RESEARCH ARTICLE

Genome-wide characterization of folate transporter proteins of eukaryotic pathogens [version 1; peer review: 2 approved with reservations]

Mofolusho Falade¹¹, Benson Otarigho¹¹²

¹Cellular Parasitology Programme, Department of Zoology, University of Ibadan, Ibadan, Nigeria
²Department of Biological Science, Edo University, Iyamho, Nigeria

Abstract

Background: Medically important pathogens are responsible for the death of millions every year. For many of these pathogens, there are limited options for therapy and resistance to commonly used drugs is fast emerging. The availability of genome sequences of many eukaryotic protozoa is providing important data for understanding parasite biology and identifying new drug and vaccine targets. The folate synthesis and salvage pathway are important for eukaryote pathogen survival and organismal biology and may present new targets for drug discovery.

Methods: We applied a combination of bioinformatics methods to examine the genomes of pathogens in the EupathDB for genes encoding homologues of proteins that mediate folate salvage in a bid to identify and assign putative functions. We also performed phylogenetic comparisons of identified proteins.

Results: We identified 234 proteins to be involved in folate transport in 63 strains, 28 pathogen species and 12 phyla, 60% of which were identified for the first time. Many of the genomes examined contained genes encoding transporters such as folate-binding protein YgfZ, folate/pteridine transporter, folate/biopterin transporter, reduced folate carrier family protein, folate/methotrexate transporter FT1. The mitochondrion is the predicted location of the majority of the proteins, with 15% possessing signal peptides. Phylogeney computation shows the similarity of the proteins identified.

Conclusion: These findings offer new possibilities for potential drug development targeting folate-salvage proteins in eukaryotic pathogens.

Keywords
Folate transporter, Eukaryotic pathogens, Drug discovery, Putative homologues

Open Peer Review

Reviewer Status

Invited Reviewers

---

First published: 12 Jan 2017, 6:36
https://doi.org/10.12688/f1000research.10561.1

Latest published: 13 Jul 2017, 6:36
https://doi.org/10.12688/f1000research.10561.2

---

1 Raphael D. Isokpehi, Bethune-Cookman University, Daytona Beach, USA
2 Gajinder Singh, International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi, India

Any reports and responses or comments on the article can be found at the end of the article.
Corresponding author: Mofolusho Falade (mo.falade@ui.edu.ng)

Competing interests: No competing interests were disclosed.

Grant information: B.O. was supported by a TWAS-CNPq fellowship (FP number: 3240274297)
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2017 Falade M and Otarigho B. This is an open access article distributed under the terms of the Creative Commons Attribution License, 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: Falade M and Otarigho B. Genome-wide characterization of folate transporter proteins of eukaryotic pathogens [version 1; peer review: 2 approved with reservations] F1000Research 2017, 6:36 https://doi.org/10.12688/f1000research.10561.1

First published: 12 Jan 2017, 6:36 https://doi.org/10.12688/f1000research.10561.1
Introduction

A heterogeneous diversity of eukaryotic pathogens is responsible for the most economically important diseases of humans and animals. As a result of underdevelopment, a lack of social infrastructure and insufficient funding of public health facilities, most of these pathogens are endemic to resource-poor countries in sub-Saharan Africa, South-East Asia and South America, where they are responsible for high morbidity and mortality. Of these, parasitic protozoa form a major group, with the apicomplexans and kinetoplastid parasites represented by important members, which cause diseases such as malaria, cryptosporidiosis, toxoplasmosis, babesiosis, leishmaniasis, and American trypanosomiasis and South American trypanosomiasis or Chagas’ disease causing most of the morbidity and mortality. Other important diseases caused by protozoans include giardiasis, amoebic dysentery and trichomoniasis. A vicious cycle of poverty and disease exists for most of these parasites with a high infection and death rate in affected populations. The appreciable burden of disease caused by these parasites has been aggravated by the lack of a licensed vaccine for most of them. Furthermore, current drugs of choice for treatment for many of the parasites have significant side effects, with the added emergence of drug resistant strains. Despite the urgent demand for new therapies for control, few drugs have been developed to combat these parasites. A major limitation to the development of new drugs is the paucity of new drug targets. There is therefore a need for discovery of novel and alternate potential chemotherapeutic targets that can help in drug development efforts for disease control. A possible approach to selective antimicrobial chemotherapy has been to exploit the inhibition of unique targets, vital to the pathogen and absent in mammals.

A metabolic pathway that has been exploited considerably for the development of drugs is the folate biosynthetic pathway. Antifolate drugs target this pathway and are the most important and successful antimicrobial chemotherapies targeting a range of bacterial and eukaryotic pathogens. While most parasitic protozoa can synthesize folates from simple precursors, such as GTP, p-aminobenzoic acid (pABA) and glutamate, higher animals and humans cannot. Additionally, a few of these parasites can also salvage folate as nutrient from their host. These folate compounds are important for synthesis of DNA, RNA and membrane lipids and are transported via receptor-mediated or carrier-mediated transmembrane proteins; folate transporters. Importantly, antifolate chemotherapies that target the biosynthesis and processing of folate cofactors have been effective in the chemotherapy of bacterial and protozoan parasites. More importantly, the folate pathway has also been confirmed as being essential in some eukaryotic pathogens such as Plasmodium, trypanosomes and Leishmania.

In addition to the folate biosynthesis pathway, proteins that mediate transport of useful nutrients such as folic acid have been identified as important chemotherapeutic drug targets. Hence, the folate pathway, metabolites and transporters continue to be extensively studied for identification of new enzymes including transporters, which may serve as new drug targets. Recent estimates have ascribed eight different membrane transporters to eukaryotes.

Proteins that mediate transportation of folates have been well studied in a few parasites such as Plasmodium falciparum, Trypanosoma brucei, Leishmania donovani and Leishmania major. These studies have provided information on mode of action of drugs in addition to studies describing mechanisms of parasite drug resistance. However, folate transport proteins remain unidentified and uncharacterized in many other eukaryotic pathogens. This is despite the sequencing of the genomes of most eukaryotic pathogens, which has produced a vast wealth of data that could aid in identification of drugable pathogen-specific proteins. It is therefore imperative to search and identify from these parasite genomes additional proteins such as folate transporters that may serve as novel drug targets.

Therefore, in an attempt to identify and characterize targets for novel therapeutics, we report herein an extensive search of folate transporters from pathogen genomes. In addition, we investigated the evolutionary relationship of these transporters in a bid to determine similarities and differences that make them attractive drug targets. The knowledge provided may assist in the design of new antifolates for protozoan parasites.

Methods

Our experiment workflow is depicted in Figure 1. We extracted protein sequences of approximately 200 pathogens that mediate transportation or salvage of folates from Eukaryotic Pathogen Genome Database Resources and from the literature using a key-word search. We also searched the 200 pathogen genome sequences archived at the Eukaryotic Pathogen Genome Database Resources. The search was for all proteins that mediate transportation or folate salvage alone or folate salvage and related compounds (such as pteridine, biopterin and methotrexate) together. This database gives public access to most sequenced emerging/re-emerging infectious pathogen genomes. We utilized the word “folate” for search on the gene text and “folic acid” was used to confirm the hits. Hit results containing proteins annotated as folate-binding protein YgfZ, folate/pteridine transporter, folate/biopterin transporter, reduced folate carrier family protein, folate/methotrexate transporter FT1, Folate transporters alone and other folate related proteins were retrieved. The complete list of proteins extracted from Eupathdb is presented in Dataset 1. The folate transporters were classified based on type of transporter, number of transmembrane helix (TMH) and localization (either cell or mitochondrial membrane) of transporter. Gene sequences were obtained in FASTA format for transporter proteins using the sequence download tool on Eupathdb.

To ensure that most of the retrieved proteins had not been previously studied, we performed a literature search on PubMed and Google Scholar using the query “folate transporter + Parasite name”. The protein sequence information (Dataset 1 and Table 1) obtained from literature search was used for a BLAST search on EupathDB, UniprotDB (http://www.uniprot.org) and GeneDB (http://www.genedb.org/Homepage). Sequence data were edited on textEdit mac version.
and uploaded to Molecular Evolutionary Genetics Analysis (MEGA) platform version 7.0 obtained from http://www.megasoftware.net. The 234 sequences were aligned using muscle tools with large alignment (Max iterations = 2) selected while other settings were left at defaults. Evolutionary history was inferred using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) was also analysed. The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the number of differences method. While uniform rate and complete deletion was selected for substitution rates and data subset, respectively. Other parameters were at default settings. All positions containing gaps and missing data were eliminated. The newick format of the tree was exported and opened on FigTree 1.4.2 platform downloaded from http://tree.bio.ed.ac.uk/software/figtree. The final tree was constructed using radial tree layout. Additional analysis consisted of sub-phylogenies based on the transporter type. Since folate-binding protein YgfZ, folate/pteridine transporter, folate/biopterin transporter, putative, reduced folate carrier family protein, folate/methotrexate transporter FT1, putative folate transporters alone and others have 10, 25, 132, 2, 7, 49 and 9. So we decided to reconstruct the phylogeny based folate transporter, folate-biopterin transporter after considering the identification number, the species diversity in each category.

**Results**

A methodological search for folate transporters in all eukaryotic pathogen genomes we examined under EupathDB with validation via GenBank, GeneDB and Uniprot contained a total of 234 proteins (detail features of proteins are presented in Dataset 1). We identified these transporters in 28 pathogen species (containing 63 strains) cutting across 12 phyla (Table 1). The parasites with the highest number of folate transporters are *Phytophthora parasitica* INRA-310, *P. infestans* T30–4 and *Leptomonas pyrrhocoris* H10 with 20, 16 and 16 proteins, respectively. While *Aspergillus clavatus* NRRL 1, *A. flavus* NRRL3357, *A. macrogynus* ATCC 38327, *Crithidia fasciculata* strain Cf-CI and others have one folate transporter protein each (Table 1). The different proteins identified to be involved in folate salvage or related molecules were folate-binding protein YgfZ, folate/pteridine transporter, folate/biopterin transporter, reduced folate carrier family protein, folate/methotrexate transporter having 4%, 11%, 56%, 1%, 3% and 21% identity, respectively. Proteins that did not belong to these groups were classified as others (4%) (Figure 2A).

![Figure 1. Conceptual hierarchical methodology workflow showing the various steps employed in this study to identify, characterize and curate protozoan folate transporters.](http://tree.bio.ed.ac.uk/software/figtree)
| No. | Species/Strain | Genera            | Class                  | Order                  | Family               | Confirmation Search (References) |
|-----|----------------|-------------------|------------------------|------------------------|----------------------|----------------------------------|
| 1   | A. capsulatus G186AR | Actinobacillus    | Gammaproteobacteria    | Pasteurellales         | Streptobacillales     | X                                |
| 2   | A. clavatus NRRL1   | Aspergillus       | Eurotiomycetes         | Eurotiales             | Trichocomaceae       | X                                |
| 3   | A. flavus NRRL3357  | Aspergillus       | Eurotiomycetes         | Eurotiales             | Trichocomaceae       | X                                |
| 4   | A. macrogynus ATCC38227 | Allomyces        | Blastocladiomycetes    | Blastocladiaceae      | Blastocladiomycetes  | X                                |
| 5   | C. fasciculata strain Cf-Cl | Cryptosporidium | Cryptosporidomycetes   | Cryptosporidiidae     | Chromerida           | X                                |
| 6   | C. immitis RS       | Cryptosporidium   | Cryptosporidomycetes   | Cryptosporidiidae     | Chromerida           | X                                |
| 7   | C. neoformans var. grubii H99 | Cryptococcus     | Basidiomycetes         | Cryptobasidomycetes   | Cryptobasidomycetes  | X                                |
| 8   | C. velia CCMP2878   | Cryptococcus      | Cryptobasidomycetes    | Cryptobasidomycetes   | Cryptobasidomycetes  | X                                |
| 9   | E. acervulina        | Eimeria           | Microsporidia          | Microsporida           | Microsporida         | X                                |
| 10  | E. brunetti         | Eimeria           | Microsporidia          | Microsporida           | Microsporida         | X                                |
| 11  | E. cuniculi GB-M1   | Eimeria           | Microsporidia          | Microsporida           | Microsporida         | X                                |
| 12  | E. maxima Houghton   | Eimeria           | Microsporidia          | Microsporida           | Microsporida         | X                                |
| 13  | E. necatrix Houghton | Eimeria           | Microsporidia          | Microsporida           | Microsporida         | X                                |
| 14  | E. praecox Houghton  | Eimeria           | Microsporidia          | Microsporida           | Microsporida         | X                                |
| 15  | E. velia strain Houghton | Eimeria         | Microsporidia          | Microsporida           | Microsporida         | X                                |
| 16  | F. tularensis       | Eimeria           | Microsporidia          | Microsporida           | Microsporida         | X                                |
| N° | Species/Strain | Genera | Family | Order | Class | Phylum/Division | Kingdom | Confirmation Search (References) |
|----|----------------|--------|--------|-------|-------|----------------|---------|----------------------------------|
| 17 | F. graminearum PH-1 | Gibberella | Neotyphodium | Hypocreales | Eurotiales | Ascomycota | Animalia | X |
| 18 | H. hammondii strain H.H.34 | Hammondia | Sarcocystidae | Trypanosomatida | Trypanosomatidae | Kinetoplastida | Eukaryota | X |
| 19 | L. braziliensis MHOM/BR/75/RC904 | Leishmania | Trypanosomatida | Trypanosomatida | Trypanosomatida | Kinetoplastida | Euglenozoa | X |
| 20 | L. donovani BPK282A1 | Leishmania | Trypanosomatida | Trypanosomatida | Trypanosomatida | Kinetoplastida | Euglenozoa | X |
| 21 | L. infantum JPCM5 | Leishmania | Trypanosomatida | Trypanosomatida | Trypanosomatida | Kinetoplastida | Euglenozoa | X |
| 22 | L. major strain Friedlin | Leishmania | Trypanosomatida | Trypanosomatida | Trypanosomatida | Kinetoplastida | Euglenozoa | X |
| 23 | L. mexicana MHOM/GT/2001/U1103 | Leishmania | Trypanosomatida | Trypanosomatida | Trypanosomatida | Kinetoplastida | Euglenozoa | X |
| 24 | L. pyrolophora | Leptomonas | Leptomonadida | Leptomonadida | Leptomonadida | Kinetoplastida | Euglenozoa | X |
| 25 | L. seymouri ATCC 30220 | Leptomonas | Leptomonadida | Leptomonadida | Leptomonadida | Kinetoplastida | Euglenozoa | X |
| 26 | L. seymouri ATCC 30220 | Leptomonas | Leptomonadida | Leptomonadida | Leptomonadida | Kinetoplastida | Euglenozoa | X |
| 27 | N. bombycis CQ1 | Nosema | Nosematida | Nosematida | Nosematida | Diniplasiea | Microsporidia | Fungi |
| 28 | N. caninum Liverpool | Neospora | Sarcocystidae | Trypanosomatida | Trypanosomatida | Kinetoplastida | Euglenozoa | X |
| 29 | N. fischeri NRRL 1811 | Neospora | Sarcocystidae | Trypanosomatida | Trypanosomatida | Kinetoplastida | Euglenozoa | X |
| 30 | N. fowleri ATCC 39338 | Naegleria | Vahlkampfiidae | Schizophysida | Vahlkampfiidae | Heterolobosea | Fungi | X |
| 31 | P. berghei ANKA | Plasmodium | Plasmodiida | Plasmodiida | Plasmodiida | Chromerida | Apicomplexa | X |
| 32 | P. chabaudi chabaudi | Plasmodium | Plasmodiida | Plasmodiida | Plasmodiida | Chromerida | Apicomplexa | X |
| 33 | P. falciparum 3D7 | Plasmodium | Plasmodiida | Plasmodiida | Plasmodiida | Chromerida | Apicomplexa | X |
| No. | Strain/Species | Genus | Family | Order | Class | Phylum/Division | Kingdom | Confirmation Search (References) |
|-----|----------------|-------|--------|-------|-------|----------------|---------|---------------------------------|
| 34  | P. falciparum IT | Plasmodium | Plasmodiidae | Chromatorida | Aconoidasida | Apicomplexa | Eukaryota | √ |
| 35  | P. infestans T30-4 | Phytophthora | Pythiaceae | Peronosporales | Oomycota | Heterokontophyta | Chromalveolata | X |
| 36  | P. knowlesi strain H | Plasmodium | Plasmodiidae | Chromatorida | Aconoidasida | Apicomplexa | Eukaryota | X |
| 37  | P. parasitica INRA-310 | Phytophthora | Pythiaceae | Peronosporales | Oomycetes | Oomycota | Fungi | X |
| 38  | P. reichenowi CDC | Plasmodium | Plasmodiidae | Chromatorida | Aconoidasida | Apicomplexa | Eukaryota | X |
| 39  | P. ultimum DAOM BR144 | Pythium | Pythiaceae | Pythiales | Oomycetes | Heterokontophyta | Chromalveolata | X |
| 40  | P. vivax Sal-1 | Plasmodium | Plasmodiidae | Chromatorida | Aconoidasida | Apicomplexa | Eukaryota | X |
| 41  | P. yoelii yoelii 17X | Plasmodium | Plasmodiidae | Chromatorida | Aconoidasida | Apicomplexa | Eukaryota | X |
| 42  | P. yoelii yoelii YM | Plasmodium | Plasmodiidae | Chromatorida | Aconoidasida | Apicomplexa | Eukaryota | X |
| 43  | S. neurona SN3 | Sarcocystis | Sarcocystidae | Eucoccidiorida | Conoidasida | Apicomplexa | Eukaryota | X |
| 44  | S. punctatus DAOM BR117 | Spizellomyces | Powellomyctecetaceae | Spizellomycetales | Chytridiomycota | Chromalveolata | X |
| 45  | T. brucei gambiense DAL972 | Trypanosoma | Trypanosomatidae | Kinetoplastea | Euglenozoa | Excavata | X |
| 46  | T. brucei Lister strain 427 | Trypanosoma | Trypanosomatidae | Kinetoplastea | Euglenozoa | Excavata | X |
| 47  | T. brucei TRBC/22 | Trypanosoma | Trypanosomatidae | Kinetoplastea | Euglenozoa | Excavata | X |
| 48  | T. cruzi CL Brener Non-Esmeraldia-like | Trypanosoma | Trypanosomatidae | Kinetoplastea | Euglenozoa | Excavata | X |
| 49  | T. cruzi CL Brener Esmeraldia-like | Trypanosoma | Trypanosomatidae | Kinetoplastea | Euglenozoa | Excavata | X |
| 50  | T. cruzi L3000 | Trypanosoma | Trypanosomatidae | Kinetoplastea | Euglenozoa | Excavata | X |
| 51  | T. cruzi Dm28c | Trypanosoma | Trypanosomatidae | Kinetoplastea | Euglenozoa | Excavata | X |
| S/ N | Species/Strain               | No. of FTP | Genera     | Family            | Order               | Class               | Phylum/Division     | Kingdom     | Confirmation Search (References) |
|------|-----------------------------|-----------|------------|-------------------|----------------------|---------------------|---------------------|-------------|---------------------------------|
| 52   | T. cruzi marinkellei strain B7 | 6         | Trypanosoma| Trypanosomatida    | Trypanosomatida      | Kinetoplastea       | Euglenozoa          | Excavata    | √                               |
| 53   | T. cruzi Sylvio X10/1        | 2         | Trypanosoma| Trypanosomatida    | Trypanosomatida      | Kinetoplastea       | Euglenozoa          | Excavata    | √                               |
| 54   | T. equi strain WA            | 4         | Theileria  | Theileriida        | Proplasmida          | Aconoidasida        | Apicomplexa         | Eukaryota   | √                               |
| 55   | T. evansi strain STIB 805    | 3         | Trypanosoma| Trypanosomatida    | Trypanosomatida      | Kinetoplastea       | Euglenozoa          | Excavata    | X                               |
| 56   | T. gondii GT1                | 3         | Toxoplasma | Sarcozystida       | Eucoecidiorida       | Conoidasida         | Apicomplexa         | Eukaryota   | √                               |
| 57   | T. gondii ME49               | 4         | Toxoplasma | Sarcozystida       | Eucoecidiorida       | Conoidasida         | Apicomplexa         | Eukaryota   | √                               |
| 58   | T. gondii VEG                | 3         | Toxoplasma | Sarcozystida       | Eucoecidiorida       | Conoidasida         | Apicomplexa         | Eukaryota   | √                               |
| 59   | T. grayi ANR4                | 6         | Trypanosoma| Trypanosomatida    | Trypanosomatida      | Kinetoplastea       | Euglenozoa          | Excavata    | √                               |
| 60   | T. parva strain Muguga       | 2         | Theileria  | Theileriida        | Proplasmida          | Aconoidasida        | Apicomplexa         | Eukaryota   | X                               |
| 61   | T. rangeli SC58              | 3         | Trypanosoma| Trypanosomatida    | Trypanosomatida      | Kinetoplastea       | Euglenozoa          | Excavata    | √                               |
| 62   | T. vivax Y486                | 1         | Trypanosoma| Trypanosomatida    | Trypanosomatida      | Kinetoplastea       | Euglenozoa          | Excavata    | X                               |
| 63   | V. brassicafornis CCMP3155   | 12        | Vitrella    | Vitrellaceae       | –                    | –                   | Chromerida          | Eukaryota   | X                               |

| Total | 234                  |

**Keys**

| √     | Folate transporters found from literature search |
| X     | Folate transporters identified from this study   |
| FTP   | Folate Transporter Protein                      |
Our literature search for parasite folate transporters on PubMed and Google Scholar indicated 60% (38 out 63) of the proteins were identified for the first time as presented in Table 1 and Figure 2C, while 40% have been previously investigated. Besides, the *Leishmania* folate transporters we came across were not found on the EupathDB resource. We thus performed a BLAST search of Kinetoplastida on EupathDB, the returned hits were folate/biopterin transporter for *L. infantum*. The only *Plasmodium* species with results for proteins that salvage folate was *P. falciparum*. Our study, however, describes for the first time the presence of these transporters in other *Plasmodium* species. There were no transporter proteins deposited in EupathDB for *P. malariae* and *P. ovale*. However, folate transporters I and II were retrieved from our search of GeneDB for *P. malariae* and *P. ovale curtisi*, respectively.
Figure 2. Categorization of proteins identified. This is based on [A] Transporter Type [B] Number of TMMs [C] Novel folate transporters [D] Localization [E] Presence/absence of Signal peptide. TMM, Transmembrane Helix;
Our analysis of folate transporters indicate the presence in *Plasmodium* species of two proteoforms; folate transporter I and II (Dataset 1\(^4\)). All *Leishmania* species identified possess folate/biotin-erin transporters and not folate transporters. Trypanospora species have both folate/pteridine and folate transporters; *T. cruzi* Dm28c, *T. cruzi* Sylvio X10/1 and *T. cruzi* CL Brener Esmeraldo-like, *T. cruzi* CL Brener Non-Esmeraldo-like and *T. cruzi* marinkellei strain B7 all have folate/pteridine transporter while *T. brucei* TREVU927, *T. brucei* Lister strain 427, *T. brucei gambiense* DAL972, *T. congolense* IL3000 possess folate transporters. Eukaryotic parasites like *Eimeria acervulina* Houghton, *E. brunetti* Houghton, *E. maxima* Weybridge, *E. necatrix* Houghton, *E. praecox* Houghton, *E. tenella* strain Houghton and *Neospora caninum* Cotton all boast folate/methotrexate transporter FT1. The folate-binding protein YgFZ was found in the fungus, *Allomyces macrognymus* ATCC 38327, the protist *C. fasciculata* strain CF-CI, *C. immitis* RS, the feline protozoon, *Hammondia hammondi* strain H.H.34, *Sarcocystis neurona* SN3, *S. punctatus* DAOM BR117, *T. gondii* GT1, *T. gondii* ME49, *T. gondii* VEG and *T. brucei* TREVU927. Parasites such as *Microsporidium daphniae* UGP3 and the amoeba *Naegleria fowleri* ATCC 30863 possess the reduced folate carrier family protein (Figure 2D). We observed that 7% of the identified proteins are localized on the mitochondrial membrane of some pathogens such as the fungi *Aspergillus clavatus* NRRL 1, *A. flavus* NRRL3357, *C. immitis* RS, the yeast *Cryptococcus neoformans* var. grubii H99, *Fusarium graminearum* PH-1, *A. capsulatus* G186AR, *Leptomonas pyrrhocoris* H10, the food fungus *Neosartorya fischeri* NRRL 181, * Phytophthora parasitica* INRA-310 and *P. ultimum* DAOM BR144. The remaining proteins are localized on the plasma membrane (Dataset 1\(^4\) and Figure 2D).

Approximately 15% (34/234) of the folate transporters identified possess signal peptides (Figure 2E) with the trypanosomes with the most signal peptides. Deductions can be made of the probable destination within the cell of any transporter by its signal peptide sequence; thus, further work may seek to decipher the sequence of the targeting signal to determine its localization. The proteins identified all have transmembrane helices with the exception of the alveolate *Chromera velia* CCMP2878, apicomplexan *P. berghei* ANKA, *S. neurona* SN3, the kinetoplastid *T. brucei* TREVU927, *T. grayi* ANR4 and protist *Virella brassicaformis* CCM3155 with Gene ID’s Cvel_17766, PBANKA_0713700, SN3_01500005, Tb927.8.6480, Tgr.2739.1000 and Vbra_15327, respectively (Dataset 1\(^4\)).

The phylogenetic tree (Figure 3) shows the evolutionary position, history and relationship of all the folate transporters identified in this work. The type of transporter or species/strain was used for constructing phylogenetic trees, with the 234 proteins identified forming two clades, a major and minor. The major clade lacked a sub-clade, while the minor clade possessed a sub-clade. All proteins identified were distributed between the two major clades; except for folate/methotrexate transporter and mitochondrial folate transporter, with the latter present on the major clade and the former on the minor clade exclusively. All the species are represented on both clades, however, *V. brassicaformis* CCM3155, *Plasmodium* species, *A. clavatus* NRRL, *A. flavus* NRRL 3357, *A. macrognymus* ATCC 38327, *C. fasciculata* strain Cf-CI, *C. immitis* RS, *C. maris* RN66, *C. neoformans* var. grubii H99, *C. neoformans* var. grubii H99, *Leishmania* species, *N. bombycis* CQ1, *N. caninum* Liverpool, *F. graminearum* PH-1 and *H. hammondi* strain H.H.34 are exclusively on the major clade. There are some parasites that were identified once, as shown in Dataset 1\(^4\); these are mostly in the large clade. Some of these pathogens include *P. ultimum* DAOM BR144, which has mitochondrial folate transporter/carrier proteins similar to *Homo sapiens*, *E. cuniculi* GB-M1, which has proteins similar to folate transporter, and *S. punctatus* DAOM BR117, which has folate-binding protein YgFZ. These were the only proteins of the aforementioned species identified in this work. However, *M. daphniae* UGP3, which had reduced folate carrier domain containing protein, was the only parasite that was found in the small clade. Improving on our phylogenetic analysis, we performed a sub-phylogenetic reconstruction (Figure 4–Figure 6) based on the substrate type of the transport proteins. After phylogenetic analysis each sub-phylogeny show a clear characterization except for folate-biotinier transporters (Figure 5), which fell in a different clade save for *Leptomonas* species and *C. velia*.

**Discussion**

Folate transporters are important proteins involved in the salvage of folate, cofactors and related molecules in eukaryotic pathogens important for metabolism and survival in their respective hosts\(^21\). We identified proteins that could mediate the salvage of folates into cells and/or mitochondria from eukaryotic pathogen genomes in EupathDB. Many of these proteins are involved in folate biosynthesis or transport and are present in many of the eukaryotic pathogens we queried. In this study, 234 genes encoding homologues of folate salvaging proteins were identified in the genome of 64 strains, representing 28 species of eukaryotic pathogens. Some of the pathogens include *P. falciparum* 3D7 and IT, *P. knowlesi* H, *P. berghei* ANKA, *P. chabaudi* chabaudi, *T. brucei* Lister 427, *T. brucei* TREVU927, *T. brucei gambiense* DAL972, *Encephalitozoon cuniculi* GB-M1. The pathogens range from bacteria through to fungi, intracellular parasites such as *Plasmodium* and *leishmania* species, to extracellular parasites such as trypanosome species. This suggests a widespread presence of the proteins cutting across a range of pathogens that infect humans and animals.

A few of the proteins we identified have been previously identified and characterized in parasites such as *Plasmodium falciparum*\(^22,30\). Trypanosome species\(^30\), *Leishmania* species and *Toxoplasma gondii*\(^36\). It has been estimated that over half of the drugs currently on the market target integral membrane proteins of which membrane transporters are a part, but unfortunately, these transporters have not been adequately explored as drug targets\(^35\). Folate transporters therefore represent attractive drug targets for treatment of infectious diseases. Thus their identification from other eukaryotic pathogens could open a window for novel chemotherapeutics for disease control\(^12,32\).
Figure 3. Phylogenetic tree showing relatedness of folate transporters compared across all studied pathogen species. Tree was constructed using Neighbor-Joining method (Bootstrap test was a 1000 replicates).
Figure 4. Phylogenetic tree showing relatedness of proteins involved in only folate transport across all studied pathogen species. Tree was constructed using Neighbor-Joining method (Bootstrap test was a 1000 replicates).
Figure 5. Phylogenetic tree showing relatedness of proteins involved in either folate or biopterin transport across all studied pathogen species. Tree was constructed using Neighbor-Joining method (Bootstrap test was a 1000 replicates).
Figure 6. Phylogenetic tree showing relatedness of proteins involved in only folate-binding protein YgfZ transport across all studied pathogen species. Tree was constructed using Neighbor-Joining method (Bootstrap test was a 1000 replicates).
In Plasmodium two folate transporters have been identified, namely PfT1 and PfT2. These transporters have been shown to mediate the salvage of folate derivatives and precursors in P. falciparum and proposed blocking of their salvage activity may improve the antimalarial efficacy of several classes of antimalarial drugs. In our work we identified folate transporters for other plasmodial species, which, like P. falciparum, may also be chemotherapeutic targets. Transport of folate in higher eukaryotes is made possible by a high affinity folate-biotin transporter FBT or BT1 family. In the trypanosomes and related kinetoplastids, a member of these transporters, the folate biotin transporter (FBT) family of proteins was identified in Leishmania. It is thought that MFS proteins are related to the FBT. These proteins have been characterized in a few protozoa and cyanobacteria. Results from our study describing the presence of these transporters across several phyla corroborate the conservation of folate transport function among FBT family proteins from species from plants and protists. We also identified folate salvaging proteins from fungi such as Coccidioides immitis and A. clavatus, fungi found in soil, vegetable, and waters in tropical and subtropical areas. These fungi are known to occasionally become pathogenic and act as opportunistic pathogens for animals and man. Coccidioidomycosis caused by C. immitis in association with AIDS has been known to be a fatal disease. Treatment of acute and chronic infections with antifungals such as amphotericin B have not been adequate, hence folate transporters may present new targets in these group of pathogens. Identification was also made on pathogens such as C. fasciculata that parasitize several species of insects including mosquitoes and has been widely used to test new therapeutic strategies against parasitic infections. As a model organism, folate transporters identified in C. fasciculata may be useful in research for developing new drugs in medically important Kinetoplasts as has been shown for other targets in this protozoan parasite.

We noticed that P. parasitica INRA-310 and L. pyrrhocoris H10 had the highest number of folate transporters identified. Their utility as model fungal (P. parasitica) and monoxenous kinetoplast may provide models instrumental for developing new antifolates for fungal and protozoan diseases. The relatedness of these proteins across the different pathogens shows that there are two major phylogenetically distinct clades in the eukaryotic pathogens examined. The clustering of these proteins suggests that these transport proteins have highly conserved regions often required for basic cellular function or stability. Thus, antifolate chemotherapeutic drugs that are effective against one pathogen might have some effect on others.

Conclusion
In summary, we have identified and classified 234 proteins after an extensive search of pathogens genome in eukaryotic pathogen resource databases, though experimental studies will be required to confirm the expression and function of these proteins in parasites. Our results show that these proteins that mediate the transportation of folate are widely distributed in different pathogen species examined in various phyla. The identification of folate salvage proteins in diverse eukaryotes extend the evolutionary diversity of these proteins and suggests they might offer new possibilities for potential drug development targeting folate-salvaging routes in eukaryotic pathogens.

Data availability
Dataset 1: Complete list of proteins extracted from Eupthadb and literature search, including their properties. These data are available in a .xlsx file. Doi, 10.5256/f1000research.10561.d148742

Author contributions
M.O.F. and B.O. Conceptualized and Designed the study. M.O.F and B.O. structured methodology. B.O. performed analysis. M.O.F and B.O. wrote manuscript.

Competing interests
No competing interests were disclosed.

Grant information
B.O. was supported by a TWAS-CNpq fellowship (FP number: 3240274297).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
Open Peer Review

Current Peer Review Status: ?

Version 1

Reviewer Report 06 March 2017

https://doi.org/10.5256/f1000research.11380.r20419

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

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

Below are my major concerns:

1. The abstract does not provide an adequate summary of the article.

   The authors have claimed much higher scope of their work than actually reported. While their work is restricted to folate transporters, they have claimed to work on whole folate pathway. In the Abstract: "We applied a combination of bioinformatics methods to examine the genomes of pathogens in the EupathDB for genes encoding homologues of proteins that mediate folate salvage in a bid to identify and assign putative functions." "These findings offer new possibilities for potential drug development targeting folate-salvage proteins."

2. No proper justification for the work is provided.

   a) While folate pathway is well established as drug target, the authors have only identified folate transporters. So please give examples of drugs (with names) which are known to target folate transporters in any organism. If no sufficient information is provided, the usefulness of the work is severely reduced.

   b) How many of folate transporters are essential in species where essentiality data is available such as P. berghei and T. gondii?

3. The methodology to identify transporters is not comprehensive.

   Since authors have used key-word searches to identify folate transporters, they are likely to miss many transporters not labelled as such. An appropriate methodology will incorporate profile (such as HMM) based searches. Thus the number of transporters identified by authors is most likely to be an underestimate.
4. The methodology is not clear.

Authors write that "we utilized the word “folate” for search on the gene text and “folic acid" was used to confirm the hits", then how only transporters were retrieved? The Figure 1 is confusing. Where is BLAST used here?

5. The manuscript is written very poorly with so many scientific and grammatical mistakes that it is very difficult for the reader to follow the manuscript. Below are some examples.

a) "The mitochondrion is the predicted location of the majority of the proteins, with 15% possessing signal peptides." - how can mitochondria be majority location if only 15% have signal peptides and even less with mitochondrial signal peptide? Shouldn't the majority then be cytoplasmic?

b) "We identified 234 proteins to be involve in folate transport".

c) "Since folate-binding protein YgfZ, folate/pteridine transporter, folate/biopterin transporter, putative, reduced folate carrier family protein, folate/methotrexate transporter FT1, putative folate transporters alone and others have 10, 25, 132, 2, 7, 49 and 9." What are these numbers?

d) "So we decided to reconstruct the phylogeny based folate transporter, folate-biopterin transporter after considering the identification number, the species diversity in each category."

e) "The different proteins identified to be involved in folate salvage or related molecules were folatebinding protein YgfZ, folate/pteridine transporter, folate/biopterin transporter, reduced folate carrier family protein, folate/methotrexate transporter FT1 and folate transporters having a 4%, 11%, 56%, 1%, 3% and 21% identity, respectively." What does this statement mean?

f) Does Table 1 really need to be 4 page long?

g) "The only Plasmodium species with results for proteins that salvage folate was P. falciparum"

h) "However, folate transporters I and II were retrieved from our search of GeneDB for P. malariae and P. ovale curtisi, respectively." What are these transporter classes?

i) "Some of these pathogens include P. ultimum DAOM BR144, which has mitochondrial folate transporter/carryer proteins similar to Homo sapiens, E. cuniculi GB-M1, which has proteins similar to folate transporter, and S. punctatus DAOM BR117, which has folate-binding protein YgfZ."

j) "After phylogenetic analysis each sub-phylogeny show a clear characterization except for folate-biopterin transporters"

h) "In this study, 234 genes encoding homologues of folate salvaging proteins were identified in the genome of 64 strains, representing 28 species of eukaryotic pathogens. Some of the pathogens include P. falciparum 3D7 and IT, P. knowlesi H, P. bergheri ANKA, P. chabaudi chabaudi, T. brucei Lister 427, T. brucei TREU927, T. brucei gambiensis DAL972, Encephalitozoon cuniculi GB-M1. The pathogens range from bacteria through to fungi, intracellular parasites such as Plasmodium and leishmania species, to extracellular parasites such as trypanosome species" Which bacteria was included in the study?
i) "It has been estimated that over half of the drugs currently on the market target integral membrane proteins of which membrane transporters are a part, but unfortunately, these transporters have not been adequately explored as drug targets. Folate transporters therefore represent attractive drug targets for treatment of infectious diseases." Please tell us how many drugs are available in the market which target folate transporters, which is a more relevant statistic with respect to this study.

j) "In the trypanosomes and related kinetoplastids, a member of these transporters, the folate biopterin transporter (FBT) family of proteins was identified in Leishmania."

k) "It is thought that MFS proteins are related to the FBT." What is MFS?

l) "Results from our study describing the presence of these transporters across several phyla corroborate results other researches, establishing the conservation of folate transport function among FBT family proteins from species from plants and protists".

m) "The clustering of these proteins suggests that these transport proteins have highly conserved regions often required for basic cellular function or stability". The clustering does not suggest that these transport proteins have highly conserved regions often required for basic cellular function or stability.

n) " We also performed phylogenetic comparisons of identified proteins. .".

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

I confirm that I have read this submission and 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.

Reviewer Report 27 February 2017

https://doi.org/10.5256/f1000research.11380.r20535

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

**Raphael D. Isokpehi**
College of Science, Engineering and Mathematics (CSEM) , Bethune-Cookman University , Daytona Beach, FL, USA

**Summary of Referee's Report**
The manuscript presents a strong justification for research on folate transporter proteins as drug targets for diseases caused by eukaryotic pathogens including the malaria parasite. The manuscript reports a data curation effort that involves the use of the Eukaryotic Pathogens Database (EuPathDB) Resource. Several novel results to guide future research are included such as (i) a list of 234 folate transporter proteins from 63 eukaryotic microbes including eukaryotic pathogens; and (ii) phylogenetic trees of relatedness of the protein sequences. The authors observed the clustering of the protein sequences that
indicate the possibility that antifolate drugs could be effective for multiple eukaryotic pathogens.

Major concerns are:
(i) The need for a clearer description of the workflow for the construction of the protein list.
(ii) There is inadequate support for the statement that “60% of the proteins were identified for the first time”.
(iii) Confusion between retrieval and identification of protein sequences. The workflow diagram indicates retrieval of sequences but narrative text describes identification in multiple sections. The potential drug targeting categorization of the retrieved protein sequences is a key contribution of the research study.

Some minor concerns include (i) typographic errors such as spelling of abbreviations (e.g. EuPathDB, PubMed and UniProt); (ii) classification of A. (Ajellomyces) capsulatus G186AR as a bacteria; and (iii) need to quantify the statistics associated with observations (Examples: in abstract “Many of the genomes”; in discussion: “A few of the proteins”).

Title and Abstract:
Is the title appropriate for the content of the article?
The manuscript title is “Genome-wide characterization of folate transporter proteins of eukaryotic pathogens”. “Genome-wide characterization” does not effectively describe the accomplishments of the research reported. The manuscript in the conclusion section (Page 15) states “we have identified and classified 234 proteins...”. The workflow (Figure 1) provides categorization of proteins by features such as cellular location, presence of signal peptide and number of transmembrane helices. Figure 2 has the title “Categorization of proteins identified”. A suggested revised title is “Categorization of potential drug targeting folate transporter proteins from eukaryotic pathogens”. The “potential drug targeting” is obtained from the conclusion section.

Does the abstract represent a suitable summary of the work?
There are sentences in the abstract that should be revised to accurately represent a suitable summary of the research performed. Comments/suggestions are provided below.
1. Quantify the observations described as many. For example “genome sequences of many”: How many?

2. “eukaryotic protozoa” > “eukaryotic microbes”. The term “eukaryotic microbes” encompasses pathogens and non-pathogens (e.g. Chromera velia and Vitrella brassicaformis).

3. “important data” > “critical biological information”

4. “pathway are important for” > “pathways are necessary for”

5. “Methods: We applied a combination of bioinformatics methods to examine the genomes of pathogens in the EupathDB for genes encoding homologues of proteins that mediate folate salvage in a bid to identify and assign putative functions.” > “Methods: We developed automated search strategies in the Eukaryotic Pathogen Database Resources (EuPathDB) to construct a protein list and retrieve protein sequences of folate transporters encoded in the genomes of 200 eukaryotic microbes. The folate transporters were categorized according to features including mitochondrial localization, number of transmembrane helix, and protein sequence relatedness.

6. Provide key result(s) of the protein list retrieval and phylogenetic comparison of the retrieved proteins. For example, We constructed a list of 234 folate transporter proteins associated with 63
eukaryotic microbes including ??? algae, ??? fungi and ??? protozoa. Seven percent of the proteins were predicted to localize on the mitochondrial membrane. Phylogenetic tree revealed major (??? proteins) and minor (??? Proteins) clades. All the folate transporter sequences from the malaria parasite, *Plasmodium*, belonged to the major clade.

7. “The mitochondrion is the predicted location of the majority of the proteins”. This statement is not supported by Figure 2D, where 7% of the protein sequences are labelled as Mitochondrial folate transporters.

**Article content:**

**Have the design, methods and analysis of the results from the study been explained and are they appropriate for the topic being studied?**

**Design and Methods:**

1. Figure 1 presents a conceptual hierarchical methodology. The rectangle labelled “Protein names/sequences verification” has arrows to PubMed, UniProt, Membranetransporters.org, NCBI, GeneDB, Google Scholar and Phylogenetic analyses. It appears that the integrated results from the search strategies in the databases provided the input for the phylogenetic analyses. Please clarify.

2. Which step of the workflow resulted in the list of 234 proteins?

3. How many proteins were retrieved from the initial search using EuPathDB?

4. Protein Features Retrieved rectangle: Was the retrieval of protein features performed on only the 234 proteins?

5. There is adequate explanation of the methods for phylogenetic analyses. Please provide the Newick format phylogenetic tree as a supplementary dataset.

**Analysis of the Results:**

1. Table 1 is a major curation effort presented in the manuscript.
   a. The title of Table 1 should be updated to “Eukaryotic microbes from which folate transporters were identified”. The list includes non-pathogens.
   b. The content of the table (especially the Kingdom entries) should be checked for accuracy. The Kingdom column could be updated to Eukaryotic Microbe Group with entries as algae, protozoa or fungi.
   c. The column entries for *A. capsulatus* G186AR (mislabelled as bacteria) should be updated as the organism name is for a fungus (genus *Ajellomyces*). This update will also affect the Phylogenetic Tree (Figure 3). The node labelled *Actinobacillus* clusters with *Ajellomyces macrogynus*.
   d. An updated Table 1 should be presented as Dataset 2 in a spreadsheet file. This would enable secondary data analysis by other researchers.
   e. A new Table 1 could consist of columns for Eukaryotic Microbe Group, Genera of Eukaryotic Microbe, List of Species/Strain and Number of Folate Transport Proteins. This will provide reader with an overview of how the 234 proteins is distributed into the genera of the eukaryotic microbes.
   f. References listed for confirmation searches. Page 8, Paragraph 1, Sentence 1: “Our literature search for parasite folate transporters on PubMed and Google Scholar indicated 60% (38 out 63) of the proteins were identified for the first time as presented in Table 1.”
Comment: Among the references included in Table 1, only eight references (22, 73 to 77 and 82) on the basis of the article title provide experimental assessments of the folate transporters. Reference 85 is a reference for MEGA7 software. In the sentence “proteins” should be eukaryotic microbes. The proportion of eukaryotic microbes whose folate transporter(s) have been previously investigated with functional assays should be revised.

2. Figure 2. Categorization of proteins identified.
   Authors should consider representing Figure 2A and 2B as bar graphs. Figures 2C, 2D and 2E have only two categories that can be described in the Results section.

3. Discussion
   a. “genomes of 64 strains”. Table 1 has 63 eukaryotic microbes.
   b. Discuss Chromera velia and Vitrella brassicaformis as organismal systems for investigating folate transporter function. See Woo et al. 1

4. Conclusion
   Consider revising “In summary, we have identified and classified 234 proteins...” to “In summary, we have retrieved information on 234 folate transporter proteins from the Eukaryotic Pathogen Database (EuPathDB) resources. The folate transporter proteins were categorized into potential drug targeting features including mitochondrial localization, number of transmembrane helix, and protein sequence relatedness."

Data (if applicable):
Has enough information been provided to be able to replicate the experiment? Are the data in a usable format/structure and have all the data been provided?
   1. Table 1 needs to be revised and converted to a Dataset.

   2. Please provide the Newick format phylogenetic tree as a supplementary dataset.

   3. The Gene Identifiers [Gene ID] in Dataset 1 can be used to retrieve the protein sequences from EuPathDB.

References
1. Woo YH, Ansari H, Otto TD, Klinger CM, et al.: Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites.Elife. 2015; 4: e06974 PubMed Abstract | Publisher Full Text

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and 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.

The benefits of publishing with F1000Research:
• Your article is published within days, with no editorial bias
• You can publish traditional articles, null/negative results, case reports, data notes and more
• The peer review process is transparent and collaborative
• Your article is indexed in PubMed after passing peer review
• Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com