RESEARCH NOTE

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

Gajinder Pal Singh

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

v1 **First published:** 09 Jan 2017, **6**:23 (doi: 10.12688/f1000research.10559.1)
Latest published: 09 Jan 2017, **6**:23 (doi: 10.12688/f1000research.10559.1)

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.

Open Peer Review

Referee Status: ? ?

	Invited Referees 1	2
version 1 published 09 Jan 2017	? report	? report

1 **Gregory J. Crowther**, University of Washington USA

2 **Didier Picard**, University of Geneva Switzerland

Discuss this article

Comments (0)

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

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; referees: 2 approved with reservations]** *F1000Research* 2017, **6**:23 (doi: 10.12688/f1000research.10559.1)

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.

Competing interests: No competing interests were disclosed.

First published: 09 Jan 2017, **6**:23 (doi: 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[1]. The emergence and spread of *Plasmodium falciparum* strains resistant to all currently used anti-malarial drugs[2] 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[3]. There are two main approaches for drug-discovery against pathogens: Phenotype screening and target-based approach[4]. 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[5]. Some of these inhibitors have progressed to clinical trials[6]. 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[6]. However, recent large-scale screening of about 2500 genes in a rodent malaria parasite *P. berghei* has identified about 1200 essential genes[7,8]. A recent genome-scale CRISPR screen in a related apicomplexan parasite *Toxoplasma gondii* has identified about 3000 essential genes[9]. 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* genes during infection of human fibroblasts cells was obtained from Sidik *et al.*[9]. The authors defined $\log_2$ fold change in abundance of single guide RNA (sgRNA) targeting a given gene as the "phenotype" score for that gene[9]. 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[9]. 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 the PlasmoGEM database[7,8] (http://plasmogem.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[7]. 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[7].

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* Sal1, *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 EuPathDB[11]: *Cryptosporidium hominis* TU502 (http://cryptodb.org/common/downloads/release-29/ChominisTU502/); *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 algorithm[12].

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, dN/S values were obtained from Hall *et al.*[14]. Gene Ontology information for *P. falciparum* was obtained from PlasmoDB[10], 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 TMHMM[15]. 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 essential[7,8]. Similarly, *in vivo* relative fitness of knockout mutants for 8151 *T. gondii* genes have been measured[9], 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[16]).

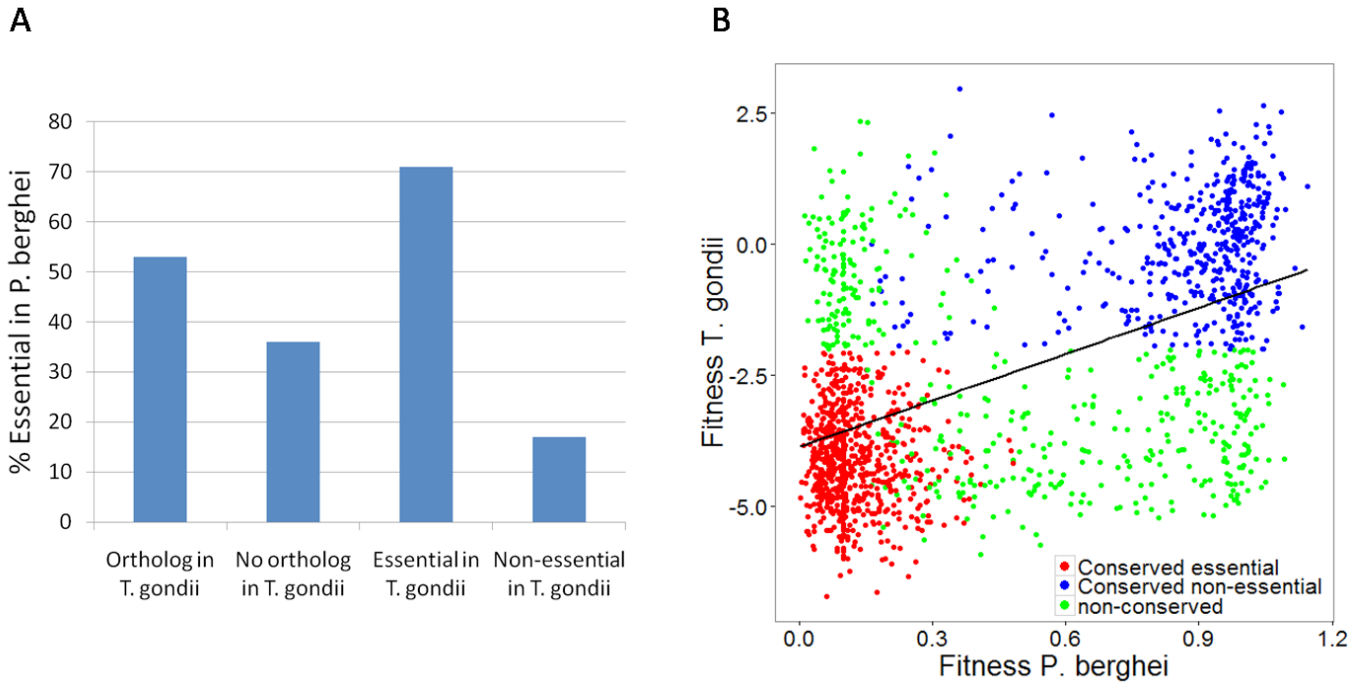

Figure 1. Conservation of essentiality between *Plasmodium berghei* and *Toxoplasma gondii*. (**A**) *P. berghei* genes with an ortholog in *T. gondii* were more likely to be essential, compared to *P. berghei* genes without an ortholog in *T. gondii* (Fisher test p = 7e-18). *P. berghei* genes with 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* (Fisher test p = 6e-59). (**B**) 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). Genes classified as essential in both species are colored red. Genes classified as non-essential in both species are colored blue. Genes that are essential in only one of the species are colored green.

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[7], 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 in human malarial species, such as *P. falciparum* and *P. vivax*; 4) 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[17]. 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[16]). 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[16]).

Properties of essential *P. berghei* genes

Essential genes show different expression, evolutionary and functional properties[9]. 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).

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

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

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).

Dataset 1. Fitness, expression, functionality, conservation and evolutionary information of *Plasmodium berghei* genes

http://dx.doi.org/10.5256/f1000research.10559.d148698

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 antimalarial 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

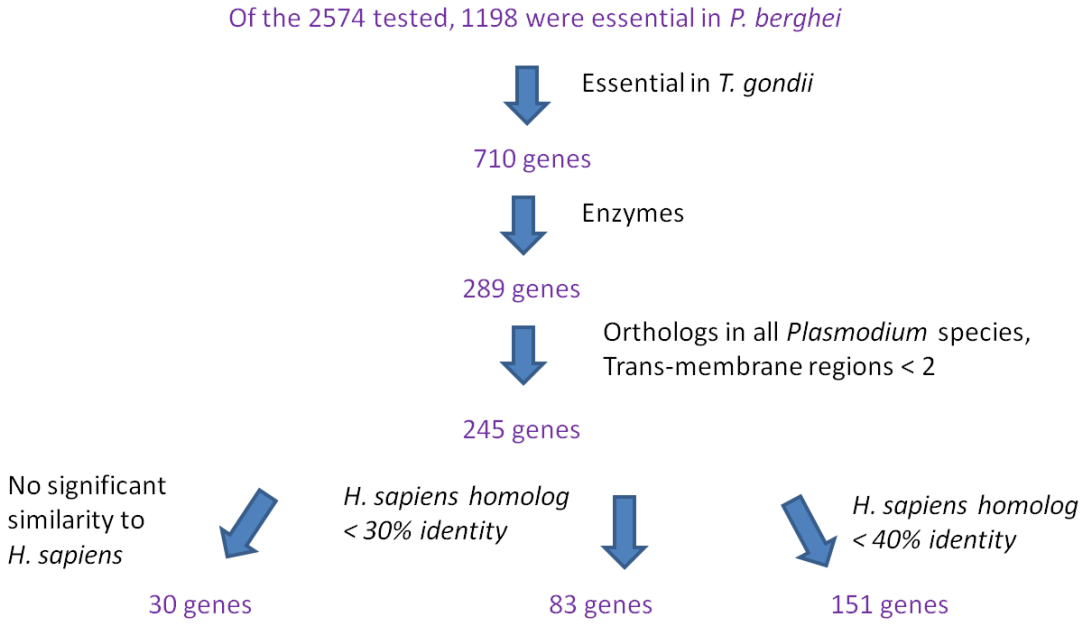

Figure 2. Selection of potential drug targets in *Plasmodium*.

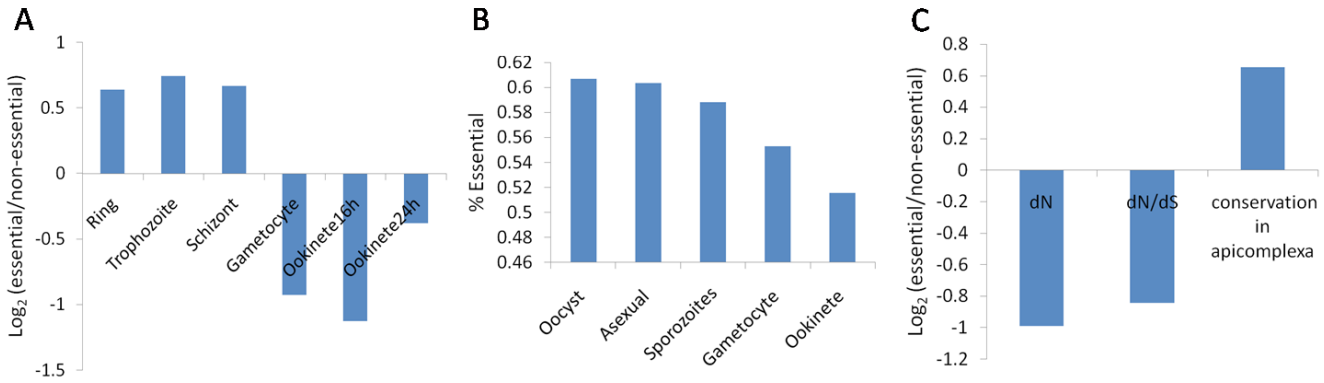

Figure 3. Properties of essential *Plasmodium berghei* genes. (**A**) Essential *P. berghei* genes showed higher mRNA expression levels in asexual stages, but lower mRNA expression levels in sexual stages. The mean FPKM values for the essential and non-essential genes were calculated for different development stages and their $\log_2$ 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 $\log_2$ ratio was taken. This data was taken from Hall *et al.*[14]. 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 $\log_2$ 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.

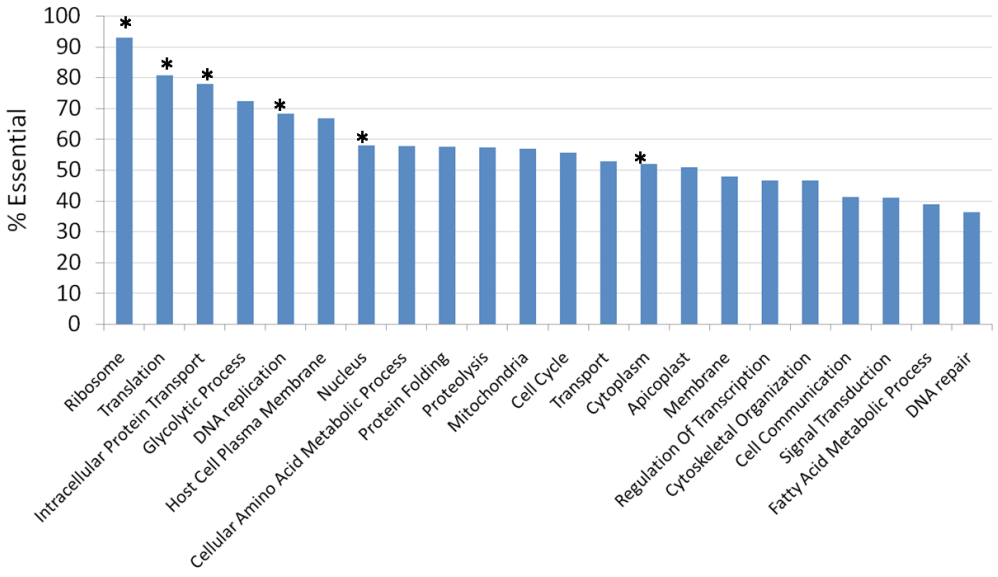

Figure 4. Prevalence of essential genes in different functional classes. The Gene Ontology information for *Plasmodium falciparum* genes was obtained from PlasmoDB[10] and assigned to their *P. berghei* orthologs. Classes with a significant difference (Chi-square test; $p < 0.05$) in essential genes are marked with *.

that are essential in multiple species and stages of *Plasmodium* are particularly desirable[17]. 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[18,19]. 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[20]. 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[16]

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

1. World Health Organization: **The World Malaria Report 2015.** 2015.
Reference Source
2. Duru V, Witkowski B, Ménard D: ***Plasmodium falciparum* Resistance to Artemisinin Derivatives and Piperaquine: A Major Challenge for Malaria Elimination in Cambodia.** *Am J Trop Med Hyg.* 2016; **95**(6): 1228–1238.
PubMed Abstract | Publisher Full Text | Free Full Text
3. Alonso PL, Brown G, Arevalo-Herrera M, *et al.*: **A research agenda to underpin malaria eradication.** *PLoS Med.* 2011; **8**(1): e1000406.
PubMed Abstract | Publisher Full Text | Free Full Text
4. Gilbert IH: **Drug discovery for neglected diseases: molecular target-based and phenotypic approaches.** *J Med Chem.* 2013; **56**(20): 7719–7726.
PubMed Abstract | Publisher Full Text | Free Full Text
5. Spangenberg T, Burrows JN, Kowalczyk P, *et al.*: **The open access malaria box: a drug discovery catalyst for neglected diseases.** *PLoS One.* 2013; **8**(6): e62906.
PubMed Abstract | Publisher Full Text | Free Full Text
6. Wells TN, Hooft van Huijsduijnen R, Van Voorhis WC: **Malaria medicines: a glass half full?** *Nat Rev Drug Discov.* 2015; **14**(6): 424–442.
PubMed Abstract | Publisher Full Text
7. Gomes AR, Bushell E, Schwach F, *et al.*: **A genome-scale vector resource enables high-throughput reverse genetic screening in a malaria parasite.** *Cell Host Microbe.* 2015; **17**(3): 404–413.
PubMed Abstract | Publisher Full Text | Free Full Text
8. Schwach F, Bushell E, Gomes AR, *et al.*: ***Plasmo*GEM, a database supporting a community resource for large-scale experimental genetics in malaria parasites.** *Nucleic Acids Res.* 2015; **43**(Database issue): D1176–D1182.
PubMed Abstract | Publisher Full Text | Free Full Text
9. Sidik SM, Huet D, Ganesan SM, *et al.*: **A Genome-wide CRISPR Screen in *Toxoplasma* Identifies Essential Apicomplexan Genes.** *Cell.* 2016; **166**(6): 1423–1435.e12.
PubMed Abstract | Publisher Full Text | Free Full Text
10. Aurrecoechea C, Brestelli J, Brunk BP, *et al.*: **PlasmoDB: a functional genomic database for malaria parasites.** *Nucleic Acids Res.* 2009; **37**(Database issue): D539–D543.
PubMed Abstract | Publisher Full Text | Free Full Text
11. Aurrecoechea C, Barreto A, Basenko EY, *et al.*: **EuPathDB: the eukaryotic pathogen genomics database resource.** *Nucleic Acids Res.* 2016; **45**(D1): D581–D591, pii: gkw1105.
PubMed Abstract | Publisher Full Text
12. Wolf YI, Koonin EV: **A tight link between orthologs and bidirectional best hits in bacterial and archaeal genomes.** *Genome Biol Evol.* 2012; **4**(12): 1286–1294.
PubMed Abstract | Publisher Full Text | Free Full Text
13. Otto TD, Böhme U, Jackson AP, *et al.*: **A comprehensive evaluation of rodent malaria parasite genomes and gene expression.** *BMC Biol.* 2014; **12**: 86.
PubMed Abstract | Publisher Full Text | Free Full Text
14. Hall N, Karras M, Raine JD, *et al.*: **A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses.** *Science.* 2005; **307**(5706): 82–86.
PubMed Abstract | Publisher Full Text
15. Krogh A, Larsson B, von Heijne G, *et al.*: **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.** *J Mol Biol.* 2001; **305**(3): 567–580.
PubMed Abstract | Publisher Full Text
16. Singh G: **Dataset 1 in: Conservation of gene essentiality in Apicomplexa and its application for prioritization of anti-malarial drug targets.** *F1000Research.* 2017.
Data Source
17. Burrows JN, van Huijsduijnen RH, Möhrle JJ, *et al.*: **Designing the next generation of medicines for malaria control and eradication.** *Malar J.* 2013; **12**: 187.
PubMed Abstract | Publisher Full Text | Free Full Text
18. Magariños MP, Carmona SJ, Crowther GJ, *et al.*: **TDR Targets: a chemogenomics resource for neglected diseases.** *Nucleic Acids Res.* 2012; **40**(Database issue): D1118–D1127.
PubMed Abstract | Publisher Full Text | Free Full Text
19. Crowther GJ, Shanmugam D, Carmona SJ, *et al.*: **Identification of attractive drug targets in neglected-disease pathogens using an *in silico* approach.** *PLoS Negl Trop Dis.* 2010; **4**(8): e804.
PubMed Abstract | Publisher Full Text | Free Full Text
20. Doyle MA, Gasser RB, Woodcroft BJ, *et al.*: **Drug target prediction and prioritization: using orthology to predict essentiality in parasite genomes.** *BMC Genomics.* 2010; **11**: 222.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Referee Status: ? ?

Version 1

Referee Report 13 February 2017

doi:10.5256/f1000research.11378.r19792

? **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 conext I also agree with point 2 of the referee report by Gregory Crowther[1].

4. While I agree with Gregory Crowther's comment 3[1] 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[1] 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

1. Crowther GJ: Referee Report For: Conservation of gene essentiality in Apicomplexa and its application for prioritization of anti-malarial drug targets [version 1; referees: 1 approved with reservations]. *F1000Research*. 2017; **6** (23).

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.

Competing Interests: No competing interests were disclosed.

Referee Report 16 January 2017

doi:10.5256/f1000research.11378.r19085

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[1] 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

1. Gomes AR, Bushell E, Schwach F, Girling G, Anar B, Quail MA, Herd C, Pfander C, Modrzynska K, Rayner JC, Billker O: A genome-scale vector resource enables high-throughput reverse genetic screening in a malaria parasite.*Cell Host Microbe*. 2015; **17** (3): 404-13 PubMed Abstract | Publisher Full Text
2. Shears MJ, Botté CY, McFadden GI: Fatty acid metabolism in the Plasmodium apicoplast: Drugs, doubts and knockouts.*Mol Biochem Parasitol*. **199** (1-2): 34-50 PubMed Abstract | Publisher Full Text
3. Wells TN, Hooft van Huijsduijnen R, Van Voorhis WC: Malaria medicines: a glass half full?. *Nat Rev Drug Discov*. 2015; **14** (6): 424-42 PubMed Abstract | Publisher Full Text

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.

Competing Interests: No competing interests were disclosed.