parasites. Similarly, host responses to infections at a single cell level have been documented for bacterial and viral infections, but for parasitology, the initial studies have focused on expression profiles of the parasites at a single cell level. For example, a recent paper used scRNAseq to identify tissue-specific cell types by profiling individual *T. brucei* trypanosomes from midgut, proventriculus, and salivary

glands of infected tsetse flies (Howick et al., 2022). scRNAseq profiles have also been reported for infective and adult stage of *S. mansoni* (Diaz Soria et al., 2020; Wendt et al., 2020), two important stages of the parasite. Even though there is evidence for gene expression based on bulk or scRNAseq data, many of the genes are taxonomically restricted thus hypothetical. CRISPR-Cas9 genome editing approaches and RNAi screening, as reverse genetics methods that are particularly useful to functionally annotate taxonomically restricted parasite genes that do not have orthologs in other organisms. However, while these are robust and high-throughput in some parasites (e.g., kinetoplastids), they still pose a challenge for some parasites with more complex life cycle, and thus have been successfully performed for only a small number of genes in a few helminth species.

Parasites do not exist in isolation in the niche they occupy, so cross kingdom interactions involving the microbiome have also been studied over the last decade. Other OMICS data types have also started to emerge for the study of parasites, including spatial transcriptomics, epigenomics, glycomics, cytogenomics, immunomics to name a few.

Applied omics—Translational aspects

As we have been entering into a post-genomic era, for some parasites, emerging OMICS technologies have been used to advance translational research and in practical applications. For example, traditional bulk genomic and transcriptomic sequencing and analyses have been biased toward the dominant genotype in samples, masking cell-to-cell variation and rare variants. For some species, this issue is being addressed with single-cell genome sequencing approaches, facilitating the quantification of genetic diversity and kinship in complex parasite populations and capturing *de novo* genetic variation (Dia and Cheeseman, 2021). Malaria transmission patterns are also being studied using whole genome sequencing of field isolates, and such approaches are starting to be applied to helminth infections, where efforts are underway to generate genetic tools to tackle post-treatment recrudescence of helminth infections, parasite emergence and spread, and drug resistance evolution. OMICS driven identification of targets essential for parasite survival has been proven to be correct in many cases, based on computational and experimental approaches [e.g., (Taylor et al., 2013; International Helminth Genomes Consortium, 2019; Tyagi et al., 2021; Ferreira et al., 2022)], and taking advantage of parasite specific molecular features to design drug-like compounds with higher potency compared to the host counterpart, and of taxonomically conserved targets to identify cross-clade efficacy against a broad spectrum of parasitic species (Tyagi et al., 2018). Such studies have closed the gap

between genomics and actionable drug discovery, and progress toward lead identification and optimization for development of new antiparasitic drugs with broad spectrum activity and a novel mechanism of action. One of the challenges for global elimination of some parasitic infections such as onchocerciasis is the lack of highly sensitive, specific, and accurate diagnostic tools to detect adult female worms, and thus inform mass drug administration programs. Multi-omics profiling have provided datasets that could be mined for identification of candidates for improved diagnostics (Bennuru et al., 2016); indeed, some studies have mined the data for candidates for serodiagnosis (McNulty et al., 2015), and subsequent studies have fully characterized candidates for different infections (Curtis et al., 2021; Greene et al., 2022).

Different parasites in the ever evolving “OMICS ERA” are on different position of the OMICS spectrum, from resequencing the genomes of field isolates as an applied genomics approach for molecular surveillance, to the development of a systems biology approaches that can significantly contribute to more rational design for vaccines, therapeutics, and diagnostics.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

This study was supported by NIH Grants R01AI59450, R01AI144161, R01AI146353, and R01EY033195. The funding agencies had no role in the design, execution, or publication of this study.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Bennuru, S., Cotton, J. A., Ribeiro, J. M., Grote, A., Harsha, B., Holroyd, N., et al. (2016). Stage-Specific transcriptome and proteome analyses of the filarial parasite *onchocerca volvulus* and its wolbachia endosymbiont. *mBio* 7, e02028–16. doi: 10.1128/mBio.02028-16
- Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., et al. (2005). The genome of the African trypanosome *Trypanosoma brucei*. *Science* 309, 416–422. doi: 10.1126/science.1112642
- Berriman, M., Haas, B. J., LoVerde, P. T., Wilson, R. A., Dillon, G. P., Cerqueira, G. C., et al. (2009). The genome of the blood fluke *Schistosoma mansoni*. *Nature* 460, 352–358. doi: 10.1038/nature08160
- C. elegans Sequencing Consortium (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282, 2012–2018. doi: 10.1126/science.282.5396.2012
- Clifton, S. W., and Mitreva, M. (2009). Strategies for undertaking expressed sequence tag (EST) projects. *Methods Mol. Biol.* 533, 13–32. doi: 10.1007/978-1-60327-136-3_2
- Curtis, K. C., Fischer, K., Choi, Y. J., Mitreva, M., Weil, G. J., Fischer, P. U., et al. (2021). Characterization and localization of antigens for serodiagnosis of human paragonimiasis. *Parasitol. Res.* 120, 535–545. doi: 10.1007/s00436-020-06990-z
- Dia, A., and Cheeseman, I. H. (2021). Single-cell genome sequencing of protozoan parasites. *Trends Parasitol.* 37, 803–814. doi: 10.1016/j.pt.2021.05.013
- Diaz Soria, C. L., Lee, J., Chong, T., Coghlan, A., Tracey, A., Young, M. D., et al. (2020). Single-cell atlas of the first intra-mammalian developmental stage of the human parasite *Schistosoma mansoni*. *Nat. Commun.* 11, 6411. doi: 10.1038/s41467-020-20092-5
- El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A. N., et al. (2005a). The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 309, 409–415. doi: 10.1126/science.1112631
- El-Sayed, N. M., Myler, P. J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., et al. (2005b). Comparative genomics of trypanosomatid parasitic protozoa. *Science* 309, 404–409. doi: 10.1126/science.1112181
- Ferreira, L. L. G., de Moraes, J., and Andricopulo, A. D. (2022). Approaches to advance drug discovery for neglected tropical diseases. *Drug Discov. Today* 27, 2278–2287. doi: 10.1016/j.drudis.2022.04.004
- Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., et al. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419, 498–511. doi: 10.1038/nature01097
- Ghedini, E., Wang, S., Spiro, D., Caler, E., Zhao, Q., Crabtree, J., et al. (2007). Draft genome of the filarial nematode parasite *Brugia malayi*. *Science* 317, 1756–1760. doi: 10.1126/science.1145406
- Greene, S. E., Fischer, K., Choi, Y. J., Curtis, K. C., Budge, P. J., Mitreva, M., et al. (2022). Characterization of a novel microfilarial antigen for diagnosis of *Wuchereria bancrofti* infections. *PLoS Negl. Trop. Dis.* 16, e0010407. doi: 10.1371/journal.pntd.0010407
- Howick, V. M., Peacock, L., Kay, C., Collett, C., Gibson, W., Lawniczak, M. K. N., et al. (2022). Single-cell transcriptomics reveals expression profiles of *Trypanosoma brucei* sexual stages. *PLoS Pathog.* 18, e1010346. doi: 10.1371/journal.ppat.1010346
- International Helminth Genomes Consortium (2019). Comparative genomics of the major parasitic worms. *Nat. Genet.* 51, 163–174. doi: 10.1038/s41588-018-0262-1
- Ivens, A. C., Peacock, C. S., Worthey, E. A., Murphy, L., Aggarwal, G., Berriman, M., et al. (2005). The genome of the kinetoplastid parasite, *Leishmania major*. *Science* 309, 436–442. doi: 10.1126/science.1112680
- Kim, K., and Weiss, L. M. (2008). Toxoplasma: the next 100 years. *Microbes Infect.* 10, 978–984. doi: 10.1016/j.micinf.2008.07.015
- Logan, J., Pearson, M. S., Manda, S. S., Choi, Y. J., Field, M., Eichenberger, R. M., et al. (2020). Comprehensive analysis of the secreted proteome of adult *Necator americanus* hookworms. *PLoS Negl. Trop. Dis.* 14, e0008237. doi: 10.1371/journal.pntd.0008237
- Magrini, V., Gao, X., Rosa, B. A., McGrath, S., Zhang, X., Hallsworth-Pepin, K., et al. (2018). Improving eukaryotic genome annotation using single molecule mRNA sequencing. *BMC Genomics* 19, 172. doi: 10.1186/s12864-018-4555-7
- McNulty, S. N., Rosa, B. A., Fischer, P. U., Rumsey, J. M., Erdmann-Gilmore, P., Curtis, K. C., et al. (2015). An integrated multiomics approach to identify candidate antigens for serodiagnosis of human onchocerciasis. *Mol. Cell Proteomics* 14, 3224–3233. doi: 10.1074/mcp.M115.051953
- Taylor, C. M., Wang, Q., Rosa, B. A., Huang, S. C., Powell, K., Schedl, T., et al. (2013). Discovery of anthelmintic drug targets and drugs using chokepoints in nematode metabolic pathways. *PLoS Pathog.* 9, e1003505. doi: 10.1371/journal.ppat.1003505
- Tyagi, R., Bulman, C. A., Cho-Ngwa, F., Fischer, C., Marcellino, C., Arkin, M. R., et al. (2021). An integrated approach to identify new anti-filarial leads to treat river blindness, a neglected tropical disease. *Pathogens* 10, 71. doi: 10.3390/pathogens10010071
- Tyagi, R., Maddirala, A. R., Elfawal, M., Fischer, C., Bulman, C. A., Rosa, B. A., et al. (2018). Small molecule inhibitors of metabolic enzymes repurposed as a new class of anthelmintics. *ACS Infect. Dis.* 4, 1130–1145. doi: 10.1021/acscinfed.8b00090
- Wendt, G., Zhao, L., Chen, R., Liu, C., O'Donoghue, A. J., Caffrey, C. R., et al. (2020). single-cell RNA-seq atlas of *Schistosoma mansoni* identifies a key regulator of blood feeding. *Science* 369, 1644–1649. doi: 10.1126/science.abb7709
- Whitman, J. D., Sakanari, J. A., and Mitreva, M. (2021). Areas of metabolomic exploration for helminth infections. *ACS Infect. Dis.* 7, 206–214. doi: 10.1021/acscinfed.0c00658