

VIEWPOINTS

Parasite genomics—Time to think bigger

Carlos Talavera-López[‡], Björn Andersson*

Department of Cell and Molecular Biology, Karolinska Institutet, Berzelius väg 35 SE, Stockholm, Sweden

[‡] Current address: The Francis Crick Institute, London, United Kingdom.

* Bjorn.Andersson@ki.se

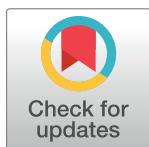
It has been more than 20 years since genomics approaches began to be applied to the problem of tropical neglected parasitic diseases in the form of pilot genome sequencing projects. The *Plasmodium falciparum* genome sequence was published in 2002 [1] as the result of an international collaborative effort, and the first reference genome sequences of *Trypanosoma cruzi*, *Trypanosoma brucei* and *Leishmania major* were published in 2005 [2]. Today, most medically relevant pathogenic protozoan have been the subject of genome sequencing projects.

The initial benefit of many sequencing projects was to boost the field of parasite molecular biology. With most genes of these organisms available for study, a large number of new functional genomics studies with focus on drug resistance, drug target identification and gene evolution could be undertaken in parallel, which has greatly increased activity in these fields [3]. As a result, the knowledge of parasite biology is increasing, and there are many promising approaches for the identification of new candidates for chemotherapeutic agents and vaccines.

The introduction of next generation sequencing technologies has now provided multiple new ways to utilize genomics for the study of parasitic diseases [4] and the integration with other types of large-scale biological data. These include comparative genome sequencing, transcriptomics, proteomics, metabolomics and epigenetics [5] of the parasite, as well as multiple aspects of the host biology. With these approaches, specific biological questions, such as drug resistance, epidemiology, genetic exchange, immune evasion mechanisms, gene function and others can be addressed at a large scale using integrative-omics approaches. In the case of human and many other pathogens such as *Plasmodium* parasites, many such studies have been undertaken and are in progress [6]. For the other neglected diseases, mostly smaller-scale pilot efforts have thus far been carried out [7].

There are several reasons why this is the case and some of these will be discussed below. For certain organisms, progress has been limited by the lack of a complete and reliable reference genome sequence. As a general rule, the greater the repeat content and complexity of the genome, the lower the quality of the initial genome sequence, due to the impossible task of assembling the repetitive regions using short sequencing reads. An example of this is *T. cruzi*, where 38% of the original reference genome sequence, containing thousands of biologically important surface molecule genes, was either missing or misassembled. The use of new long-read sequencing techniques have begun to resolve repetitive regions, resulting in vastly improved genome sequences for other species [8].

There are also other aspects of parasite biology that are relatively unexplored for many parasites and much basic research is needed before pathogenicity issues can be addressed effectively. These include a large proportion of genes where the function is unknown, unclear understanding of the parasite population structure, unresolved issues regarding the extent of genetic exchange, and even missing data on many of the basics of the mechanisms involved in disease causation. While many of these issues are perfect cases for the application of integrative-omics approaches, additional work is also needed for a complete characterisation of



OPEN ACCESS

Citation: Talavera-López C, Andersson B (2017) Parasite genomics—Time to think bigger. PLoS Negl Trop Dis 11(4): e0005463. <https://doi.org/10.1371/journal.pntd.0005463>

Editor: Judd L. Walson, University of Washington, UNITED STATES

Published: April 20, 2017

Copyright: © 2017 Talavera-López, Andersson. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

individual targets and gene families. The extent to which this applies varies depending on the biological complexity of certain parasites, the technological challenges and the resources (or lack thereof) that have been available to researchers in the past.

Sampling is a very important—and often difficult—aspect for the study of many parasitic diseases. While there are many collections available in many cases, these are often not collected in a standardized way, they possess limited metadata and very often the sampling strategy does not have a specific biological question in mind. In order to address a particular biological question, large sample collections are often required to provide enough statistical support and the sampling strategy must be clear with stringent criteria for the characteristics of each sample category, as it has been clearly stated in other fields [9]. Sampling often requires field studies with visits to remote locations that are logistically complex and very costly. In addition, some of the sampled parasites, such as *Leishmania*, *T. cruzi*, *T. rangeli*, among other, cannot be studied without growth in culture and even cloning through limiting dilution. This introduces a level of uncertainty, as parasites may change in culture, and the clones that dominate in culture may not be the dominant ones in the host.

Single-cell genome and transcriptome sequencing has recently been developed [10], and this could potentially be a solution to this problem. Indeed, these studies have helped to expose a more accurate picture of the host-pathogen interactions in simple microbes [11], and they can be expanded to study more complex unicellular protozoan parasites while integrating different types of data [12]. While the cost of carrying out integrative-omics studies is steadily dropping, large scale studies are still expensive from the perspective of the field of neglected diseases, where research grants are usually limited. Notwithstanding, small to medium-scale studies, using dozens to hundreds of samples, can now be carried out using standard-level research grants aided by the massive capacity of modern high-throughput equipment.

Many questions will require the collection of host data, such as the full characterisation of host and/or vector immune responses with high dimensional flow cytometry [13], host microbiome profiling and even host genomic data [14], which will drastically increase the overall cost of the study.

This will need to be solved by a continuing reduction of technical costs, including data production and analysis, and—hopefully—increased funding from private and public agencies for properly designed-omics projects in neglected diseases.

As evidenced by the solutions to long-standing problems mentioned above and the progress made for other organisms, we are now in position to address many biological and medical issues for many neglected parasitic diseases. This advances will translate into applied studies for the development of vaccines and drug targets, which are heavily needed among susceptible populations.

This will require large collaborative efforts involving different areas of expertise, like epidemiology, parasite molecular biology, computational biologists, and big data analyst. Future studies will require the collection of thousands of samples, the production of different-omics data from parasites and hosts, as well as careful validation and follow-up of each finding using *in vitro* and animal model studies.

We can now transition from small pilot studies with unclear results to well-designed large studies with the aim to solve the important, long standing biological and medical questions for these neglected diseases.

References

1. 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: 498–511. <https://doi.org/10.1038/nature01097> PMID: 12368864

2. El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, et al. Comparative genomics of trypanosomatid parasitic protozoa. *Science*. 2005; 309: 404–409. <https://doi.org/10.1126/science.1112181> PMID: 16020724
3. Choi J, El-Sayed NM. Functional genomics of trypanosomatids. *Parasite Immunol*. 2012; 34: 72–79. <https://doi.org/10.1111/j.1365-3024.2011.01347.x> PMID: 22132795
4. Thompson RCA, Lymbery AJ. Let's not forget the thinkers. *Trends Parasitol*. 2013; 29: 581–584. <https://doi.org/10.1016/j.pt.2013.10.006> PMID: 24211214
5. Preidis GA, Hotez PJ. The newest “omics”—metagenomics and metabolomics—enter the battle against the neglected tropical diseases. *PLoS Negl Trop Dis*. 2015; 9: e0003382. <https://doi.org/10.1371/journal.pntd.0003382> PMID: 25675250
6. Kirchner S, Power BJ, Waters AP. Recent advances in malaria genomics and epigenomics. *Genome Med*. 2016; 8: 92. <https://doi.org/10.1186/s13073-016-0343-7> PMID: 27605022
7. McNulty SN, Rosa BA, Fischer PU, Rumsey JM, Erdmann-Gilmore P, Curtis KC, et al. An Integrated Multiomics Approach to Identify Candidate Antigens for Serodiagnosis of Human Onchocerciasis. *Mol Cell Proteomics*. 2015; 14: 3224–3233. <https://doi.org/10.1074/mcp.M115.051953> PMID: 26472727
8. Chien J-T, Pakala SB, Geraldo JA, Lapp SA, Humphrey JC, Barnwell JW, et al. High-Quality Genome Assembly and Annotation for *Plasmodium coatneyi*, Generated Using Single-Molecule Real-Time Pac-Bio Technology. *Genome Announc*. 2016; 4.
9. Gomez-Cabrero D, Abugessaisa I, Maier D, Teschendorff A, Merckenschlager M, Gisel A, et al. Data integration in the era of omics: current and future challenges. *BMC Syst Biol*. 2014; 8 Suppl 2: I1.
10. Linnarsson S, Teichmann SA. Single-cell genomics: coming of age. *Genome Biol*. 2016; 17: 97. <https://doi.org/10.1186/s13059-016-0960-x> PMID: 27160975
11. Avraham R, Haseley N, Brown D, Penaranda C, Jijon HB, Trombetta JJ, et al. Pathogen Cell-to-Cell Variability Drives Heterogeneity in Host Immune Responses. *Cell*. 2015; 162: 1309–1321. <https://doi.org/10.1016/j.cell.2015.08.027> PMID: 26343579
12. Bock C, Farlik M, Sheffield NC. Multi-Omics of Single Cells: Strategies and Applications. *Trends Biotechnol*. 2016; 34: 605–608. <https://doi.org/10.1016/j.tibtech.2016.04.004> PMID: 27212022
13. Levine JH, Simonds EF, Bendall SC, Davis KL, Amir E-AD, Tadmor MD, et al. Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell*. 2015; 162: 184–197. <https://doi.org/10.1016/j.cell.2015.05.047> PMID: 26095251
14. Foth BJ, Tsai IJ, Reid AJ, Bancroft AJ, Nichol S, Tracey A, et al. Whipworm genome and dual-species transcriptome analyses provide molecular insights into an intimate host-parasite interaction. *Nat Genet*. 2014; 46: 693–700. <https://doi.org/10.1038/ng.3010> PMID: 24929830