

## PARASITE GENOMICS

## Opening up a large can of worms

The genomes for essentially all worm parasites of humans are in hand thanks to a new study. Now we have to use them.

Paul W. Sternberg

Helminths—worm parasites of the round, flat or tape variety—have plagued humans for millennia and even now affect one-third of humans. While better civil infrastructure and hygiene could prevent most human helminth infections, this is not accomplished in practice and, regardless, many of our animals would still suffer. For the past 20 years since the completion of the first animal genome (a roundworm, coincidentally), there has been the promise that the parts list, in particular the set of protein-coding genes inferred from a quality genome sequence, would be greatly informative about biology and aid in many research foci. In the past ten years, it has been appreciated that, as far as genomes go, more is better. However, given the sheer number of helminths, it had been hard to make a serious dent in the required number. A recent study<sup>1</sup> remedies this situation: the International Helminth Genomes Consortium reports the sequencing of ~45 genomes and an analysis of over 80. This is no mean feat because many of these parasites are hard for researchers to obtain. What are the implications of this tour de force?

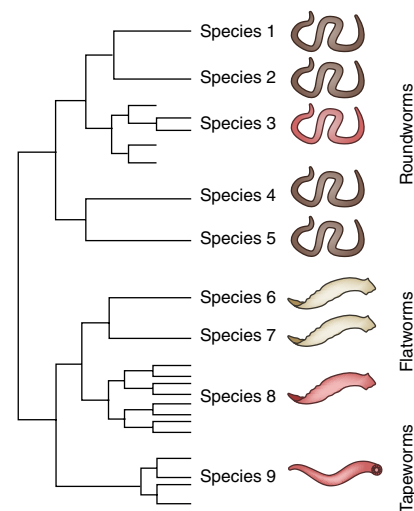
The large phylum of nematodes, estimated to have over a million species, includes parasites of humans, animals and plants as well as free-living marine and terrestrial worms. The free-living worms are found in extreme environments such as goldmines a kilometer underground<sup>2</sup> and cold methane seeps half a kilometer undersea<sup>3</sup>; parasites can be thought of as thriving in extreme environments such as the human intestine. Parasitic nematodes derive from multiple clades, whereas among Platyhelminthes (flatworms) the parasites seem to be in two clades. Tapeworms (cestodes) are at least 6,000 species in number, all parasitic, and there are 10,000 species of flukes (parasitic flatworms or trematodes) in addition to the many free-living flatworms such as the well-known *Planaria*. In spite of the sheer number of worm species, most of the human parasites are now in hand.

## 1.4 million worm genes

Genomes are obtained by massive sequencing, careful assembly and tight quality control, with difficulty caused by repeats and heterozygosity. A high-quality genome assembly comprises most genes—with completeness assessed by those known from other analyses or by orthology—in chromosome-sized pieces. The complement of genes in a genome is inferred from prediction of genes using cDNA sequencing (RNA-seq) as a guide. Each genome, unless closely related to an already known genome (think human–chimpanzee), will have hundreds to thousands of new gene families (a set of orthologs and paralogs). Indeed, in this analysis, among the 1.4 million worm protein-coding genes, the authors found over 100,000 protein families. The function of most proteins is inferred from other related proteins that have been studied somewhere by someone. Half of the protein families had no functional annotation and no obvious protein domains<sup>4</sup> or Gene Ontology annotations<sup>5</sup> associated with them (Fig. 1).

While the function of a single protein from a recognized family could evolve to have a crucial role in parasitism, identifying such proteins is difficult. By contrast, gene family expansions that are specific to parasitic species or clades stand out. Much like vocabulary expansion in human language indicates where a culture chooses to use more nuanced terms, gene family expansions indicate either that the ensuing subfunctionalization is important or that the sheer volume of protein expression is; in either case, it is good heuristic evidence in determining which proteins to examine. By these criteria, the SCP/TAPS family of secreted proteins has long been recognized as potentially having deep involvement in parasitism and association with worms that immunomodulate their hosts<sup>6</sup>. The current study reinforces this view. Other families are expanded in the independently derived parasitic worms, among which ABC transporters and protease inhibitors are prime examples.

Another route to drug discovery is via understanding parasite metabolism:



**Fig. 1 | Gene family expansions shared among independently evolved parasites suggest protein families involved in the parasitic lifestyle.**

Schematic of a protein family tree with expansions in particular species, each of which is a parasite.

the enzymes carrying out metabolic steps that are distinct from the host and that are essential for parasite growth and reproduction become potential drug targets. The large dataset now available allowed an extensive reconstruction of parasitic metabolism. For example, blood-feeding flatworms and nematodes showed expansion of genes in the lactate dehydrogenase pathway, which produces ATP where there is glucose but limited oxygen.

## Not just blood-sucking parasites

One of the goals of this study and for the deeper understanding of helminth parasitism (as well as of the many novel gene families) is to understand the functions of each parasite-specific gene, as well as the functions of common genes, for which it is necessary to know the genomes of a species' closest relatives in order to define the expansions that are likely specific to parasitism rather than due to other causes. Understandably, this impressive study could not be expected to have a few

hundred additional genomes. However, with this current analysis in hand, the most informative genomes for outgroups and ancestors can be reasonably specified. Because genome analysis is continually becoming more facile, we can expect that the requisite outgroups will be analyzed, thereby sharpening up understanding of the evolution of worm parasitism and further filtering the candidate vaccine targets worth pursuing. Functional studies and vaccine development are highly cost intensive, whereas genomes are becoming less costly; thus, it makes sense to continue to analyze helminth genome content by additional genomics to make the functional studies

more efficient. Additional functional studies are certainly needed.

Genome analysis papers describing a single species cannot do justice to the amount of information included, much less a paper describing 45 species from distinct clades. Thus, the student of worms must interrogate the data in the context of an information resource, which for these data is likely to be WormBase's ParaSite<sup>7</sup> (<http://parasite.wormbase.org/>).

**Paul W. Sternberg** 

*Division of Biology and Biological Engineering,  
California Institute of Technology, Pasadena, CA, USA.  
e-mail: pws@caltech.edu*

Published online: 4 December 2018

<https://doi.org/10.1038/s41588-018-0317-3>

#### References

1. International Helminth Genomes Consortium. *Nat. Genet.* <https://doi.org/10.1038/s41588-018-0262-1> (2018).
2. Borgonie, G. et al. *Nature* **474**, 79–82 (2011).
3. Sapir, A. et al. *Front. Microbiol.* **5**, 43 (2014).
4. Mitchell, A. et al. *Nucleic Acids Res.* **43**, D213–D221 (2015).
5. Gene Ontology Consortium. *Nucleic Acids Res.* **45** (D1), D331–D338 (2017).
6. Cantacessi, C. et al. *Biotechnol. Adv.* **27**, 376–388 (2009).
7. Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P. & Berriman, M. *Mol. Biochem. Parasitol.* **215**, 2–10 (2017).

#### Competing interests

The author declares no competing interests.