Protozoan genomics for drug discovery

Kshitiz Chaudhary & David S Roos

Comparing the metabolic pathways of parasites and their hosts facilitates the identification of new drug targets.

As the world grows smaller through increasing economic integration and travel, health is becoming a global issue, particularly in the realm of infectious diseases. Infection with microorganisms now has international repercussions, as seen in the emergence of AIDS, SARS and avian influenza; the spread of antibiostatic-resistant malaria, tuberculosis and nosocomial infections; and bioterrorism concerns. A series of papers in recent issues of *Science* report the complete genome sequences of five important protozoan parasites: two species of *Theileria*, which cause East Coast fever and theileriosis in cattle, and three members of the phylum Kinetoplastida—*Leishmania* major, which leads to skin lesions (and is related to *L. donovani*, which causes visceral leishmaniasis); *Trypanosoma brucei*, which produces sleeping sickness in humans and wasting disease in cattle; and *T. cruzi*, the cause of Chagas disease. Comparative study of the genomes of these and other parasites promises to elucidate the biology of many devastating tropical diseases and help identify parasite-specific targets for drug development.

Parasitic organisms are a major cause of disease worldwide, but because they are concentrated in poorer parts of the world, they receive relatively little attention from the pharmaceutical industry and major scientific funding agencies. Of the ten diseases targeted as research priorities by the World Health Organization's Special Program for Research and Training in Tropical Diseases (http://www.who.int/tdr/), one is caused by a virus (dengue fever), two by bacterial pathogens (leprosy, tuberculosis), three by helminth parasites (filariasis, onchocerciasis, schistosomiasis) and four by protozoan parasites (malaria, leishmaniasis, Chagas disease, African trypanosomiasis). Malaria alone affects over 500 million individuals annually, with 1–3 million deaths. Beyond their direct effects on human health, parasites such as *Theileria* and *Trypanosoma brucei* also exert a significant economic impact as veterinary pathogens of agriculturally important animals.

The five new genome sequences complement previously reported sequences of other apicomplexan parasites, including *Plasmodium falciparum* (malaria), several rodent malaria-causing species, and *Cryptosporidium parvum* and *C. hominis*. The sequences of *Plasmodium vivax* (http://www.tigr.org), *Toxoplasma gondii* (http://ToxoDB.org) and *Eimeria tenella* (http://www.sanger.ac.uk) are approaching completion. The genome of the amitochondrial eukaryote *Entamoeba histolytica* has also been published recently, and sequencing of the *Giardia lambia* (http://www.mbl.edu/Giardia) and *Trichomonas*

---

**Figure 1** Eukaryotic genome projects. Effectively complete genomes are available or in preparation for each of the taxa shown; parentheses indicate availability of genomes for multiple species. Asterisks indicate genome sequences of parasitic organisms; crosses indicate taxa in which some of the available genomes are derived from parasites (in other cases related species are parasitic, e.g., *Naegleria fowleri*, *Tetrahymena pyriformis*). Shaded boxes represent five distinct eukaryotic groups; the precise relationship among these lineages, and where they diverged from prokaryotes, is uncertain. The vast majority of eukaryotic diversity is protozoan, and most of the sequenced protozoa are parasitic. Recent publications report the genomes of kinetoplastida, *Trypanosoma brucei*, *T. cruzi*, *Leishmania*; dark blue) and apicomplexans including *Theileria*, *Plasmodium*, *Toxoplasma* and *Cryptosporidium*; red.
Vaginalis (http://www.tigr.org) genomes is in progress. In aggregate, this work helps fulfill a decade-old proposal to sequence all major parasitic pathogens, providing "the sequence of every virulence determinant, every protein antigen, and every drug target, for all time". But how is this information to be exploited to produce new therapeutics?

The development of drugs, vaccines and diagnostics can be pursued from an evolutionary perspective, taking advantage of comparative genomics to identify and exploit characteristics that distinguish pathogen and host. Contrary to popular opinion, there is no shortage of effective antibiotics, but those that kill the patient are not particularly useful! From this perspective, eukaryotic pathogens pose a more difficult challenge than prokaryotes, as they share many metabolic pathways with humans. For example, macrolides, tetracyclines and many other inhibitors of prokaryotic protein synthesis do not inhibit cytoplasmic protein synthesis in eukaryotic parasites (but see below for a surprising twist on this story), and β-lactam antibiotics—specific inhibitors of bacterial cell-wall synthesis—are not active against eukaryotic pathogens, which lack a cell wall.

Building on available sequence data, phylogenetic analysis has defined five supergroups of eukaryotes (Fig. 1): the plant and red/green algal lineage; a clade comprised of animals, fungi, slime molds and amoebozoans; and three groups that are entirely protozoan. The chromalveolates include the parasitic phyllum apicomplexa (dark red), along with ciliates, diatoms and many taxa for which no complete genomes are available; excavates include the kinetoplastid parasites (dark blue) and other lineages (many anaerobic and/or parasitic). The availability of multiple apicomplexan and kinetoplastid sequences provide a rich resource for examining the evolution and biology of these pathogens and for developing antimicrobial agents based on parasite-specific targets.

Species- or genus-specific genes may help define unique aspects of parasite lifestyle or pathogenicity, such as the striking lymphoproliferation induced by Theileria or the cytoadherence of erythrocytes infected by Plasmodium falciparum—effects that underlie parasite virulence. Among the kinetoplastida, Leishmania is an intracellular parasite that has adapted to survival within the lysosomes of professional phagocytic cells (macrophages), whereas T. brucei, another intracellular parasite, escapes into the cytoplasm. T. brucei lives in the blood, using its spectacular antigenic variability to evade the immune response. More ancient taxon-specific features offer the prospect of developing broad-spectrum antiparasitic agents, such as inhibitors that block endocytosis through the distinctive flagellar pocket of kinetoplastid parasites or that inhibit the establishment or function of the intracellular parasitophorous vacuole within which most apicomplexans reside.

Comparative genomic analysis identifies significant differences between parasites and their host or vector species. Metabolic pathways have traditionally provided attractive targets for drug development and are therefore of particular interest. Many parasites have dispensed with pathways that are essential for free-living organisms. For example, all parasites are purine auxotrophs, and the consequent dependence on salvage pathways underlies the use of allopurinol as a treatment for leishmaniasis (a toxic purine analog). In contrast, most parasites have lost the ability to salvage pyrimidines, relying on de novo pyrimidine biosynthesis—the target of sulfonamides. (Only Cryptosporidium can salvage pyrimidines, as discussed below.) All apicomplexan and kinetoplastid parasites also lack at least some components of the urea cycle, and several (Theileria, Cryptosporidium, trypanosomes) have lost the ability to synthesize heme. Theileria and Toxoplasma are unable to synthesize polyamines, and Cryptosporidium has even eliminated the TCA cycle.

Parasites have also developed several alternative strategies for sterol and lipid synthesis. The kinetoplastida synthesize isopentenyl pyrophosphate using HMG-CoA reductase, but they then convert squalene to ergosterol rather than cholesterol, providing a promising target for chemotherapeutic intervention. Antifungal agents already in the clinic (triazole inhibitors of C14α sterol demethylyase) provide lead compounds against protozoan parasites. Apicomplexan parasites lack the HMG-CoA reductase pathway, but most are capable of synthesizing isoprenoids through the 1-deoxy-D-xylulose 5-phosphate (DOXP) pathway, which is attractive as the target of known herbicides. Plasmodium, Toxoplasma and all of the kinetoplastids synthesize fatty acids using a type II fatty acyl synthase (FAS) complex, the target of triclosan and other antibiotics. In contrast, Cryptosporidium, like humans, employs FAS-I. Surprisingly, the Toxoplasma genome encodes both FAS-I and FAS-II, whereas the Theileria genome has neither.

Striking differences also exist in parasite metabolism of amino acids. As shown in Figure 2, human cells synthesize 10 amino acids (green boxes), via the enzymatic steps indicated as green arrows above. Remarkably, each of these pathways is missing in one or more of the kinetoplastid or apicomplexan parasites (white arrows), resulting in many distinct patterns of auxotrophy. Parasites with the most reduced genomes (Cryptosporidium, Theileria) have lost multiple pathways and presumably compensate by using proteases and permeases evident in their genomes. Another convenient parasitic strategy involves salvage from the host. Plasmodium, for example, eats hemoglobin, and the specialized proteases responsible for this activity provide a possible drug target.

In some cases, gene losses within a single lineage account for the observed distribution of
amino acid auxotrophy, as in the loss of enzymes that convert methionine to cysteine in all of the apicomplexans. In other cases, multiple independent losses are more likely, as in the loss of phenylalanine hydroxylase in the trypanosomes, the Plasmodium-Theileria lineage and Cryptosporidium. A gene encoding this enzyme appears to have been retained in the Leishmania genome, although its activity may be insufficient to meet all of the parasite’s needs.\(^{26}\)

Parasite genomes also reveal metabolic pathways that are lacking in humans\(^{1–5,7–12}\), including pathways for amino acid biosynthesis (Fig. 2, red arrows). The Kinoplastida can synthesize cysteine from serine, and both methionine and threonine from aspartate semialdehyde. Cryptosporidium can metabolize serine to tryptophan, and Toxoplasma harbors the entire diaminopimelate pathway, enabling conversion of aspartate to lysine. In many cases, such parasite-specific pathways bear a striking resemblance to prokaryotic metabolism, and phylogenetic reconstruction suggests their acquisition by horizontal gene transfer. For example, several genes in the glycolytic pathway and TCA cycle of kinetoplastids (including, GAPDH and isocitrate dehydrogenase) exhibit close phylogenetic affinity to bacterial orthologs\(^{4,16}\). As noted above, Cryptosporidium is highly unusual among protozoan parasites in expressing a thymidylate synthase (β subunit), thymidine kinase, IMPDH and several other proteins in Cryptosporidium appear to be derived from a eubacterial ancestor.\(^{27}\)

Such transfers are far more common in protozoa than in ‘higher’ eukaryotes, particularly in intestinal parasites such as Cryptosporidium. It may be that the coexistence of these pathogens with diverse gut flora provides a suitable environment for horizontal transfer. Overall, lateral gene transfer has significantly shaped the metabolic repertoire of protozoan parasites.

Perhaps the most intriguing instance of horizontal transfer is that of the apicomplexan plastid, or “apicoplast,” a secondary endosymbiotic organelle acquired when an ancient chromalveolate engulfed (or was invaded by) a eukaryotic alga and retained the algal plastid as a distinct organelle.\(^{28}\) The apicoplast has lost all photosynthetic function but is nevertheless essential for parasite survival\(^{29}\) (except in Cryptosporidium, which has lost this organelle, but has retained several plant-like enzymes that may derive from the algal endosymbiont).\(^{30}\) Approximately 15% of the P. falciparum genome encodes proteins destined for the apicoplast, including those involved in producing isoprenoids and fatty acids through the DOXP and FAS-II pathways; in addition, the apicoplast has a crucial role in compartmentalizing heme biosynthesis\(^{31}\). Housekeeping enzymes associated with this organelle are particularly attractive as drug targets, as apicoplast DNA replication, RNA transcription and protein synthesis are all prokaryotic in nature, providing targets for drugs such as ciprofloxacin, rifamycin and clindamycin (used clinically to treat toxoplasmosis)\(^{29,31}\).

Genome sequences are being deciphered at an ever-increasing rate, with each new entry providing numerous opportunities for comparison with those already available. Comparative genomics applications are now driving the development of new bioinformatics tools\(^{32}\), making it possible to map entire metabolic pathways in silico and complementing methods for genetic, cell biological, biochemical and pharmacological validation at the bench\(^{33–35}\) and in the clinic. Such advances are bound to expedite the development of new drugs for old pathogens.

---

1. Gardner, M.J. et al. Science 309, 134–137 (2005).
2. Pain, A. et al. Science 309, 131–133 (2005).
3. Ivors, A.C. et al. Science 309, 436–442 (2005).
4. Beriman, M. et al. Science 309, 416–422 (2005).
5. El-Sayed, N.M. et al. Science 309, 409–415 (2005).
6. Snow, R.W. et al. Nature 434, 214–217 (2005).
7. Gardner, M.J. et al. Nature 419, 498–511 (2002).
8. Carlton, J.M. et al. Nature 419, 512–519 (2002).
9. Hall, N. et al. Science 307, 82–86 (2005).
10. Abrahamsen, M.S. et al. Science 304, 441–445 (2004).
11. Xu, P. et al. Nature 431, 1107–1112 (2004).
12. Loftus, B. et al. Nature 433, 865–868 (2005).
13. Bloom, B.R. Nature 378, 236 (1995).
14. Keeling, P.J. et al. Trends Ecol. Evol. (in the press).
15. Roos, D.S. Science 309, 72–73 (2005).
16. El-Sayed, N.M. et al. Science 309, 404–409 (2005).
17. Dobbeleare, D.A. & Kuenzi, P. Curr. Opin. Immunol. 16, 524–530 (2004).
18. Craig, A. & Scherf, A. Molec. Biochem. Parasitol. 115, 129–143 (2001).
19. Borst, P. Cell 109, 5–8 (2002).
20. Overath, P. & Engstler, M. Molec. Microbiol. 53, 735–744 (2004).
21. Gull, K. Curr. Opin. Microbiol. 6, 365–370 (2003).
22. Sibley, L.D. Science 304, 248–253 (2005).
23. Roberts, C.W. et al. Mol. Biochem. Parasitol. 126, 129–142 (2003).
24. Borman, S. et al. J. Infect. Dis. 190, 1534–1540 (2004).
25. Kiemba, M. & Goldberg, D.E. Annu. Rev. Biochem. 71, 275–305 (2002).
26. Steiger, R.F. & Steiger, E. J. Protozool. 40, 437–441 (1977).
27. Striepen, B. et al. Proc. Natl. Acad. Sci. USA 101, 3154–3159 (2004).
28. Roos, D.S. et al. Curr. Opin. Microbiol. 2, 426–432 (1999).
29. Fichera, M.E. & Roos, D.S. Nature 389, 407–409 (1987).
30. Zhu, G. et al. Microbiol. 146, 315–321 (2000).
31. Ralph, S.A. et al. Nature Rev. Microbiol. 2, 203–216 (2004).
32. Li, S. et al. Genome Res. 13, 2178–2190 (2003).
33. Beverley, S.M. Nature Rev. Genet. 4, 11–19 (2003).
34. Roos, D.S. et al. Methods Cell Biol. 45, 27–63 (1994).
35. Carvalho, T.G. & Menard, R. Curr. Issues Molec. Biol. 7, 39–55 (2005).