D706–D713 *Nucleic Acids Research, 2013, Vol. 41, Database issue* *Published online 27 November 2012*
doi:10.1093/nar/gks1139

Library of Apicomplexan Metabolic Pathways: a manually curated database for metabolic pathways of apicomplexan parasites

Achchuthan Shanmugasundram[1,2,*], Faviel F. Gonzalez-Galarza[1], Jonathan M. Wastling[2], Olga Vasieva[1] and Andrew R. Jones[1]

[1]Department of Functional and Comparative Genomics, Institute of Integrative Biology, University of Liverpool, Biosciences Building, Crown Street, Liverpool L69 7ZB and [2]Department of Infection Biology, Institute of Infection and Global Health, University of Liverpool, Liverpool Science Park Innovation Centre 2, 146 Brownlow Hill, Liverpool L3 5RF, UK

Received August 15, 2012; Revised October 9, 2012; Accepted October 24, 2012

ABSTRACT

The Library of Apicomplexan Metabolic Pathways (LAMP, http://www.llamp.net) is a web database that provides near complete mapping from genes to the central metabolic functions for some of the prominent intracellular parasites of the phylum *Apicomplexa*. This phylum includes the causative agents of malaria, toxoplasmosis and theileriosis—diseases with a huge economic and social impact. A number of apicomplexan genomes have been sequenced, but the accurate annotation of gene function remains challenging. We have adopted an approach called metabolic reconstruction, in which genes are systematically assigned to functions within pathways/networks for *Toxoplasma gondii*, *Neospora caninum*, *Cryptosporidium* and *Theileria* species, and *Babesia bovis*. Several functions missing from pathways have been identified, where the corresponding gene for an essential process appears to be absent from the current genome annotation. For each species, LAMP contains interactive diagrams of each pathway, hyperlinked to external resources and annotated with detailed information, including the sources of evidence used. We have also developed a section to highlight the overall metabolic capabilities of each species, such as the ability to synthesize or the dependence on the host for a particular metabolite. We expect this new database will become a valuable resource for fundamental and applied research on the *Apicomplexa*.

INTRODUCTION

The *Apicomplexa* is a large phylum of obligate intracellular eukaryotic protozoa, part of the *Chromalveolata*. They are defined by possessing an apical complex at their anterior end. The *Apicomplexa* originate from a photosynthetic predecessor and some members of the phylum retain a non-photosynthetic plastid called the apicoplast. The different branches of *Apicomplexa* occupy different ecological niches and possess varying metabolic capabilities owing to their early divergence from a common ancestor. It is estimated that *Coccidia* (*Toxoplasma*, *Neospora*, *Eimeria*, *Sarcocystis*, *Cryptosporidium*) separated from *Aconoidasida* (*Plasmodium*, *Theileria* and *Babesia*) ~705 million years ago (Mya) (1,2), although Douzery *et al*. calculated it as 495 Mya (2,3). These variations are because of the sensitivity of different phylogenetic classifications. The former estimation considers *Cryptosporidia* to be a branch of *Coccidia*, whereas Douzery *et al*. considered *Cryptosporidia* to be an independent group that separated from other apicomplexans around 580 Mya (2,3). The divergence time between the closely related apicomplexans *Toxoplasma gondii* and *Neospora caninum* is estimated to be about 28 Mya (4). These parasites present a considerable health and economic burden on society. Malaria, a disease of the tropical world, caused around 216 million cases of fever and 655 000 deaths in 2010 (5) and it is one of the major causes of poverty, especially in sub-Saharan Africa where it costs US $12 billion each year (6). *T. gondii*, which can infect any warm blooded animal, is found in around a third of human populations and can be fatal in immuno-compromised individuals. Toxoplasmosis can

*To whom correspondence should be addressed. Tel: +44 1 51 79 54 555; Fax: +44 1 51 79 54 410; Email: ashanmu@liv.ac.uk

be transmitted *in utero* through the placenta, causing congenital toxoplasmosis in humans and other animals (7). Neosporosis and East Coast Fever are important bovine infections caused, respectively, by *N. caninum* (8) and *Theileria parva* (9) and lead to substantial losses in the dairy and beef industry.

There are around 15 apicomplexan genomes sequenced so far from different branches of the phylum including several *Plasmodium* (10–15) species, *T. gondii* (16), *N. caninum* (4), *Eimeria tenella* (16), *Babesia bovis* (17), *Cryptosporidium* (18–21) and *Theileria* (22,23) species. These genomes and their annotations are available in EuPathDB (24) and GeneDB (25). The annotation of genes to biological functions remains challenging. The sole use of homology-based tools, such as Basic Local Alignment Search Tool (BLAST), can be insufficient and often produces incorrect assignments, especially where gene families exist. It is therefore important to perform careful and detailed functional annotation of these genomes to support both fundamental research and the development of therapeutics and vaccines.

The association of genes to the specific role they play within metabolic networks or structural complexes helps to increase the precision of functional assignments, in a process known as 'metabolic reconstruction' first demonstrated in bacterial genomes such as *Escherichia coli* (26–28). Pinney *et al*. (29) demonstrated the need for a similar approach in the functional annotation of eukaryotic pathogens. Metabolic reconstruction approaches can identify incorrect annotations made by sequence homology-based methods and discover missing components of metabolic pathways.

Metabolic reconstruction can be carried out either automatically or more slowly by manual curation. Automated resources available for the *Apicomplexa* include the Kyoto Encyclopedia of Genes and Genomes (KEGG) (30), ApiCyc (31) and MetaTIGER (32). The Malaria Parasite Metabolic Pathways (MPMP) resource exists for the intra-erythrocyte stage of *Plasmodium falciparum* (33). It is the only manually curated web database available for any apicomplexan and is considered to be a gold standard resource (34). The use of MPMP and other automatically generated resources for the genome-scale metabolic reconstruction and metabolic flux-balance analysis of *P. falciparum* by Plata *et al*. (35) and Huthmacher *et al*. (36) demonstrate the importance of well-curated resources for systems-level analysis and computational evaluation of drug targets. MPMP pathways are reconstructed on the basis of having at least three to four enzymes acting consecutively in a pathway and having been annotated in the genome (33,37). A comparison of *P. falciparum* metabolic pathways from automatic reconstruction resources with MPMP showed that the automated resources are highly error prone (37). The pitfalls of PlasmoCyc (part of ApiCyc) include that it has a larger number of polypeptides than predicted in the official genome release at PlasmoDB, as it was built from an earlier gene model release, more than half of these pathways do not meet the definition of a metabolic pathway and there are large numbers of pathway holes (37). In addition, some of the pathways that were experimentally proven to be absent, such as the mevalonate pathway of isoprene biosynthesis, have been included (37,38). In KEGG, annotated paralogs of enzyme coding genes are missing and many of the pathways are incomplete with fewer genes annotated to functions. The percentage of pathways overlapping in MPMP with KEGG is also <25% (37).

The problems associated with automatic reconstruction resources reported for *P. falciparum* are relevant for other *Apicomplexa* including *T. gondii*. The goal of the Library of Apicomplexan Metabolic Pathways (LAMP) web database is to provide a curated resource for all sequenced *Apicomplexa* of medical and veterinary importance, except for *Plasmodium*, which is covered by MPMP. LAMP contains metabolic pathways constructed in a semi-automatic process, in which information from genome databases and experimental evidence available in the scientific literature has been integrated. In addition, sequence homology searches, protein functional motif identification, information on organelle localization, experimental proteome identification and biochemical/physiological evidence have all been used for metabolic pathway creation (Supplementary File 1 and Figure S1). The curated metabolic pathway annotations and all associated data can be freely accessed without registration by any individual through the http://www.llamp.net portal.

WEBSITE ORGANIZATION AND CONTENT

This website is divided into four main sections (Figure 1C). The introductory section provides a general overview of metabolic reconstruction, the methods used in the analysis and the guidance for understanding metabolic pathways. It also presents a comparative overview of metabolic capabilities present in different apicomplexan species. The second section is *Toxoplasma* and *Neospora*, which presents the pathways for *T. gondii* and *N. caninum* grouped under the metabolic pathway super families of carbohydrate and energy metabolism, amino acids metabolism, lipids and glycan metabolism, nucleotide metabolism, vitamins, cofactors and other substrates metabolism and other organellar pathways. The third and fourth sections are for the metabolic pathways of *Cryptosporidia* (*Cryptosporidium muris*, *Cryptosporidium parvum* and *Cryptosporidium hominis*) and *Piroplasma* (*B. bovis*, *T. parva* and *Theileria annulata*) species and are organized in the same way as *Toxoplasma* and *Neospora*. These organism-specific sections reflect the organization of genomes in different databases, ToxoDB (16), CryptoDB (21) and PiroplasmaDB (24), in EuPathDB.

The pages of each metabolic pathway have four main sections in general. They comprise a literature overview, table of gene annotations, metabolic pathway diagram and a table containing source and fate pathways of metabolites (Figure 1 D–G). The literature overview presents a textual description of the pathway we have written, describing the importance of the metabolic capability and any biochemical/physiological evidence available for the organism or its close relatives. It also includes links to the PubMed pages of relevant publications, thus acting as source of detailed information for users wishing to

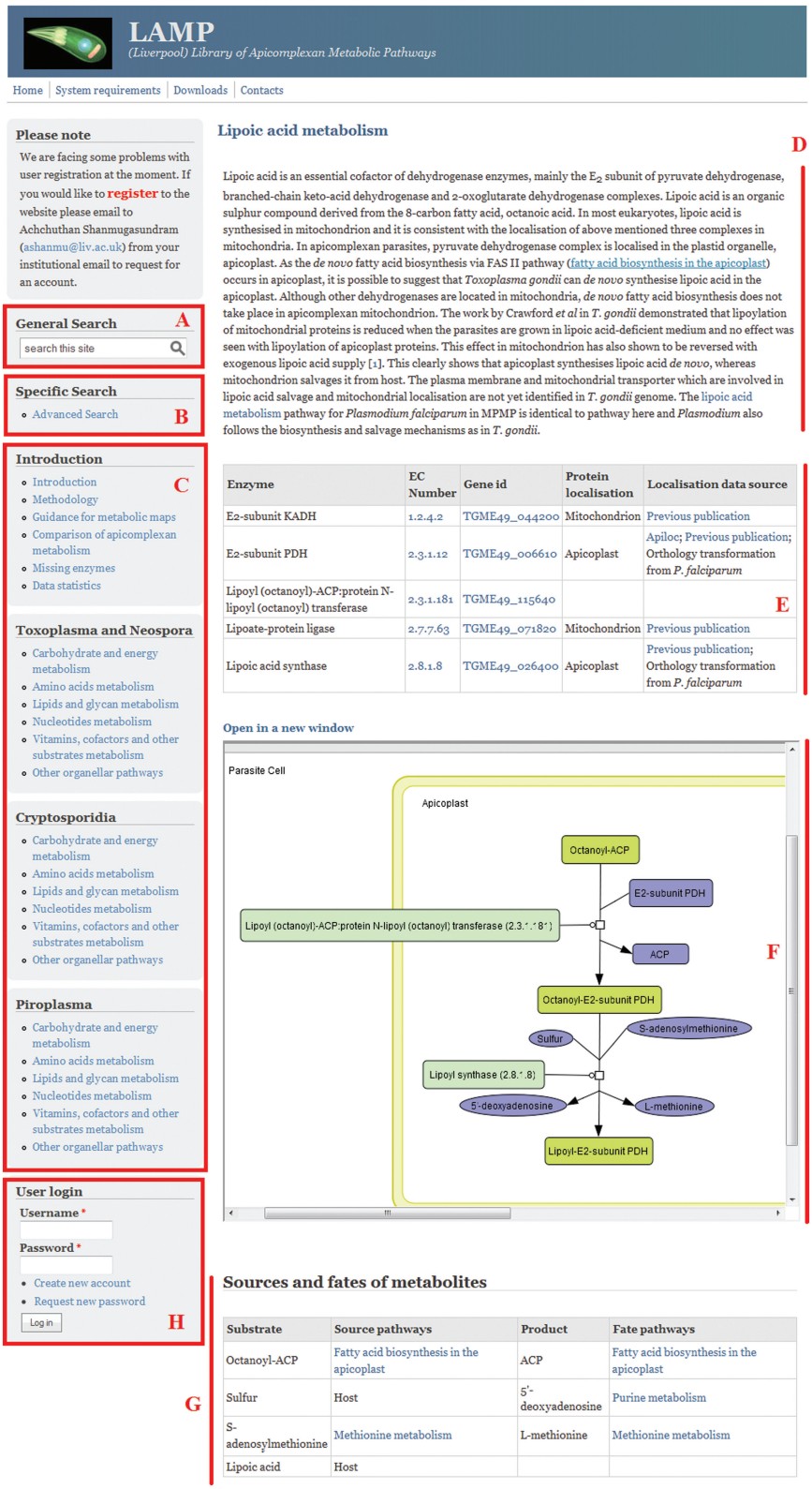

Figure 1. Screen shot of *T. gondii* lipoic acid metabolism pathway page in LAMP. (**A**) The general search box. (**B**) This block shows the specific search option, which allows the search to be carried by choosing one or more of the parameters such as organism, EC number, gene ID and pathway name. (**C**) The blocks showing the four main sections of the website. (**D**) Introductory text of the metabolic pathway page. (**E**) The enzyme annotation table with enzyme names, EC numbers, annotated gene IDs, any available localization evidence and the source of localization evidence. The EC numbers are linked out to ExPASy and the gene IDs are linked out to respective pages in ToxoDB. (**F**) The metabolic pathway diagram is present in a scrollable window with a link to open the diagram in a new window. The apicoplast-based *de novo* biosynthesis branch is visible in the screen shot. (**G**) The table showing the substrates and products of the pathway with their origin and fate pathways. The pathway names have hyperlinks to the respective pathway pages. (**H**) The block for signing up or signing in to the website, required for adding comments and curation and not for accessing pathways or data downloads. 119 × 219 mm (300 × 300 DPI).

gain a summary of past research. The annotation table comprises the list of enzymes/proteins in the pathway with their full or partial Enzyme Commission (EC) numbers available and the gene ids. If any enzyme required for the completion of the pathway is missing, it is noted in the table with 'missing in annotation' in the column for gene ids. The full EC numbers and gene ids are linked to their respective pages in the enzyme database (39) of ExPASy (40) and EuPathDB, respectively. In addition, the organelle localization and the source of the localization (with links to the source) are provided for *T. gondii*, if available. The organelle localization for other species will be added in future releases as and when more experimental data becomes available. One of the main limitations of existing resources such as KEGG and metaTIGER is that they do not provide enzyme names and other metabolites such as cofactors in the metabolic maps. In LAMP, these are taken into account and the cofactors are represented with correct stoichiometry in our pathways. The maps also include enzyme names with EC numbers. Hyperlinks to KEGG compounds are provided for each metabolite to allow users to access the chemical formulas and structures. The enzymes in the pathway are hyperlinked to the KEGG enzyme database, where the full EC number is assigned. When the full EC number is not assigned for a particular enzyme, the links are provided to the KEGG ontology database. The presence of hyperlinks for each reaction to the respective page of KEGG reactions will help distinguish the alternative reactions catalysed by the same enzyme. The pathways' diagrams can be viewed either in the annotation page with the use of scrollable bar or in a new window. Upstream pathway(s) (sources of the pathway precursors) and downstream pathways (ways of utilization of the pathway's products) are shown in an accompanying table. The pathways thus provide a mechanism by which users can learn which metabolites can be produced *de novo* by the parasite and which require uptake from the host. Such information can be helpful in the search for new therapeutic targets.

LAMP can be searched through two different mechanisms. The general search box (Figure 1A) allows the retrieval of data via text mining and therefore the output of the search will provide links to all pages where the searched term is present. The advanced search option (Figure 1B) allows the retrieval of gene annotations by choosing one or more of the parameters from drop boxes (Supplementary Figure S2), querying the underlying relational database of annotations only and not the textual descriptions of pathways.

The core metabolic pathways for the eight species mentioned are available in LAMP. Where possible, it also includes a comparison with MPMP for *P. falciparum*. By design, annotations of proteins to other biological processes such as DNA replication, DNA damage repair, RNA degradation and translation are not present, along with enzymes that do not fit into metabolic pathways such as proteases and peptidases for which more traditional forms of functional annotation may be appropriate. The genes annotated to metabolic pathways in LAMP constitutes around 7–10% of total protein coding genes of the organisms (Table 1). The number of missing enzymes in the annotation for *T. gondii* is low (20) and approximately the same as that for *P. falciparum* in MPMP (19), suggesting that the accuracy of metabolic reconstruction and the underlying gene models for these species are both high.

We have performed a survey to summarize the missing enzymes, which is available on the website (http://www.llamp.net/?q=Missing%20enzymes). Some enzymes are missing in all apicomplexans, including glucosamine-phosphate *N*-acetyltransferase, acyl-Acyl Carrier Protein (ACP) thiolesterase, phosphatidylglycerophosphatase and 3-octaprenyl-4-hydroxybenzoate carboxy-lyase. The enzymes dolichol kinase and dolichylphosphatase are only missing in *Coccidians—T. gondii* and

Table 1. A survey of the data available for the different apicomplexan genomes in LAMP

Organism	Total number of protein coding genes[a]	No. of pathways	No. of unique EC numbers[b]	No. of unique enzymes[c]	No. of total genes in pathways	No of missing enzymes[d]
P. falciparum[e]	5418	42	294	316	520	19
T. gondii ME49	7934	51	386	417	666	20
N. caninum	7080	51	381	411	659	26
C. muris	3934	31	208	224	302	15
C. parvum	3805	28	191	207	270	10
C. hominis	3886	28	184	200	261	17
B. bovis	3706	32	203	216	322	11
T. parva	4082	32	199	213	323	17
T. annulata	3795	32	200	214	313	16

[a]Total numbers of protein coding genes for *Apicomplexa* are obtained from the respective databases of EuPathDB (24).
[b]Unique EC numbers represents the total number of unique enzyme activities with full EC numbers assigned by IUBMB present in the metabolic pathways annotated.
[c]Unique Enzymes represents total unique enzyme activities annotated to be present in the pathways for an organism. This includes enzyme functions with full and partial EC numbers and without EC number annotations.
[d]Missing enzymes represents the enzymes need to be present to complete the metabolic pathways. They may either be missing in the gene model predictions or may be absent in the organism.
[e]The metabolic pathways for *P. falciparum* are not present in LAMP. The numbers for *P. falciparum* do not reflect the total number of genes annotated to all pathways in MPMP, but only those genes annotated to core metabolic functions.

N. caninum. Dolichyl-phosphate-mannose-glycolipid alpha-mannosyltransferase is missing in all *Piroplasma* species and adenosine deaminase is missing in *Theileria* genomes. These missing enzymes could be owing to a novel enzyme or branch yet to be discovered in these pathways. There is also a set of enzymes that is randomly missing in one or two species, suggesting errors in the current gene models. These include inosine monophosphate (IMP) dehydrogenase, asparagine-linked glycosylation (ALG5) and ALG11 in *C. muris*, phosphopantothenoylcysteine decarboxylase in *T. parva* and 6-pyruvoyltetrahydropterin synthase in *N. caninum*. The complete set of metabolic pathway annotations and the metabolic pathway maps of these organisms are available in Microsoft Excel and Systems Biology Markup Language (SBML) files for download at the website (http://www.llamp.net/?q = Downloads). In addition, detailed comparison of metabolic capabilities between these apicomplexans and comparison with human metabolism is also available in the 'downloads' section, which we anticipate should be helpful in identifying putative novel drug targets.

IMPROVEMENTS IN METABOLIC RECONSTRUCTION WITH MANUAL CURATION

The combined use of bioinformatics resources and the wealth of information available in the literature has improved the annotation of individual enzyme functions or led to the validation of overall pathways in LAMP. The majority of pathways present in LAMP are at least partly supported by experimental evidence in either *T. gondii* or *P. falciparum*, and a process of manual curation has been performed on every pathway. Curation tasks typically involve reading the scientific literature, checking for pathway holes, identification of mistaken automated assignments and writing textual summaries for each pathway to give users confidence that the pathway is a true reflection of the known biology of the organism. In the following section, we provide some examples of pathways in which manual curations have led to improvements.

An example of validation of a metabolic pathway with biochemical evidence is lipoic acid metabolism in *T. gondii*. The dissection of the *T. gondii* gene models suggested the presence of enzymes involved in both *de novo* lipoic acid synthesis (lipoyl-ACP:protein *N*-lipoyl transferase and lipoic acid synthase) and salvage (lipoate–protein ligase). The work by Crawford *et al.* (41) in *T. gondii* demonstrated that lipoylation of mitochondrial proteins is reduced when the parasites are grown in lipoic acid–deficient medium and no effect was seen with lipoylation of apicoplast proteins. This effect in mitochondria has also shown to be reversed with exogenous lipoic acid supply. This suggests the presence of apicoplast-based synthesis and mitochondrial salvage pathways.

Aromatic amino acid hydroxylase and phosphatidylserine synthase are examples of enzymes for which annotation of substrate specificity and enzyme function were improved using biochemical evidence. The two genes encoding an aromatic amino acid hydroxylase enzyme were annotated as either phenylalanine 4-hydroxylase (EC number: 1.14.16.1) or tyrosine 3-monooxygenase (EC number: 1.14.16.2) in different resources. However, it has been demonstrated experimentally that the two enzymes can accept both phenylalanine and tyrosine as substrates, although with higher preference for tyrosine (42). In LAMP, we have therefore annotated the two genes with both functions.

A further example of manual curation performed in LAMP involves phosphatidylserine, which can be synthesized either from cytosine diphosphate (CDP)-diacylglycerol or from phosphatidylethanolamine. In MPMP, the *P. falciparum* phosphatidylserine synthase has been annotated as CDP-diacylglycerol dependent (EC number: 2.7.8.8). Conversely, published enzyme assays in *T. gondii* have allowed us to improve the EC number annotation in LAMP, as a study by Gupta *et al.* (43) showed the presence of base-exchange–dependent activity from phosphatidylethanolamine (EC number: 2.7.8.29) rather than CDP-diacylglycerol–dependent activity.

The pathways that have been added without any biochemical evidence and with missing enzymes for *T. gondii* are lysine biosynthesis and threonine biosynthesis pathways; these were added on the basis of the presence of predicted genes for the majority of enzymes and having proteomics evidence for at least one of the enzymes. It remains an open question as to whether *T. gondii* can synthesize lysine and threonine, as we have searched both the gene models and the raw genome sequence for evidence of the missing functional domains required to complete these pathways and, as yet, found no evidence. It remains possible that there are novel enzymatic functions or new pathway branches still to be discovered.

UNDERSTANDING HOST–PARASITE BIOLOGY FROM METABOLIC PATHWAYS

The comparative analysis of metabolic capabilities between different *Apicomplexa* suggests that the variations are attributable to the different ecological niches they occupy and the divergent environmental stresses they undergo. The *Coccidians T. gondii* and *N. caninum* possess greater metabolic capabilities than other *Apicomplexa* species, which is essential for the generalist life style of these *Coccidia* infecting almost any cell type. *P. falciparum* does not have some of the capabilities present in *Toxoplasma*, the main ones being synthesis of some of the amino acids, as they can salvage these from haemoglobin digestion (Figure 2). The absence of genes for lysine and threonine biosynthesis pathways in monoxenic *Coccidia E. tenella* (pathways to be added to LAMP in a later release) may be suggestive of the stress *T. gondii* and *N. caninum* undergo in the tissue cyst (bradyzoite) stage. The presence of starch metabolism (polysaccharide storage) and trehalose synthesis pathways in *Coccidia* and *Cryptosporidia* is indicative of the challenges oocysts face in the external environment. The favourable host environment of *Cryptosporidia* (epithelial cells of the small intestine in *C. parvum* and *C. hominis* and the gastric glands of the stomach in *C. muris*) and the

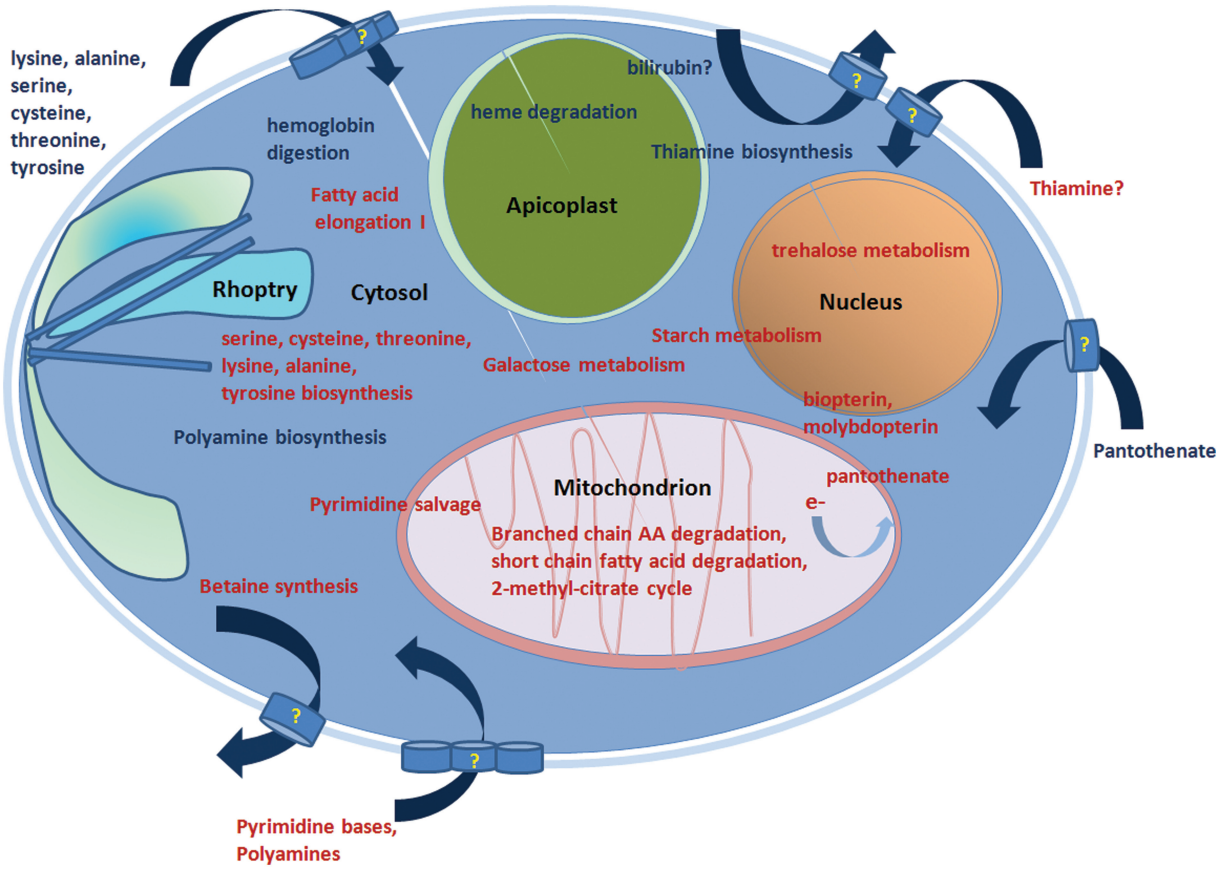

Figure 2. A schematic diagram of *Apicomplexa* illustrating the differences in the metabolic capabilities of *T. gondii* and *P. falciparum*. The metabolic capabilities present in *T. gondii* and absent in *P. falciparum* are shown in red, and the capabilities present in *P. falciparum* and absent in *T. gondii* are in blue. The pathways present in both are not shown. Although *P. falciparum* can synthesize thiamine *de novo*, it also salvages it from host. The pathways are shown in the organelles where they are predicted to occur in the *Apicomplexa*. If part of a pathway is predicted to occur in an organelle and another part in the cytosol, it is shown in midway between the organelle and cytosol. The complete comparison of metabolic capabilities of these species and other species in LAMP is available at http://www.llamp.net/?q = Apicomplexan%20comparison. 53 × 38 mm (300 × 300 DPI).

absence of formation of the parasitophorous vacuole in *Piroplasma* have led to reduced metabolic capabilities in these species including the synthesis of many amino acids, vitamins and cofactors. A striking difference between *C. parvum*/*C. hominis* and *C. muris* is the absence of Tricarboxylic acid (TCA) cycle in the former species, likely due to the increased nutrient availability in the small intestine. An integrated *Apicomplexa* database such as LAMP will help understand these differences and provide a community tool for selecting both organism-specific and general drug targets.

DISCUSSION

By providing an integrated metabolic pathway database for apicomplexan parasites, we aim to provide the medical, veterinary and scientific community with a resource to study the biochemistry of these parasites and to investigate the host–parasite biology. This resource will also be useful for researchers working on flux-balance analysis and metabolomics of these apicomplexan parasites in addition to providing interesting targets for enzyme characterization and drug development. We believe that the inclusion of biochemical evidence and links from genes to the regularly updated EuPathDB gene models will provide users considerable added value beyond those data provided in complementary resources. Links back from EuPathDB gene models to the pathways in LAMP are currently under development and will be available in an up-coming release. The development of LAMP does not only provide a metabolic pathway resource for *Apicomplexa* but also serves as an exemplar for other annotation projects to follow. For example, there are many more parasites of medical and veterinary significance such as Entamoeba, Kinetoplastids and parasitic helminths with genomes currently being sequenced that would be appropriate for following a similar methodology.

The *P. falciparum* resource MPMP is of high quality, with much effort given to manual curation. It is an organism-specific resource, and the usefulness of it to wider *Apicomplexa* community is limited. LAMP provides a balance between a highly curated organism-specific resource and general automatic reconstruction resource such as KEGG. The metabolic reconstruction of *T. gondii* was performed through a hugely detailed manual curation effort and thus we believe pathways and mappings are accurate to a high level. The pathways for *Cryptosporidia* and *Piroplasma* were developed using

the *P. falciparum* and *T. gondii* pathways as a framework, with less manual curation than *T. gondii* because there is less transcriptomics, proteomics and biochemical evidence available for these organisms. However, we believe these pathways are still accurately curated to a level beyond another similar resource based purely on automated reconstruction.

We also encourage the active participation of the apicomplexan research community. Although access to the pathway or the download of files does not require any registration, researchers with specialist knowledge in *Apicomplexa* biochemistry will be able to contribute to the curation by freely registering to the website with an institutional email address. Researchers can also make comments on annotation and metabolic pathways through registration. This database is still under active development and improvements will be applied in response to user feedback. In conclusion, we believe that LAMP acts as an important resource for researchers wishing to understand the metabolic capabilities of this important group of parasites.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary File and Supplementary Figures 1 and 2.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. Hagai Ginsburg for kindly providing his MPMP annotation data.

FUNDING

LAMP is indirectly funded through several grants from Biotechnology and Biological Sciences Research Council. A.R.J. and J.M.W. would like to acknowledge funding from BBSRC for proteomics analysis of the Apicomplexa [BB/G010781/1 to A.R.J. and J.M.W.]. A.S. is funded by a BBSRC DTG studentship awarded to the University of Liverpool. F.G. is funded by BBSRC [BB/I00095X/1 to A.R.J.] Funding for open access charge: Biotechnology and Biological Sciences Research Council UK DTG studentship awarded to the University of Liverpool.

Conflict of interest statement. None declared.

REFERENCES

1. Bhattacharya,D., Yoon,H.S., Hedges,S.B. and Hackett,J.D. (2009) In: Hedges,S.B. and Kumar,S. (eds), *The Timetree of Life*. OUP, New York, pp. 116–120.
2. Hedges,S.B., Dudley,J. and Kumar,S. (2006) TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics*, **22**, 2971–2972.
3. Douzery,E.J.P., Snell,E.A., Bapteste,E., Delsuc,F. and Philippe,H. (2004) The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Proc. Natl Acad. Sci. USA*, **101**, 15386–15391.
4. Reid,A.J., Vermont,S.J., Cotton,J.A., Harris,D., Hill-Cawthorne,G.A., Könen-Waisman,S., Latham,S.M., Mourier,T., Norton,R., Quail,M.A. *et al.* (2012) Comparative genomics of the apicomplexan parasites toxoplasma gondii and neospora caninum: coccidia differing in host range and transmission strategy. *PLoS Pathog.*, **8**, e1002567.
5. WHO. (2011) World Malaria Report 2011.
6. Greenwood,B.M., Bojang,K., Whitty,C.J.M. and Targett,G.A.T. (2005) Malaria. *Lancet*, **365**, 1487–1498.
7. Montoya,J.G. and Liesenfeld,O. (2004) Toxoplasmosis. *Lancet*, **363**, 1965–1976.
8. Trees,A.J., Davison,H.C., Innes,E.A. and Wastling,J.M. (1999) Towards evaluating the economic impact of bovine neosporosis. *Int. J. Parasitol.*, **29**, 1195–1200.
9. Olwoch,J.M., Reyers,B., Engelbrecht,F.A. and Erasmus,B.F.N. (2008) Climate change and the tick-borne disease, Theileriosis (East Coast fever) in sub-Saharan Africa. *J. Arid Environ.*, **72**, 108–120.
10. Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W., Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S. *et al.* (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, **419**, 498–511.
11. Carlton,J.M., Angiuoli,S.V., Suh,B.B., Kooij,T.W., Pertea,M., Silva,J.C., Ermolaeva,M.D., Allen,J.E., Selengut,J.D., Koo,H.L. *et al.* (2002) Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii* yoelii. *Nature*, **419**, 512–519.
12. Carlton,J., Silva,J. and Hall,N. (2005) The genome of model malaria parasites, and comparative genomics. *Curr. Issues Mol. Biol.*, **7**, 23–37.
13. Carlton,J.M., Adams,J.H., Silva,J.C., Bidwell,S.L., Lorenzi,H., Caler,E., Crabtree,J., Angiuoli,S.V., Merino,E.F., Amedeo,P. *et al.* (2008) Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature*, **455**, 757–763.
14. Pain,A., Bohme,U., Berry,A.E., Mungall,K., Finn,R.D., Jackson,A.P., Mourier,T., Mistry,J., Pasini,E.M., Aslett,M.A. *et al.* (2008) The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature*, **455**, 799–803.
15. Aurrecoechea,C., Brestelli,J., Brunk,B.P., Dommer,J., Fischer,S., Gajria,B., Gao,X., Gingle,A., Grant,G., Harb,O.S. *et al.* (2009) PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res.*, **37**, D539–D543.
16. Gajria,B., Bahl,A., Brestelli,J., Dommer,J., Fischer,S., Gao,X., Heiges,M., Iodice,J., Kissinger,J.C., Mackey,A.J. *et al.* (2008) ToxoDB: an integrated *Toxoplasma gondii* database resource. *Nucleic Acids Res.*, **36**, D553–D556.
17. Brayton,K.A., Lau,A.O.T., Herndon,D.R., Hannick,L., Kappmeyer,L.S., Berens,S.J., Bidwell,S.L., Brown,W.C., Crabtree,J., Fadrosh,D. *et al.* (2007) Genome sequence of babesia bovis and comparative analysis of apicomplexan hemoprotozoa. *PLoS Pathog.*, **3**, e148.
18. Bankier,A.T., Spriggs,H.F., Fartmann,B., Konfortov,B.A., Madera,M., Vogel,C., Teichmann,S.A., Ivens,A. and Dear,P.H. (2003) Integrated mapping, chromosomal sequencing and sequence analysis of *Cryptosporidium parvum*. *Genome Res.*, **13**, 1787–1799.
19. Abrahamsen,M.S., Templeton,T.J., Enomoto,S., Abrahante,J.E., Zhu,G., Lancto,C.A., Deng,M., Liu,C., Widmer,G., Tzipori,S. *et al.* (2004) Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science*, **304**, 441–445.
20. Xu,P., Widmer,G., Wang,Y., Ozaki,L.S., Alves,J.M., Serrano,M.G., Puiu,D., Manque,P., Akiyoshi,D., Mackey,A.J. *et al.* (2004) The genome of *Cryptosporidium hominis*. *Nature*, **431**, 1107–1112.
21. Heiges,M., Wang,H., Robinson,E., Aurrecoechea,C., Gao,X., Kaluskar,N., Rhodes,P., Wang,S., He,C., Su,Y. *et al.* (2006) CryptoDB: a *Cryptosporidium bioinformatics* resource update. *Nucleic Acids Res.*, **34**, D419–D422.
22. Gardner,M.J., Bishop,R., Shah,T., de Villiers,E.P., Carlton,J.M., Hall,N., Ren,Q., Paulsen,I.T., Pain,A., Berriman,M. *et al.* (2005) Genome sequence of theileria parva, a bovine pathogen that transforms lymphocytes. *Science*, **309**, 134–137.
23. Pain,A., Renauld,H., Berriman,M., Murphy,L., Yeats,C.A., Weir,W., Kerhornou,A., Aslett,M., Bishop,R., Bouchier,C. *et al.* (2005) Genome of the host-cell transforming parasite *Theileria annulata* cmpared with *T. parva*. *Science*, **309**, 131–133.

24. Aurrecoechea,C., Brestelli,J., Brunk,B.P., Fischer,S., Gajria,B., Gao,X., Gingle,A., Grant,G., Harb,O.S., Heiges,M. *et al.* (2010) EuPathDB: a portal to eukaryotic pathogen databases. *Nucleic Acids Res.*, **38**, D415–D419.
25. Logan-Klumpler,F.J., De Silva,N., Boehme,U., Rogers,M.B., Velarde,G., McQuillan,J.A., Carver,T., Aslett,M., Olsen,C., Subramanian,S. *et al.* (2012) GeneDB—an annotation database for pathogens. *Nucleic Acids Res.*, **40**, D98–D108.
26. Karp,P.D. and Paley,S. (1996) Integrated access to metabolic and genomic data. *J. Comput. Biol.*, **3**, 191–212.
27. Karp,P.D., Krummenacker,M., Paley,S. and Wagg,J. (1999) Integrated pathway–genome databases and their role in drug discovery. *Trends Biotechnol.*, **17**, 275–281.
28. Keseler,I.M., Collado-Vides,J., Santos-Zavaleta,A., Peralta-Gil,M., Gama-Castro,S., Muñiz-Rascado,L., Bonavides-Martinez,C., Paley,S., Krummenacker,M., Altman,T. *et al.* (2011) EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nucleic Acids Res.*, **39**, D583–D590.
29. Pinney,J.W., Papp,B., Hyland,C., Wambua,L., Westhead,D.R. and McConkey,G.A. (2007) Metabolic reconstruction and analysis for parasite genomes. *Trends Parasitol.*, **23**, 548–554.
30. Kanehisa,M., Goto,S., Sato,Y., Furumichi,M. and Tanabe,M. (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.*, **40**, D109–D114.
31. Caspi,R., Altman,T., Dale,J.M., Dreher,K., Fulcher,C.A., Gilham,F., Kaipa,P., Karthikeyan,A.S., Kothari,A., Krummenacker,M. *et al.* (2010) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.*, **38**, D473–D479.
32. Whitaker,J.W., Letunic,I., McConkey,G.A. and Westhead,D.R. (2009) metaTIGER: a metabolic evolution resource. *Nucleic Acids Res.*, **37**, D531–D538.
33. Ginsburg,H. (2006) Progress in in silico functional genomics: the malaria Metabolic Pathways database. *Trends Parasitol.*, **22**, 238–240.
34. Hung,S.S. and Parkinson,J. (2011) Post-genomics resources and tools for studying apicomplexan metabolism. *Trends Parasitol.*, **27**, 131–140.
35. Plata,G., Hsiao,T.-L., Olszewski,K.L., Llinas,M. and Vitkup,D. (2010) Reconstruction and flux-balance analysis of the *Plasmodium falciparum* metabolic network. *Mol. Syst. Biol.*, **6**, 408.
36. Huthmacher,C., Hoppe,A., Bulik,S. and Holzhutter,H.-G. (2010) Antimalarial drug targets in *Plasmodium falciparum* predicted by stage-specific metabolic network analysis. *BMC Syst. Biol.*, **4**, 120.
37. Ginsburg,H. (2009) Caveat emptor: limitations of the automated reconstruction of metabolic pathways in Plasmodium. *Trends Parasitol.*, **25**, 37–43.
38. Jomaa,H., Wiesner,J., Sanderbrand,S., Altincicek,B., Weidemeyer,C., Hintz,M., Türbachova,I., Eberl,M., Zeidler,J., Lichtenthaler,H.K. *et al.* (1999) Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science*, **285**, 1573–1576.
39. Bairoch,A. (2000) The ENZYME database in 2000. *Nucleic Acids Res*, **28**, 304–305.
40. Artimo,P., Jonnalagedda,M., Arnold,K., Baratin,D., Csardi,G., de Castro,E., Duvaud,S., Flegel,V., Fortier,A., Gasteiger,E. *et al.* (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.*, **40**, W597–W603.
41. Crawford,M.J., Thomsen-Zieger,N., Ray,M., Schachtner,J., Roos,D.S. and Seeber,F. (2006) Toxoplasma gondii scavenges host-derived lipoic acid despite its de novo synthesis in the apicoplast. *EMBO J.*, **25**, 3214–3222.
42. Gaskell,E.A., Smith,J.E., Pinney,J.W., Westhead,D.R. and McConkey,G.A. (2009) A unique dual activity amino acid hydroxylase in *Toxoplasma gondii*. *PLoS One*, **4**, e4801.
43. Gupta,N., Zahn,M.M., Coppens,I., Joiner,K.A. and Voelker,D.R. (2005) Selective disruption of phosphatidylcholine metabolism of the intracellular parasite *Toxoplasma gondii* arrests its growth. *J. Biol. Chem.*, **280**, 16345–16353.