RESEARCH NOTE

Conservation of gene essentiality in Apicomplexa and its application for prioritization of anti-malarial drug targets [version 1; peer review: 2 approved with reservations]

Gajinder Pal Singh
Molecular Medicine Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi, India

Abstract

New anti-malarial drugs are needed to address the challenge of artemisinin resistance and to achieve malaria elimination and eradication. Target-based screening of inhibitors is a major approach for drug discovery, but its application to malaria has been limited by the availability of few validated drug targets in Plasmodium. Here we utilize the recently available large-scale gene essentiality data in Plasmodium berghei and a related apicomplexan pathogen, Toxoplasma gondii, to identify potential anti-malarial drug targets. We find significant conservation of gene essentiality in the two apicomplexan parasites. The conservation of essentiality could be used to prioritize enzymes that are essential across the two parasites and show no or low sequence similarity to human proteins. Novel essential genes in Plasmodium could be predicted based on their essentiality in T. gondii. Essential genes in Plasmodium showed higher expression, evolutionary conservation and association with specific functional classes. We expect that the availability of a large number of novel potential drug targets would significantly accelerate anti-malarial drug discovery.

Keywords

Plasmodium falciparum, Toxoplasma gondii, drug targets, essential genes

Corresponding author: Gajinder Pal Singh (gajinder.pal.singh@gmail.com)

Competing interests: No competing interests were disclosed.

Grant information: The work is supported by an Early Career Fellowship to G.P.S. by the Wellcome Trust/DBT India Alliance (IA/E/15/1/502297). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2017 Singh GP. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

How to cite this article: Singh GP. Conservation of gene essentiality in Apicomplexa and its application for prioritization of anti-malarial drug targets [version 1; peer review: 2 approved with reservations] F1000Research 2017, 6:23 (https://doi.org/10.12688/f1000research.10559.1)

First published: 09 Jan 2017, 6:23 (https://doi.org/10.12688/f1000research.10559.1)
Introduction
Malaria killed an estimated half a million people in the year 2015, 70% of them were children under the age of five. The emergence and spread of Plasmodium falciparum strains resistant to all currently used anti-malarial drugs has created an urgent need to discover new drugs. New anti-malarial drugs are also needed for malaria elimination and global eradication, for which the currently available drugs are not adequate. There are two main approaches for drug-discovery against pathogens: Phenotype screening and target-based approach. In phenotype screening, compounds are identified that inhibit the cellular growth of the pathogen. Large-scale screening of millions of compounds against the erythrocytic stage of P. falciparum has identified thousands of such inhibitors. Some of these inhibitors have progressed to clinical trials. In the target-based approach, compounds are identified that inhibit the activity of a protein essential for the viability of the pathogen. Thus target-based approach requires previous knowledge about genes that are essential for the pathogen. Only a few essential genes have been identified in P. falciparum, hampering the target-based approach for anti-malarial drug discovery. Consequently, target-based approach has only identified a few anti-malarial candidates. However, recent large-scale screening of about 2500 genes in a rodent malaria parasite P. berghei has identified about 1200 essential genes. A recent genome-scale CRISPR screen in a related apicomplexan parasite Toxoplasma gondii has identified about 3000 essential genes. Here we analyse this data and find significant conservation of gene essentiality in these two pathogens. From this, we identified potential anti-malarial drug targets that exhibit conserved essentiality in apicomplexan parasites; we predict novel essential genes in Plasmodium based on the essentiality of their orthologs in T. gondii. These targets could serve as starting points for target-based anti-malarial drug discovery.

Methods
Fitness data for knockout mutants
The genome-wide CRISPR screening data on the relative fitness of T. gondii in vivo was generated in human fibroblasts cells obtained from Sidik et al.1. The authors defined log, fold change in abundance of single guide RNA (sgRNA) targeting a given gene as the “phenotype” score for that gene. It was found that for a previously determined set of 81 essential and non-essential genes, a phenotype score of less than -2 identified most of the essential genes, but none of the non-essential genes. We thus defined all genes with a phenotype score of less than -2 as essential (2870 genes). Genes with a phenotype score greater than 0 were defined as non-essential (3071 genes), while those with a phenotype score between 0 and -2 were not classified (2210 genes). The in vivo relative growth rate data for 2574 genes of P. berghei were obtained from PlasmoGEM database.8 (http://plasmojem.sanger.ac.uk/phenotypes). The authors generated knockout mutants by transfection with large pools of barcoded gene knockout vectors. The in vivo growth rate in Balb/c mice was obtained by counting barcodes by next generation sequencing daily between days 4 and 8 post transfection. Essential genes were defined as genes with a growth rate not significantly different from 0.1 (growth rate of the wild type taken as 1), while non-essential genes were defined as genes with growth rate not significantly different from 1.

Proteome data and sequence analyses
Proteome sequences of P. falciparum 3D7, P. berghei ANKA, P. chabaudi chabaudi, P. cynomolgi B, P. knowlesi H, P. reichenowi CDC, P. vivax Sall, P. yoelii 17X were downloaded from the PlasmoDB database 10 (http://plasmodb.org/common/downloads/release-27). The Proteome sequences for six apicomplexan species were obtained from EuPathDB11: Cryptosporidium hominis TUS02 (http://cryptodb.org/common/downloads/release-29/ChominiisTU5022); T. gondii GT1 (http://toxodb.org/common/downloads/release-29/TgondiiGT1); Eimeria brunetti Houghton (http://toxodb.org/common/downloads/release-29/EbrunettiHoughton); Babesia bovis T2Bo (http://piroplasmadb.org/common/downloads/release-29/BbovisT2Bo); Theileria annulata Ankara (http://piroplasmadb.org/common/downloads/release-29/TannulataAnkara); and Gregarina niphandrodes (http://cryptodb.org/common/downloads/release-29/GniphandrodesUnknown/). Proteome sequences for Homo sapiens were downloaded from EBI (http://www.ebi.ac.uk/reference_proteomes). Homologs of P. berghei genes in H. sapiens were identified with E-value cut-off of 1e-6, with soft mask set as true. Orthologous sequences were identified using best bidirectional hit algorithm12.

Functional data
RNA-seq data (FPKM values) for different stages of P. berghei was obtained from Otto et al.13. Proteomics data on different stages of P. berghei and dN, dNS values were obtained from Hall et al.14. Gene Ontology information for P. falciparum was obtained from PlasmoDB15, and these functions were assigned to their orthologous proteins in P. berghei. Enzyme Commission (EC) numbers for P. berghei and P. falciparum were also obtained from PlasmoDB. Trans-membrane regions were identified using TMHMM16. All statistical analyses were performed in the R software version 3.3.1 (https://www.r-project.org/).

Results
Conservation of gene essentiality in apicomplexan parasites
The relative in vivo growth rate of knockout mutants for 2574 P. berghei genes (out of total 5076 genes in P. berghei) has recently been measured, of which 1198 genes (46%) with very low growth rate were classified as essential17,18. Similarly, in vivo relative fitness of knockout mutants for 8151 T. gondii genes have been measured, of which 2870 genes (35%) with very low relative fitness values were classified as essential (see Methods). Of the 2574 P. berghei genes with fitness data, 1617 genes have an ortholog in T. gondii. P. berghei genes with an ortholog in T. gondii were significantly more likely to be essential, compared to P. berghei genes without an ortholog in T. gondii (53% vs. 36%; Fisher test p = 7e-18; Figure 1A). P. berghei genes with an essential ortholog in T. gondii were significantly more likely to be essential, compared to P. berghei genes with a non-essential ortholog in T. gondii (71% vs. 17%; Fisher test p = 6e-59; Figure 1A). There was a significant correlation in relative fitness values of P. berghei and T. gondii (Spearman correlation coefficient 0.47; p = 3e-89; n =1617; Figure 1B). The essentiality of 2502 P. berghei genes was not tested, but the essentiality information of T. gondii orthologs may be used to predict their essentiality in P. berghei. There were 687 genes in P. berghei with an essential ortholog in T. gondii, and thus may be predicted as essential in P. berghei (Dataset 1).
Prioritization of anti-malarial drug targets

We argue that genes identified as essential in both the apicomplexan parasites could be more useful drug targets for the following reasons: 1) Genome-scale fitness screens often involve significant false positives and false negatives, thus genes identified as essential in independent experiments in different parasites could be more confidently assigned as essential; 2) the substantial conservation of gene essentiality between the two parasites demonstrates that essentiality information in *T. gondii* offers relevant information about gene essentiality in *P. berghei*; 3) genes that are essential in both *P. berghei* and *T. gondii* should be more likely to be essential across different developmental stages of *Plasmodium*, which is a highly desirable property of *Plasmodium* drug targets. We thus identified 710 genes that were essential in both species. A total of 289 of these 710 genes encode enzymes, which are typically used as drug targets against pathogens. Of these 289 genes, 245 had an ortholog in all *Plasmodium* species and did not have more than one trans-membrane segment. We removed proteins with more than one trans-membrane segments, as these are often difficult to purify for *in vitro* assays. Of the 245 proteins, 30 showed no significant sequence similarity to any human proteins (listed in Table 1), and 83 showed less than 30% identity and 151 showed less than 40% identity to any human protein (Dataset 1). Figure 2 shows the flow chart of the selection process.

Among the *P. berghei* enzymes that were not tested for essentiality, 186 had an essential ortholog in *T. gondii* and thus may be predicted as essential in *P. berghei*. To increase the confidence of these genes to be essential in *Plasmodium*, we considered 53 genes that were conserved across *Plasmodium* and apicomplexan species. Among the enzymes tested for essentiality, such a criteria led to a set with 77% enzymes as essential, suggesting high enrichment for essentiality among predicted essential enzymes. In total, 28 of these enzymes had low sequence similarity (<40% identity) with human proteins and thus may also be considered as potential drug targets (Dataset 1).

Properties of essential *P. berghei* genes

Essential genes show different expression, evolutionary and functional properties. We thus tested whether similar patterns would be observed for *P. berghei*. Essential *P. berghei* genes showed higher mRNA expression levels in asexual stages, but lower expression levels in sexual stages compared to non-essential genes (Figure 3A). Proteins encoded by essential genes were more likely to be detected by mass-spectrometry in different developmental stages compared to non-essential genes (Figure 3B).
Essential genes showed a lower evolutionary rate (dN and dN/dS) and higher conservation in apicomplexan species (Figure 3C). Essential genes were significantly enriched in functional classes, such as “Translation”, “Ribosome”, “DNA replication”, “Intracellular protein transport”, “Cytoplasm”, and “Nucleus” (Figure 4).

**Table 1. Essential Plasmodium enzymes with no significant similarity to human proteins.**

<table>
<thead>
<tr>
<th>P. berghei ID</th>
<th>P. falciparum ID</th>
<th>Gene name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBANKA_0306300</td>
<td>PF3D7_0209200</td>
<td>3' exoribonuclease</td>
<td></td>
</tr>
<tr>
<td>PBANKA_0507000</td>
<td>PF3D7_1022800</td>
<td>GcpE</td>
<td>4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase</td>
</tr>
<tr>
<td>PBANKA_0615800</td>
<td>PF3D7_0718100</td>
<td>EST</td>
<td>exported serine/threonine protein kinase</td>
</tr>
<tr>
<td>PBANKA_0802700</td>
<td>PF3D7_0705000</td>
<td>mRNA cap guanine-N7 methyltransferase</td>
<td></td>
</tr>
<tr>
<td>PBANKA_0823300</td>
<td>PF3D7_0922400</td>
<td>pBAS</td>
<td>para-aminobenzoic acid synthetase</td>
</tr>
<tr>
<td>PBANKA_0828600</td>
<td>PF3D7_0927800</td>
<td>COX5B</td>
<td>cytochrome c oxidase subunit 5B</td>
</tr>
<tr>
<td>PBANKA_0828800</td>
<td>PF3D7_0928000</td>
<td>COX6B</td>
<td>cytochrome c oxidase subunit 6B</td>
</tr>
<tr>
<td>PBANKA_0927800</td>
<td>PF3D7_1120500</td>
<td>iRNA nucleotidyltransferase</td>
<td></td>
</tr>
<tr>
<td>PBANKA_1006500</td>
<td>PF3D7_0408900</td>
<td>KAE1</td>
<td>iRNA N6-adenosine theoneylcarbamoyltransferase</td>
</tr>
<tr>
<td>PBANKA_1017700</td>
<td>PF3D7_1426900</td>
<td>QCR6</td>
<td>cytochrome b-c1 complex subunit 6</td>
</tr>
<tr>
<td>PBANKA_1033400</td>
<td>PF3D7_1409100</td>
<td>aldo-keto reductase</td>
<td></td>
</tr>
<tr>
<td>PBANKA_1035300</td>
<td>PF3D7_1406900</td>
<td>radical SAM protein</td>
<td></td>
</tr>
<tr>
<td>PBANKA_1104700</td>
<td>PF3D7_0505100</td>
<td>TRS85</td>
<td>trafficking protein particle complex subunit 8</td>
</tr>
<tr>
<td>PBANKA_1121500</td>
<td>PF3D7_0622600</td>
<td>QCR9</td>
<td>cytochrome b-c1 complex subunit 9</td>
</tr>
<tr>
<td>PBANKA_1122100</td>
<td>PF3D7_0623200</td>
<td>FNR</td>
<td>ferredoxin--NADP reductase</td>
</tr>
<tr>
<td>PBANKA_1138500</td>
<td>PF3D7_1362500</td>
<td>exonuclease</td>
<td></td>
</tr>
<tr>
<td>PBANKA_1139700</td>
<td>PF3D7_1363700</td>
<td>conserved Plasmodium protein, unknown function</td>
<td></td>
</tr>
<tr>
<td>PBANKA_1210700</td>
<td>PF3D7_1012300</td>
<td>QCR7</td>
<td>cytochrome b-c1 complex subunit 7</td>
</tr>
<tr>
<td>PBANKA_1228500</td>
<td>PF3D7_0801700</td>
<td>SENP2</td>
<td>sentrin-specific protease 2</td>
</tr>
<tr>
<td>PBANKA_1304200</td>
<td>PF3D7_1440200</td>
<td>SPP</td>
<td>stromal-processing peptidase</td>
</tr>
<tr>
<td>PBANKA_1310600</td>
<td>PF3D7_1446800</td>
<td>HDP</td>
<td>heme detoxification protein</td>
</tr>
<tr>
<td>PBANKA_1322800</td>
<td>PF3D7_1459100</td>
<td>GTP-binding protein</td>
<td></td>
</tr>
<tr>
<td>PBANKA_1330600</td>
<td>PF3D7_1467300</td>
<td>DXR</td>
<td>1-deoxy-D-xylose 5-phosphate reductoisomerase</td>
</tr>
<tr>
<td>PBANKA_1338400</td>
<td>PF3D7_1323200</td>
<td>V-type proton ATPase subunit G</td>
<td></td>
</tr>
<tr>
<td>PBANKA_1406100</td>
<td>PF3D7_1307600</td>
<td>DNA-directed RNA polymerase alpha chain</td>
<td></td>
</tr>
<tr>
<td>PBANKA_1409500</td>
<td>PF3D7_1311000</td>
<td>ISD11</td>
<td>protein ISD11</td>
</tr>
<tr>
<td>PBANKA_1418400</td>
<td>PF3D7_1320100</td>
<td>CtpS</td>
<td>ATP-dependent Ctp protease adapter protein CtpS</td>
</tr>
<tr>
<td>PBANKA_1426700</td>
<td>PF3D7_0810800</td>
<td>PPPK-DHPS</td>
<td>hydroxymethylidihydropterin pyrophosphokinase-dihydropteroate synthase</td>
</tr>
<tr>
<td>PBANKA_1442600</td>
<td>PF3D7_1227900</td>
<td>RNA pseudouridylate synthase</td>
<td></td>
</tr>
<tr>
<td>PBANKA_1443200</td>
<td>PF3D7_1228500</td>
<td>RNA pseudouridylate synthase</td>
<td></td>
</tr>
</tbody>
</table>

**Discussion**

The recent availability of gene essentiality data from *P. berghei* and the related apicomplexan *T. gondii* provides an unprecedented opportunity to identify potential drug targets to accelerate anti-malarial drug discovery. We find a significant correlation of gene essentiality between *P. berghei* and *T. gondii* (Figure 1). Thus, the information about gene essentiality in *T. gondii* provides independent experimental support for gene essentiality in *P. berghei*, which not only increases the confidence of gene essentiality in *P. berghei*, but also increases the likelihood that these genes would be essential in other *Plasmodium* species that cause human malaria, and probably in different *Plasmodium* developmental stages. Drug targets
Of the 2574 tested, 1198 were essential in *P. berghei*. Essential in *T. gondii*:
- 710 genes
- Enzymes:
  - 289 genes
- Orthologs in all *Plasmodium* species, Trans-membrane regions < 2:
  - 245 genes

30 genes with no significant similarity to *H. sapiens*

83 genes with *H. sapiens* homolog < 30% identity

151 genes with *H. sapiens* homolog < 40% identity

**Figure 2.** Selection of potential drug targets in *Plasmodium*.

---

**Figure 3.** Properties of essential *Plasmodium berghei* genes. (A) Essential *P. berghei* genes showed higher mRNA expression levels in asexual stages, but lower mRNA expression levels in sexual stages. The mean FPKM values for the essential and non-essential genes were calculated for different development stages and their log₂ ratio was taken. All stages except ‘ookinete 24h’ showed a statistically significant difference between essential and non-essential genes (t-test; p < 0.05). The RNA-seq data was taken from Otto *et al.*[13]. (B) Proteins encoded by essential genes were more likely to be detected by mass-spectrometry in different stages compared to non-essential genes. All stages except ‘sporozoites’ showed a significant difference between essential and non-essential genes (Chi-square test; p < 0.05). Overall 47% of the tested genes were essential. The proteomics data was obtained from Hall *et al.*[14]. (C) Essential genes showed a lower evolutionary rate and higher conservation across apicomplexan species. The mean dN and dN/dS values for essential and non-essential genes was calculated and their log₂ ratio was taken. The mean number of apicomplexan species (out of six), in which an ortholog was identified, was calculated for essential and non-essential genes and their log₂ ratio was taken. dN and conservation in apicomplexan species showed a statistically significant difference between essential and non-essential genes (t-test; p < 0.05), but not dN/dS.
that are essential in multiple species and stages of Plasmodium are particularly desirable. Novel essential genes in Plasmodium could also be predicted based on the essentiality of their orthologs in T. gondii. Further prioritization of these genes could be made based on their conservation across Plasmodium and apicomplexan species, low sequence similarity to human proteins, as well as practical information, such as previous availability of clones, assays, protein structure and inhibitors. The high conservation of essentiality between P. berghei and T. gondii may allow prediction of essential genes in other apicomplexan pathogens, such as Cryptosporidium.

We found gene and protein properties significantly associated with essentiality in P. berghei. At the mRNA level, essential genes, compared to non-essential genes, were expressed at higher levels in asexual stages, but at lower levels in sexual stages (Figure 3A). Since gene essentiality was measured at the asexual stage, this might explain the positive correlation between essentiality and mRNA expression in asexual stages. Proteins encoded by essential genes were more likely to be detected by mass-spectrometry in different development stages (Figure 3B). Essential genes showed lower evolutionary rates and higher conservation across apicomplexan species (Figure 3C). The higher evolutionary conservation of essential genes is well-documented. We find Gene Ontology classes “Translation”, “Ribosome”, “DNA replication”, “Intracellular protein transport”, “Cytoplasm”, and “Nucleus” to be significantly enriched in essential genes (Figure 4). “Translation” class was also enriched in essential genes after excluding “Ribosome” genes (69% essential; Chi-square test; p = 0.0001), suggesting that enrichment of essential genes in the “Translation” category is not only due to ribosomal genes. Thus enzymes involved in protein translation may be important targets for anti-malarial drug discovery.

Data availability
The in vivo relative growth rate data for 2574 genes of P. berghei genes was obtained from PlasmoGEM database (http://plasmogem.sanger.ac.uk/phenotypes)8. The genome-wide CRISPR screening data for the relative fitness of 8151 T. gondii genes during infection of human fibroblasts cells was obtained from Sidik et al.9.

Dataset 1: Fitness, expression, functionality, conservation and evolutionary information of Plasmodium berghei genes. doi, 10.5256/f1000research.10559.d148698

Author contributions
G.P.S. conceived and designed the study, performed the research and wrote the manuscript.

Competing interests
No competing interests were disclosed.

Grant information
The work is supported by an Early Career Fellowship to G.P.S. by the Wellcome Trust/DBT India Alliance (IA/E/15/1/502297).

Acknowledgements
The author would like to acknowledge suggestions and criticism on the manuscript by Ms. Preeti Goel.
References

Open Peer Review

Current Peer Review Status: ?  ?

Version 1

Reviewer Report 13 February 2017

https://doi.org/10.5256/f1000research.11378.r19792

© 2017 Picard D. This is an open access peer review report distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Didier Picard
Department of Cell Biology, University of Geneva, Geneva, Switzerland

This Research Note reports on an interesting and potentially useful exercise to identify and to prioritize candidates for target-based drug development in Plasmodium. The whole approach is relatively straightforward and provides a list of candidates to think about, not more, not less. Additional considerations could subsequently be applied by others to home in on reasonable targets to focus on. Overall, this short report was worth publishing, but would benefit from some revisions outlined below.

Specific comments:
1. How many genes are experimentally essential in both species is mentioned in the text at a relatively late stage of the presentation. It would be helpful to mention it earlier, e.g. in the legend to Figure 1 (the number of red dots).

2. At some point, the author focuses on enzymes as targets. I do not think that enzymes are the only druggable targets. But if that's what the author wants to focus on, the term "enzyme" should be defined. Is it just based on the GO term associated with these genes/proteins?

3. 40% sequence identity is still a lot, and may be too much if active sites are even more highly conserved. Moreover, in this context I also agree with point 2 of the referee report by Gregory Crowther.

4. While I agree with Gregory Crowther's comment 3 about the relevance to drug discovery of the data in Figures 3 and 4, I still find this analysis interesting and not superfluous in the context of the overall story presented here.

5. Figure 2: I share the confusion with Gregory Crowther with respect to the math here. The text at the bottom of page 3 clearly suggests that 245 = 30+83+151, which of course cannot be. This needs to be fixed/clarified.

References

**Competing Interests:** No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

© 2017 Crowther G. This is an open access peer review report distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Gregory J. Crowther
Division of Biological Sciences, School of STEM, University of Washington, Bothell, WA, USA

This paper analyzes genome-wide data on gene essentiality from two apicomplexan parasites: Plasmodium berghei (the cause of malaria in rodents) and Toxoplasma gondii (the cause of toxoplasmosis). The paper is a new analysis of previously reported data (rather than a presentation of new wet-lab results), which is fine. Those whole-genome datasets are so rich that the papers with the original data cannot possibly cover every interesting angle, so I am happy to see interesting follow-up papers such as this one, which offers additional insight into the datasets.

The following comments go from broad to specific.

**Broad**

While the analysis is interesting, I'm not fully convinced that it advances malaria drug discovery in important ways; it might actually be most useful as an investigation of basic apicomplexan parasite biology. Target-based drug discovery researchers are certainly glad to know whether particular genes of interest (corresponding to specific enzymes or pathways in which they have expertise) are essential or not. However, the figures present genome-wide trends that, while interesting, don’t seem that helpful in prioritizing possible drug targets.

- Figure 1 is probably the most relevant to drug discovery. It shows that genes found to be essential in one species (P. berghei or T. gondii) are more likely to also be essential in the other; thus, P. berghei genes not covered by the Gomes et al. (2015) screen are fairly likely to be essential if their T. gondii orthologs are essential.

- Figure 2 shows a prioritization exercise which is not incorrect, but I don’t think sequence similarity to human proteins is an especially useful criterion. (This is also a limitation of Table 1, in my view). The hope is that we can avoid toxicity by targeting parasite proteins that are dissimilar to human proteins; however, overall sequence similarities tell us very little about whether a parasite protein
will have any binding pockets (each of which represents a small part of the total amino acid sequence) that, in three dimensions, closely resemble any binding pockets of human proteins.

- Figure 3 shows gene expression data at the level of transcripts and proteins; I don’t think this information really applies to drug discovery. (For example, I don’t think anyone should say of a particular target, “Well, this isn’t highly expressed; maybe it isn’t a good/essential target after all.”. If I recall correctly, some excellent targets such as DHFR and PfATP4 are not expressed that highly)

- Figure 4 shows that some functional classes of proteins have a higher percentage of essential proteins than others – but I don’t think this helps us choose possible drug targets either. Even the right-most categories have plenty of essential genes, which is why, for example, there is interest in targeting fatty acid metabolism, the second-lowest category in terms of percent essentiality (see, for example, Shears et al.\(^2\)). Likewise, the unimpressive-looking “transport” category (~52% essential) includes PfATP4, a red-hot target of current Plasmodium research (see Wells et al.\(^3\)). Drug discovery researchers do not usually think in terms of the big broad categories shown in Figure 4, so knowing percent essentiality by category won’t help them much with target selection.

The above observations lead me to the overall recommendation to revise the paper in one of two ways. Option 1 is to emphasize the drug-discovery stuff less and the basic biology more. Option 2 is to enhance the drug-discovery theme by addressing my concerns about the figures (i.e. explaining why they are more relevant to drug discovery than I’m giving them credit for) and/ or adding analyses that have clearer, stronger relevance to drug discovery. The paper does not currently try to combine the essentiality data with genome-wide predictions of “druggability” (which are hard!), but perhaps a collaborator could be enlisted to help with that. In general, most proteins (including most essential proteins) are not that druggable, so essentiality information in the absence of druggability information does not get us that far down the drug-discovery road.

Specific

- Figure 1B: The legend says that green dots represent “non-conserved” proteins. I think that only conserved proteins are shown in this panel, and the green dots are proteins that are neither essential in both species nor nonessential in both species. Please check.

- Figure 2: Aside from my above-mentioned concern about homology to human proteins, it might make sense to show the arrows as follows: 710 => 289 => 245 => 151 => 83 => 30, thus showing the winnowing of the targets with additional criteria. In its current form, the figure initially led me to think, incorrectly, that the 245 genes could be split into subgroups of 30, 83, and 151.

- Figure 3: For 3A and 3B, the transcriptome data (relative abundance) don’t seem to correlate that closely with the proteome data (detectable or not). For example, essential gene expression in the sexual stages looks low at the level of RNA in 3A but average-to-high at the protein level in 3B. Are such discrepancies surprising/interesting? Discuss in the Discussion! Also, briefly define dN and dS (nonsynonymous and synonymous substitutions; 3C) somewhere in the paper. Also, to improve clarity, consider using one color for the bars corresponding to the asexual stages and another color for the bars corresponding to the sexual stages.

- Figure 4: Others must have done analyses like this for other (non-apicomplexan) species, e.g., of bacteria. Please compare the Figure 4 data to previous work in the Discussion. Also, why did the “cytoplasm” category come out as statistically significant? Are there a huge number of genes in
that category?

• Can a paragraph be added to the Discussion on what sort of specific work might follow naturally from the present analysis? That would help readers appreciate the significance of the present work.

References


**Competing Interests:** No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.