

RESEARCH ARTICLE

Genome-Wide Collation of the *Plasmodium falciparum* WDR Protein Superfamily Reveals Malarial Parasite-Specific Features

Priyanka Chahar¹, Manjeri Kaushik¹, Sarvajeet Singh Gill¹, Surendra Kumar Gakhar¹, Natrajan Gopalan², Manish Datt³, Amit Sharma³, Ritu Gill¹*

1 Centre for Biotechnology, Maharshi Dayanand University, Rohtak, Haryana, India, **2** DRDO-BU Centre for Life Sciences, Bharathiar University Campus, Coimbatore, Tamil Nadu, India, **3** Structural and Computational Biology Group, International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi, India

* ritu_gill@hotmail.com



CrossMark
click for updates

OPEN ACCESS

Citation: Chahar P, Kaushik M, Gill SS, Gakhar SK, Gopalan N, Datt M, et al. (2015) Genome-Wide Collation of the *Plasmodium falciparum* WDR Protein Superfamily Reveals Malarial Parasite-Specific Features. PLoS ONE 10(6): e0128507. doi:10.1371/journal.pone.0128507

Academic Editor: Jude Marek Przyborski, Philipps-University Marburg, GERMANY

Received: October 24, 2014

Accepted: April 29, 2015

Published: June 4, 2015

Copyright: © 2015 Chahar et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: Work on malaria molecular biology in RG's laboratory is supported by UGC and DST, Government of India, India. PC and MK acknowledge the receipt of University Research Fellowship from MD University, Rohtak. AS is supported by OSRF grant from Department of Biotechnology, Government of India, India.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Despite a significant drop in malaria deaths during the past decade, malaria continues to be one of the biggest health problems around the globe. WD40 repeats (WDRs) containing proteins comprise one of the largest and functionally diverse protein superfamily in eukaryotes, acting as scaffolds for assembling large protein complexes. In the present study, we report an extensive *in silico* analysis of the WDR gene family in human malaria parasite *Plasmodium falciparum*. Our genome-wide identification has revealed 80 putative WDR genes in *P. falciparum* (*PfWDRs*). Five distinct domain compositions were discovered in *Plasmodium* as compared to the human host. Notably, 31 *PfWDRs* were annotated/re-annotated on the basis of their orthologs in other species. Interestingly, most *PfWDRs* were larger as compared to their human homologs highlighting the presence of parasite-specific insertions. Fifteen *PfWDRs* appeared specific to the *Plasmodium* with no assigned orthologs. Expression profiling of *PfWDRs* revealed a mixture of linear and nonlinear relationships between transcriptome and proteome, and only nine *PfWDRs* were found to be stage-specific. Homology modeling identified conservation of major binding sites in *PfCAF-1* and *PfRACK*. Protein-protein interaction network analyses suggested that *PfWDRs* are highly connected proteins with ~1928 potential interactions, supporting their role as hubs in cellular networks. The present study highlights the roles and relevance of the WDR family in *P. falciparum*, and identifies unique features that lay a foundation for further experimental dissection of *PfWDRs*.

Introduction

WD40 repeats (WDRs) or WD40 domain proteins comprise one of the largest and functionally diverse protein families in eukaryotes from yeast to human. The WD40 domain is among the top ten most abundant domains in eukaryotic genomes [1]. Though rare, these proteins are

present in some prokaryotes e.g. *Thermomonospora curvata* and *Cyanobacterium synechocystis* [1,2]. The WDR family is defined by a sequence repeat of ~40 amino acids that characteristically begins with a glycine-histidine (GH) pair at N-terminus and ends with a tryptophan-aspartic acid pair (WD) at C-terminus [2]. Typically the WD40 proteins are composed of several tandem repeats of WD40 motifs [3]. The WD40 repeat motif is comprised of a four-stranded anti-parallel β -sheet (a-b-c-d) and displays only limited amino acid sequence conservation [1,2,4].

Despite significant level of sequence diversity, the WD40 repeats invariably fold into a β -propeller conformation with number of blades corresponding to the number of repeating units present [4–7]. Each WD40 β -propeller blade contains the first three strands of one repeat unit and the last strand of the previous repeat (d-a-b-c). Thus, the four strands of WD40 repeat belong to two different WD40 blades. This sharing of strand between two blades and hydrophobic interactions between the blades results in stabilization of the β -propeller structure [2,8].

The members of WDR family have been characterized to play important role in multiple cellular processes, such as RNA processing, signal transduction, vesicular trafficking, regulation of cell division, apoptosis, ubiquitination/protein degradation, chromatin assembly, remodeling and transcriptional regulation [4,9–11]. The underlying common feature of all the members of WDR family is their coordination of multiprotein complex assemblies. These repeats provide a stable platform or scaffold on which large protein or protein-DNA complexes can assemble. As determined by the available WD40 domain complex structures, the WDR proteins have three distinct surfaces (top, bottom and circumference) that can be exploited for interaction with other proteins [1,11]. The ability of this domain to interact simultaneously with a number of proteins through multiple interaction surfaces makes them important players in the cellular interaction networks. The WD40 domain is one of the top interacting domains in eukaryotic genomes. It has been identified to participate in more interaction pairs than any other domain in yeast [1]. The versatile functions of WDR family seem to be the outcome of participation in multiple protein-protein interactions [1] and fusion with additional domains i.e. multidomain architecture [3,12,13].

Peculiar features of the WDR gene family are- i) low sequence conservation although high evolutionary conservation; ii) co-occurrence of the WD40 domain with other domains; iii) interaction with multiple proteins to form large complexes and; iv) the functional diversity. Despite of the global reduction in malaria mortality by 42% from 2000 to 2012, malaria continues to be a major public health problem threatening 3.4 billion people across 97 countries and causing deaths of ~627,000 people in 2012 [14]. The WDR gene family remains uncharacterized in *Plasmodium falciparum* (the most virulent species of human malaria) except for a few individual protein studies of this family e.g. *PfSec31*, *PfSec13* and *PfRACK* [15–17]. In the present study, a comprehensive genome-wide analysis of the WDR family in *P. falciparum* has been performed in order to understand the significance, functionality and evolutionary relationships. Additionally, we have investigated the expression patterns of *PfWDRs* through different developmental stages of *Plasmodium* based on the available transcriptome and proteome datasets. Homology modeling enabled three-dimensional (3D) structure prediction and comparisons to human orthologs. We have also investigated protein-protein interactions (PPIs) networks for *PfWDRs*, and highlighted their potential significance in parasite biology.

Results and Discussion

Identification and listing of *PfWDR* genes in *P. falciparum*

To identify genes of the WDR family in *P. falciparum*; we used two approaches (Fig 1A). First, HMM search was performed using WD40 family seed downloaded from Pfam which resulted in the identification of 78 genes. Second, a keyword search 'WD40' and the InterPro domain

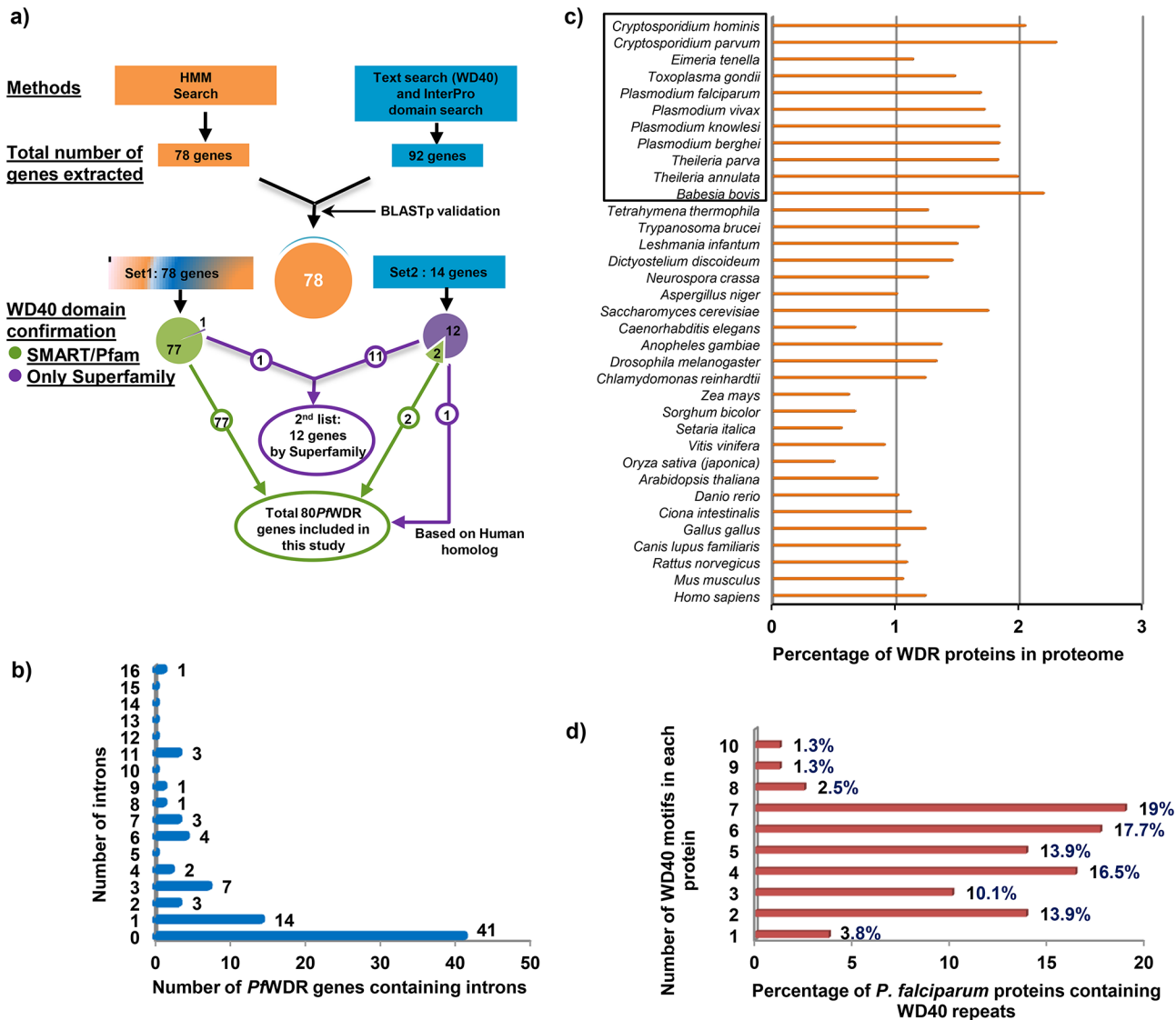


Fig 1. Extraction and characterization of the PfWDRs. a) Schematic representation of the approaches employed for the identification of PfWDR genes. b) Graphical representation of the occurrence of introns by number in PfWDR genes. c) Predicted percentage of the proteome of eukaryotic organisms devoted for the WDR proteins. Apicomplexans are boxed. d) Distribution of the WD40 motifs by number in PfWDRs.

doi:10.1371/journal.pone.0128507.g001

search (IPR017986) against PlasmoDB resulted in the extraction of 92 putative PfWDR genes. Further, BLASTp analysis with the yeast and human WD40 sequences as queries was also done to validate the above results. A careful analysis and union of both the lists resulted in the identification of 92 putative PfWDR genes. The intersection of both the lists contained 78 genes and was taken as the first set (Set1) of putative PfWDR genes (Fig 1A). Second set (Set2) composed of 14 putative PfWDR genes after subtracting first set from the total 92 putative PfWDR genes.

Subsequently, both the sets of putative PfWDR proteins were subjected to Pfam and SMART analysis to confirm the presence of WDRs. These databases confirmed the presence of WD40 repeat motifs in 79 proteins; 77 from Set 1 and 2 from Set 2 resulting in the first list of 79 putative PfWDR genes (Table A in S1 Table). In the remaining genes, WDRs were recognized by Superfamily database only and were kept in the second list of putative PfWDR genes (Table B in S1 Table) and eliminated from further analysis. However, one gene based on its

human homolog named as WDR8 was extracted from second list and included in first list resulting in the final list of 80 putative PfWDR genes (Table A in [S1 Table](#)). Presence of large number of WDR genes in *P. falciparum* is in accordance with other eukaryotic organisms, where WD40 domain has been shown to be one of the most abundant domains [1].

Further, data on CDS length, protein length, molecular weight, number of introns, and chromosome location for the putative PfWDR genes was extracted from PlasmoDB ([S1 Table](#)). Out of the 80 PfWDR genes, 39 genes (48.8%) contained introns that roughly corroborates with the reported distribution of introns in *P. falciparum* genome i.e. 53.9% of *P. falciparum* genes possess introns [18]. The number of introns in PfWDRs varied from 1 to 16 ([Fig 1B](#)) with PF3D7_1037800 (a cytoskeletal regulatory protein) having maximum number of introns. Length and molecular weight of the PfWDRs varied from 323 aa and 35.7 kDa (PF3D7_0526300, PF3D7_0826700) to 4405 aa and 526.7 kDa (PF3D7_1410300). There were large variations in the pI values, ranging from 4.37 to 9.8. These variations in pI and molecular weight of PfWDRs are comparable to the WDR proteins known in other organisms e.g. *Arabidopsis thaliana* [3], *Oryza sativa* [12], *Setaria italica* (foxtail millet) [13] and *Homo sapiens* [19] etc.

Intergenomic analysis

To perform comparative analysis of the abundance of WDR proteins in different eukaryotes, we extracted WDR proteins from several organisms as reported in various studies or as per SMART database or general text and InterPro domain search (IPR017986). There were 237 WDR proteins in *A. thaliana* [3], 200 in *O. sativa* [12], 225 in *S. italica* [13] and 267 in *H. sapiens* [19]. [Fig 1C](#) shows the percentage of WDR proteins in the proteomes of different eukaryotes ([S2 Table](#)). In comparison to plants and metazoans, apicomplexans possessed relatively higher percentage of the WDRs as per proteome size. This may be because of the participation of WDRs in basic cellular and metabolic processes and less complexity of the apicomplexans as compared to the plants and metazoans resulting in higher percentage of WDRs as per their proteome size.

Number of WDR motifs

WD40 proteins are characterized by the presence of multiple tandem repeats generally varying from 4 to 8 [2]. However, proteins with minimum 2 and maximum 16 WDRs have also been reported [3,10,20]. In our analysis, a total of 354 WDR motifs were recognized by SMART among 80 PfWDRs which range from 1–10 repeats/protein ([Fig 1D](#)). Approximately 70% of all the PfWDR proteins have 4–8 WD40 repeats. Pfam database recognized less number of WD40 proteins as well as number of repeats as compared to SMART (Table A in [S1 Table](#)) thus seems to be more stringent in the identification of members of this family.

Generation of sequence logo

To explore the characteristics and extent of sequence conservation of the WD40 repeats in *P. falciparum*, their sequence logo was generated and subsequently compared with the HMM logo of Pfam WD40 family. As per the definition of WDR superfamily, the characteristic features of the family were presence of GH dipeptide at N-terminus and WD dipeptide at C-terminus. Nevertheless, as clear in the generated logos ([Fig 2](#)), even these positions were not conserved thus making identification of the family members and all the repeats in a protein difficult by sequence comparison methods. Further, as per the Pfam logo, the most conserved positions in the WD40 superfamily were H₁₀, D₃₂ and W₃₈. These positions showed significant conservation in the PfWDRs logo but less than Pfam HMM logo, revealing more variations

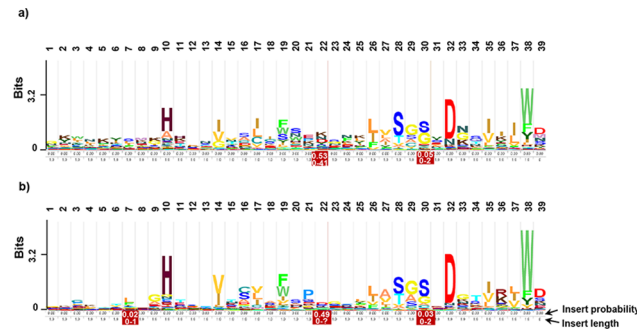


Fig 2. Sequence logos for the WD40 motif. a) The PfWDRs HMM logo based on alignment of all the identified WDRs in *P. falciparum*. b) The Pfam WD40 family HMM logo drawn from alignment of Pfam WD40 seed sequences.

doi:10.1371/journal.pone.0128507.g002

within the PfWDR superfamily at these positions. The most conserved residue in the PfWDRs was D₃₂. HMM seed of high confidence PfWDR sequences was built and has been provided as File B in [S1 File](#). This can be used to identify WDR proteins in other organisms.

Domain architecture analysis of PfWDR proteins

Domain composition analysis for the 80 PfWDR proteins as per SMART and Pfam databases revealed that 48 proteins had only WD40 domain (Class A), whereas, 32 proteins showed WD40 domain in combination with other functional domains (Class B). Class B was further divided in 21 subclasses (a-u) based on the type of additional domains present ([Fig 3A](#)). Importantly, five domain compositions of PfWDRs were found to be specific to the parasite as compared to their human host ([Fig 3B](#)). Subclass-a has one member (PF3D7_1315400) having a combination of WD40 domain with ZF_C3H1 (CCCH type zinc finger). This domain combination has not been identified in human although present in some plants, alveolates and fungi. Subclass-h has one member (PF3D7_1251200) exhibiting two unique domains of unknown functions (DUF1899 and DUF1900) interspaced by three WDRs, which is the characteristic feature of coronin gene family [21]. However, a C-terminal variable coiled coil domain responsible for oligomerization (one of the characteristic feature of short coronins) was absent in PF3D7_1251200. A medley of PX with WD40 was identified unique to *Plasmodium* as compared to its human host (subclass m- PF3D7_0704400). The PX-WD40 fusion is conserved in *Plasmodium* species, apicomplexans and restricted to only alveolates.

PF3D7_1333600 having a combination of collagen, transmembrane domain and coiled coils with three WD40 motifs was kept in subclass-p. This domain architecture is unique to *Plasmodium* although combination of WD40+collagen is present in other organisms. Another subclass-s (PF3D7_1329100) featured the presence of N-terminal myosin motor domain (head domain), multiple IQ motifs (neck domain) and coiled coils, which are usually associated with the myosin heavy chains [22]. The combination of WD40 with the N-terminal coiled coils in variable tail domain of myosin heavy chains is exclusively present in the apicomplexans. Subclass-t has one member (PF3D7_1410300) with PbH1 (parallel beta helix repeats) domain and coiled coils in addition to the WD40 domain. This domain composition is absent in human.

A comparison of the domain organization of WDRs in *P. falciparum*, *H. sapiens* (as per SMART database) and *O. sativa* [12] is illustrated in [Fig 4](#). Our results showed that ~40% and ~48% WDRs in *P. falciparum* and *H. sapiens* have multidomain combinations, respectively. However, in *O. sativa* ~27% WDRs were reported to be associated with additional domains [12]. A fusion with additional domains may contribute to the functional diversity of WDRs.

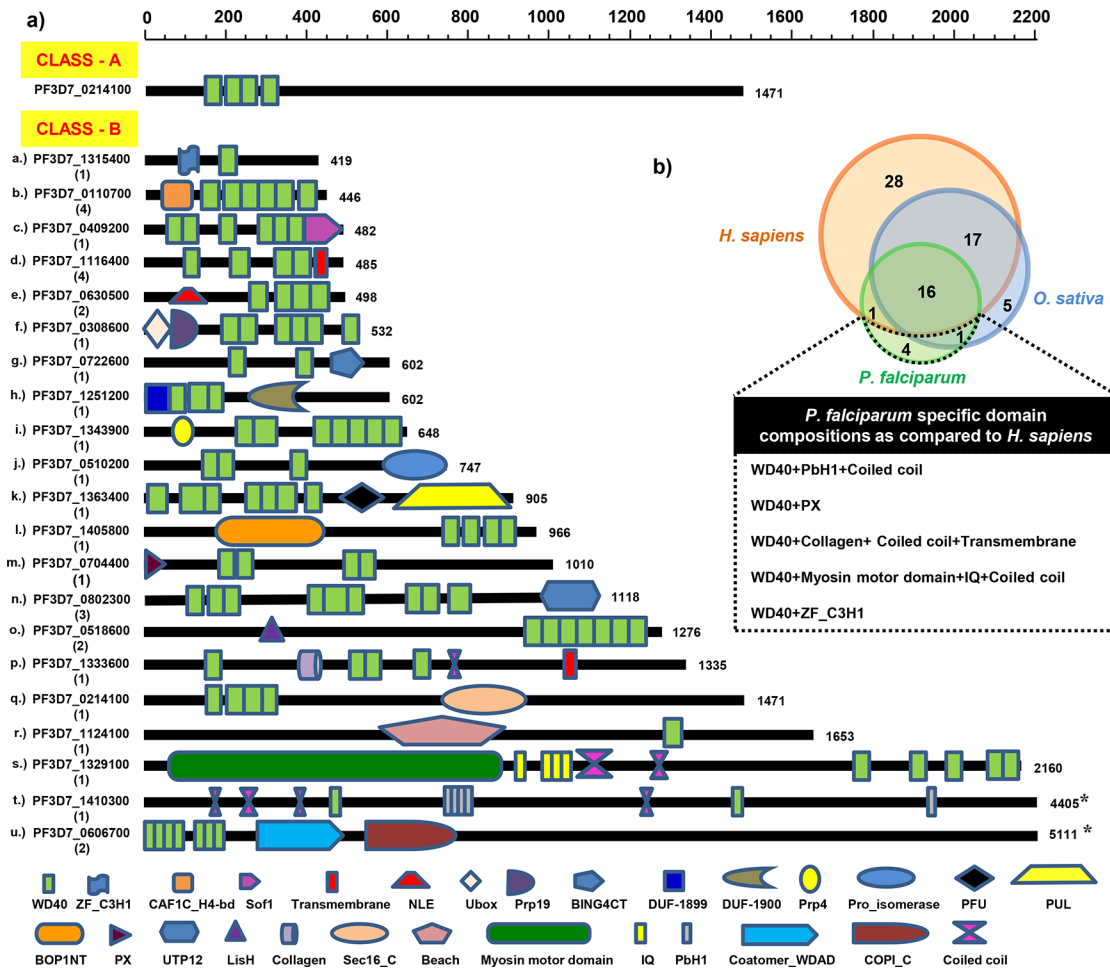


Fig 3. Domain organization of the PfWDRs a) Domain organization of the representative PfWDRs from each class and subclass based on the identification of WD40 and other additional domains by SMART, Pfam and InterPro. Subclass name and gene ID for each protein are given on the left and the number of members of each subclass is given in parenthesis. Domain positions are scaled according to the protein length bar given at the top except gene IDs marked with asterisk (*) b) Venn diagram depicting the number of shared and specific domain combinations of *P. falciparum*, *H. sapiens* and *O. sativa* WDR proteins. Inset table enlists the *P. falciparum* specific domain compositions as compared to the human host.

doi:10.1371/journal.pone.0128507.g003

Functional insights based on orthologs of PfWDRs

To investigate the putative functions of PfWDRs, we analyzed the cellular function of each listed member by exploring gene annotations in PlasmoDB and annotations of orthologs at UniProt aided by domain composition analysis and published articles. Ascribing putative roles to *P. falciparum* proteins based on homology is challenging due to lack of sufficient sequence similarity to the characterized genes in other organisms. This is because of the extremely AT-rich *P. falciparum* genome that results in an unusual amino acid composition, presence of large insertions and low complexity regions [18,23].

Orthologs for the PfWDRs were predicted using BLAST/PSI-BLAST search as mentioned in the methods and compiled in S3 and S4 Tables. Importantly, while assigning orthologs we considered retention of key domains, comparable protein sizes and annotation of the reference proteins. Notably, it was not always the first hit in BLAST that has been assigned as a homolog. Sometimes, no significant hit was observed directly in *H. sapiens*; however, homolog was assigned through BLAST searches against other organisms. For example, HsWDR12 showed a

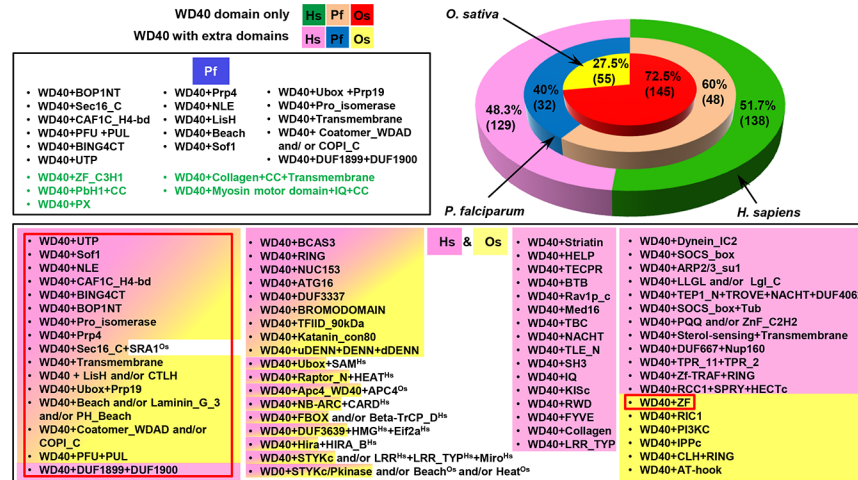


Fig 4. Comparison of percentage occurrence of the WDR proteins containing WD40 domain alone or in combination with additional domains in *H. sapiens* (Hs), *P. falciparum* (Pf) and *O. sativa* (Os) [12] shown by stacked pie diagram. *P. falciparum* specific domain compositions as compared to *H. sapiens* are highlighted in the green text. Domain compositions in pink, yellow and pink-yellow shades are of human, rice and shared human-rice, respectively as per SMART database. Superscripts 'Hs' and 'Os' represent domains specifically present in human and rice. Domain compositions of human and rice shared with *P. falciparum* are enclosed in the red box.

doi:10.1371/journal.pone.0128507.g004

weak hit against PF3D7_0630500 in BLAST (S3 Table), whereas, its yeast ortholog Ytm1 is a significant first hit (79% sequence coverage, 24% identity and 6e-18).

We were able to assign human orthologs to 61 out of 80 PfWDRs (S3 Table) though assignment of orthologs based on sequence similarity searches was poorer. Notably, most of the PfWDRs (90%) were of longer length as compared to their human orthologs highlighting the presence of species-specific insertions in *Plasmodium*. Further, the length of three PfWDRs was almost twice of their assigned orthologs (S3 Table) that may be because of the fusion of two proteins or large Pf specific insertions e.g. PfSec13 is a combination between Sec13 and Nup145C of yeast [16]. No human ortholog could be assigned to PF3D7_1333600 that has been named as U3 snoRNA-associated small subunit rRNA processing protein. Moreover, this annotation was not conserved amongst *Plasmodium* species. Brehelin et al. [24] annotated the protein based on guilt-by-association principle and assigned its yeast ortholog Utp4. Accordingly, we assigned its human ortholog Utp4; however, there was variation in length and domain structure of PF3D7_1333600 and its assigned human ortholog (S3 Table).

Orthologs for all the PfWDRs were also searched in apicomplexans *Babesia bovis* and *Toxoplasma gondii*. Except for the 17 PfWDRs, we could identify orthologs for all the others either in *B. bovis*/ *T. gondii* or both (S4 Table). Further, for the 19 PfWDRs (with no human orthologs) orthologs were searched in other *Plasmodium* species, apicomplexans, alveolates, yeast, *Drosophila* and *Arabidopsis*. This analyses revealed 1 apicomplexan specific, 2 alveolates specific, 1 alveolata and plants specific, 14 *Plasmodium* genus specific and 1 *P. falciparum* and *Plasmodium reichenowi* specific WDR proteins (Table B in S4 Table).

By homology based inter-species annotation transfer, we were able to annotate/reannotate 31 *P. falciparum* proteins which were previously unannotated or misannotated or having less refined annotation (Table 1). Further, PfWDRs were manually categorized into 13 different generic groups (Fig 5, S3 Table). A total of 16 PfWDRs, for which no function could be predicted, were kept in unknown function category. The largest fraction of PfWDRs was predicted to be involved in the RNA processing. Functional diversity of the PfWDRs is in accordance with

Table 1. List of PfWDRs with suggested annotation/refined reannotation based on orthologs.

Gene ID	Annotation at PlasmoDB	Human orthologs	Suggested reannotation
PF3D7_1243800	Microtubule associated katanin, putative	NP_079498.2-WD repeat containing protein 82 (YKL018-COMPASS component SWD2)	<i>PfSWD2</i> or <i>PfWDR82</i>
PF3D7_1347000	G-beta repeat protein, putative	NP_612467.1-WD repeat containing protein 92/Monad	<i>PfWDR92</i> or <i>Monad</i>
PF3D7_0518600	WD-repeat protein, putative	NP_079436.3-WD repeat containing protein 26 (YCL039W-Glucose-induced degradation protein 7 (GID7))	<i>PfWDR26</i>
PF3D7_1467200	Conserved Plasmodium protein, unknown function	NP_060551.1-Telomerase Cajal body protein 1 (TCAB1)/WD repeat containing protein 79	<i>PfTCAB1</i> or <i>PfWDR79</i>
PF3D7_0608000	Conserved Plasmodium protein, unknown function	NP_620133.1-Diphthamide biosynthesis protein 7 (DPH7)/WD repeat containing protein 85 (YBR246W-Diphthamide biosynthesis protein 7)	<i>PfDPH7</i> or <i>PfWDR85</i>
PF3D7_1118800	Conserved Plasmodium protein, unknown function	NP_006400.2-Actin-related protein 2/3 complex subunit 1A (YBR234C-Actin-related protein 2/3 complex subunit 1)	Actin-related protein 2/3 complex subunit 1A
PF3D7_0510800	Conserved Plasmodium protein, unknown function	NP_078808.3-Sperm associated antigen 16 protein/Pf20 protein homolog	Pf20 protein homolog
PF3D7_1348700	Conserved Plasmodium protein, unknown function	NP_659491.4-Cilia and flagella associated protein 52 (CFAP52)/WD repeat-containing protein 16	CFAP52 or <i>PfWDR16</i>
PF3D7_1406500	Conserved Plasmodium protein, unknown function	NP_689711.1-WD repeat containing protein 65 (Trypanosome homolog: Tb927.8.4870- DIGIT)	<i>PfDIGIT</i>
PF3D7_1033500	WD-repeat protein, putative	NP_060504.1-WD repeat containing protein 70 (<i>C. elegans</i> homolog: O16519-Gastrulation defective protein 1 (GAD-1))	<i>PfWDR70</i> or <i>PfGAD-1</i>
PF3D7_1105200	Conserved Plasmodium protein, unknown function	NP_060288.2-WD repeat containing protein WRAP73/WDR8	<i>PfWDR8</i>
PF3D7_1221600	Conserved Plasmodium protein, unknown function	NP_003301.1-Tumor-suppressing STF cDNA 1 protein (TSSC1)	<i>PfTSSC1</i>
PF3D7_1428400	Probable protein, unknown function	NP_055838.2-WD and tetratricopeptide repeats protein 1 (WDTC1)	<i>PfWDTC1</i>
PF3D7_0409200	40S ribosomal processing protein, putative	NP_056235.3-DDB1 and CUL4 associated factor 13 (DCAF13) (YLL011W*-nucleolar RNA-associated protein SOF1)	<i>PfDCAF13</i> or Protein SOF1
PF3D7_1237600	rRNA processing WD-repeat protein, putative	NP_008993.1-Periodic tryptophan protein 1 homolog (PWP1) (YLR196W*-Periodic tryptophan protein 1)	<i>PfPWP1</i>
PF3D7_1405800	Large subunit rRNA processing protein, putative	NP_056016.1-Ribosome biogenesis protein BOP1 (YMR049C*-Ribosome biogenesis protein ERB1)	<i>PfBOP1</i> or <i>PfERB1</i>
PF3D7_0816000	Nucleolar preribosomal assembly protein, putative	NP_113673.2-Glutamate-rich WD repeat-containing protein1 (GRWD1) (YMR131C*-Ribosome assembly protein RRB1)	<i>PfRRB1</i> or <i>PfGRWD1</i>
PF3D7_1146000	Nucleolar preribosomal assembly protein, putative	NP_060566.2-Notchless protein homolog 1 (NLE1) (YCR072C-Ribosome assembly protein 4 (RSA4))	<i>PfNLE1</i> or <i>PfRSA4</i>
PF3D7_0802300	rRNA processing WD-repeat protein, putative	NP_005040.2-Periodic tryptophan protein 2 homolog (PWP2) (YCR057C*-Periodic tryptophan protein 2 (UTP1))	<i>PfUTP1</i>
PF3D7_1333600	U3 snoRNA-associated small subunit rRNA processing associated protein, putative	NP_116219.1-Cirhin/UTP4 (YDR324C*-U3 small nucleolar RNA-associated protein 4)	<i>PfUTP4</i> [#]
PF3D7_0722600	Nucleolar rRNA processing protein, putative	NP_005443.3-WD repeat containing protein 46 (YER082C*-U3 small nucleolar RNA-associated protein 7 (UTP7))	<i>PfUTP7</i>
PF3D7_1448000	U3 snoRNA-associated small subunit rRNA processing protein, putative	NP_006775.1-WD repeat containing protein 3 (YLR129W*-U3 small nucleolar RNA-associated protein 12 (UTP12)/DIP2)	<i>PfUTP12</i>
PF3D7_1013100	U3 snoRNA-associated small subunit rRNA processing protein, putative	NP_006444.2-Transducin beta-like protein 3 (TBL3) (YLR222C*-U3 small nucleolar RNA-associated protein 13 (UTP13))	<i>PfUTP13</i>
PF3D7_1352200	Conserved Plasmodium protein, unknown function	NP_115551.2-U3 small nucleolar RNA-associated protein 15 homolog (UTP15)	<i>PfUTP15</i>
PF3D7_1357700	U3 snoRNA-associated small subunit rRNA processing protein, putative	NP_644810.1-WD repeat containing protein 36 (YLR409C*-U3 small nucleolar RNA-associated protein 21 (UTP21))	<i>PfUTP21</i>

(Continued)

Table 1. (Continued)

Gene ID	Annotation at PlasmoDB	Human orthologs	Suggested reannotation
PF3D7_1226700	Conserved Plasmodium protein, unknown function	NP_004695.1-RNA U3 small nucleolar interacting protein 2/ RRP9 homolog (YPR137W-Ribosomal RNA-processing protein 9 (RRP9))	<i>PfRRP9</i>
PF3D7_0630500	Microtubule-associated protein ytm1 homologue, putative	NP_060726.3-Ribosome biogenesis protein WDR12 (YOR272W-Ribosome biogenesis protein YTM1)	Ribosome biogenesis protein YTM1 or <i>PfWDR12</i>
PF3D7_0801500	Conserved Plasmodium protein, unknown function	NP_079170.2-Nucleolar protein10 (NOL10) (YGR145W-Ribosome biogenesis protein ENP2)	<i>PfNOL10</i> or Ribosome biogenesis protein ENP2 homologue
PF3D7_1220100	Pre-mRNA splicing factor, putative	NP_056975.1-Pre-mRNA-processing factor 17 (PRP17) (YDR364C-Pre-mRNA-processing factor 17)	<i>PfPRP17</i>
PF3D7_0302000	Golgi organization and biogenesis factor, putative	NP_002660.1-Pleiotropic regulator 1 (PLRG1) (YPL151C-Pre-mRNA-splicing factor PRP46)	<i>PfPRP46</i> Pre-mRNA-splicing factor
PF3D7_1241100	Conserved Plasmodium protein, unknown function	NP_060853.3-Pre-mRNA 3' end processing protein WDR33 (YNL317W-Polyadenylation factor subunit 2 (PFS2))	<i>PfWDR33</i> or <i>PfPFS2</i>
PF3D7_0905600	Conserved Plasmodium protein, unknown function	NP_653269.3-WD repeat containing protein 66 isoform 1	<i>PfWDR66</i>
PF3D7_1329100	Myosin C (MyoC)	-	Myosin F [^]

Saccharomyces cerevisiae/Trypanosoma brucei/Caenorhabditis elegans orthologs are mentioned (in brackets in column 3) where annotations are based on these.

'*' indicates orthologs/reannotation also suggested by Brehelin et al. [24] or Ochoa et al. [68].

'#' and '^' indicate orthologs/annotations given by Brehelin et al. [24] and Foth et al. [22] respectively.

doi:10.1371/journal.pone.0128507.t001

other eukaryotes, where WDRs were shown to be involved in a wide range of cellular functions [4,10].

Subcellular targeting of PfWDRs

Functional diversity of the PfWDRs necessitates their localization in various cellular compartments. Accordingly, PfWDRs were predicted to localize in various subcellular compartments

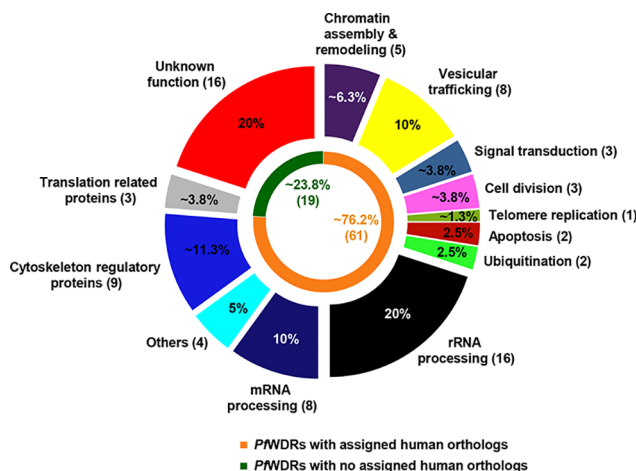


Fig 5. Pie chart representing the functional classification of PfWDRs and the number of assigned human orthologs (S3 Table). The PfWDRs are categorized into 13 functional classes. Number of proteins assigned to each class (given in parenthesis) and their percent to the total number of identified PfWDRs are indicated. Inner pie chart represents percent and number of PfWDRs with assigned human orthologs.

doi:10.1371/journal.pone.0128507.g005

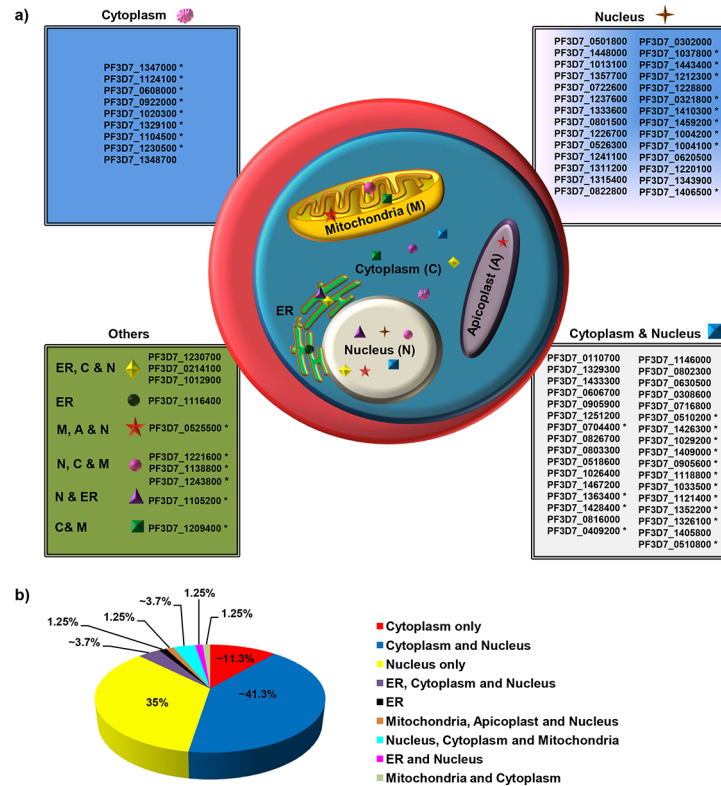


Fig 6. Predicted subcellular localization of the PfWDRs (S5 Table). a) A schematic representation of the subcellular localization of PfWDR proteins based on online programs and literature review. Abbreviations are as follows: C, cytoplasm; N, nucleus; ER, endoplasmic reticulum; M, mitochondria; A, apicoplast. b) Percentage predicted distribution of PfWDRs in different organelles within the protozoan parasite. Localization of gene IDs marked with asterisk was predicted *in silico*.

doi:10.1371/journal.pone.0128507.g006

viz. cytosol, nucleolus, nucleus, endoplasmic reticulum (ER), mitochondria and apicoplast (Fig 6). Subcellular localization was predicted on the basis of available experimental evidence/literature review either for PfWDR or its ortholog and where no supporting literature was available, prediction was based on bioinformatics tools as mentioned in the methods. Most PfWDRs (86.3%) were predicted to reside in the nucleus, which is in agreement with their major role in RNA processing, chromatin assembly and remodeling. Out of these, ~35% PfWDRs were exclusively present in the nucleus. Interestingly, subcellular targeting of most (15/17) of the Plasmodium specific conserved proteins with no assigned ortholog has been predicted to be nuclear by various programs.

Apicoplast and mitochondria are the two organelles with extra-chromosomal DNA in Plasmodium. None of the PfWDRs were predicted to localize in the apicoplast. However, PF3D7_0525500 was predicted to possess apicoplast targeting transit peptide but the protein lacks signal peptide. Five PfWDRs possessed a potential mitochondrion-targeting signal sequence as defined by Mitoprot. Further, no PfWDR was found to have PEXEL motif. Importantly, 52.5% PfWDRs were predicted to localize in more than one compartment, thus arguing for their multiple roles. Accuracy of various bioinformatics tools in predicting subcellular localization of PfWDRs seems to be low in comparison to localization inferred from direct experimentation either for the ortholog proteins or PfWDR proteins (S5 Table). For example, in comparison to experimental data (either for PfWDRs or their orthologs), Euk-mPLoc and NetNES mispredicted localization (mislocalized or not able to predict main organelle) of ~61%

and ~48% PfWDRs, respectively. Low precision of the subcellular localization prediction for PfWDRs by different programs suggests the presence of divergent signal sequences in *Plasmodium* as compared to other eukaryotes.

Expression profiles of PfWDRs at mRNA level

To examine the expression profiling of PfWDRs, we took advantage of the extensive transcriptome and proteome data available at PlasmoDB. First, we analyzed the transcriptome data from Llinas/Derisi et al. [25] and Le Roch/Winzeler et al. [26]. Derisi group represented the transcriptome profiling of intraerythrocytic developmental cycle (IDC) at one hour time resolution with 53 time points for *P. falciparum* 3D7 dataset. While, Winzeler group examined nine different stages: seven periodic erythrocytic asexual stage time points i.e. early and late rings (ER & LR), trophozoites (ET & LT), schizonts (ES & LS) and merozoites (M) with IDC synchronized by two different methods; sorbitol and temperature, the sexual stage gametocytes (G) and mosquito salivary gland sporozoites (Sp).

A phaseogram of the PfWDRs was constructed by arranging Derisi IDC transcriptome dataset according to the phase of gene expression and compared with accordingly arranged Winzeler's transcriptome heatmap produced from log₂ ratio of RMA expression value to average RMA value for all the time points for a gene (Fig 7). Overall there was good agreement between the two data sets with respect to PfWDRs. As per both transcriptome studies, most of the PfWDRs' transcripts were upregulated in R and T stage with only few PfWDRs upregulated in S stage (Fig 7). As per Winzeler dataset, some PfWDRs were found to be preferentially expressed in G and/or Sp and/or M and low expression level throughout IDC stages R, T and S e.g. most of cytoskeleton regulatory genes (PF3D7_0510800, PF3D7_0922000, PF3D7_1406500, PF3D7_1348700, PF3D7_1426300, PF3D7_1037800), two cell division genes (PF3D7_1026400, PF3D7_1105200) and some conserved proteins unknown function (PF3D7_1104500, PF3D7_1121400). Transcripts of these genes were either not detected in Derisi dataset or detected in only few time points probably because of low transcript abundance throughout IDC.

To highlight the differential transcript abundance of the PfWDRs, we reorganized Winzeler dataset in four clusters from low to high transcript abundance (S1 Fig). Of note, functionally co-related PfWDRs e.g. those involved in rRNA processing and mRNA processing were found to be co-expressed in-phase (Fig 7) in order to ensure the presence of all the products to carry out a particular function.

Proteome profiles of PfWDRs and comparison with transcriptome data

Next, we addressed the correlation between the transcriptome profiling and actual protein expression of PfWDRs by compiling the proteome data from 11 different studies available at PlasmoDB and labelled as **a** to **k** covering 7 different developmental stages of complex *Plasmodium* life cycle (Fig 7) as follows: (**a**) Florens et al. [27], (**b&c**) Lasonder et al. [28,29], (**d**) Le Roch et al. [30], (**e**) Khan et al. [31], (**f**) Silvestrini et al. [32], (**g**) Oehring et al. [33], (**h**) Linder et al. [34], (**i**) Solyakov et al. [35], (**j**) Treeck et al. [36] and (**k**) Pease et al. [37]. A total of 77 PfWDRs were identified by all the different proteome studies compiled here, which is almost equal to the total number of identified transcripts. The most comprehensive proteome analysis of PfWDRs was provided by study (**k**) identifying 58 PfWDRs in asexual stages (R,T,S).

Most of the PfWDRs classified as cytoskeleton regulatory proteins were G and Sp specific as per their functional classification (Fig 7). These PfWDRs showed low transcript expression profile throughout IDC with little or significant upregulation either in S and/or M and/or G and/or Sp (Fig 7, S1 Fig) indicating well coordination in transcriptome and proteome Myosin F (PF3D7_1329100) was observed to have good mRNA expression throughout IDC, G and Sp.

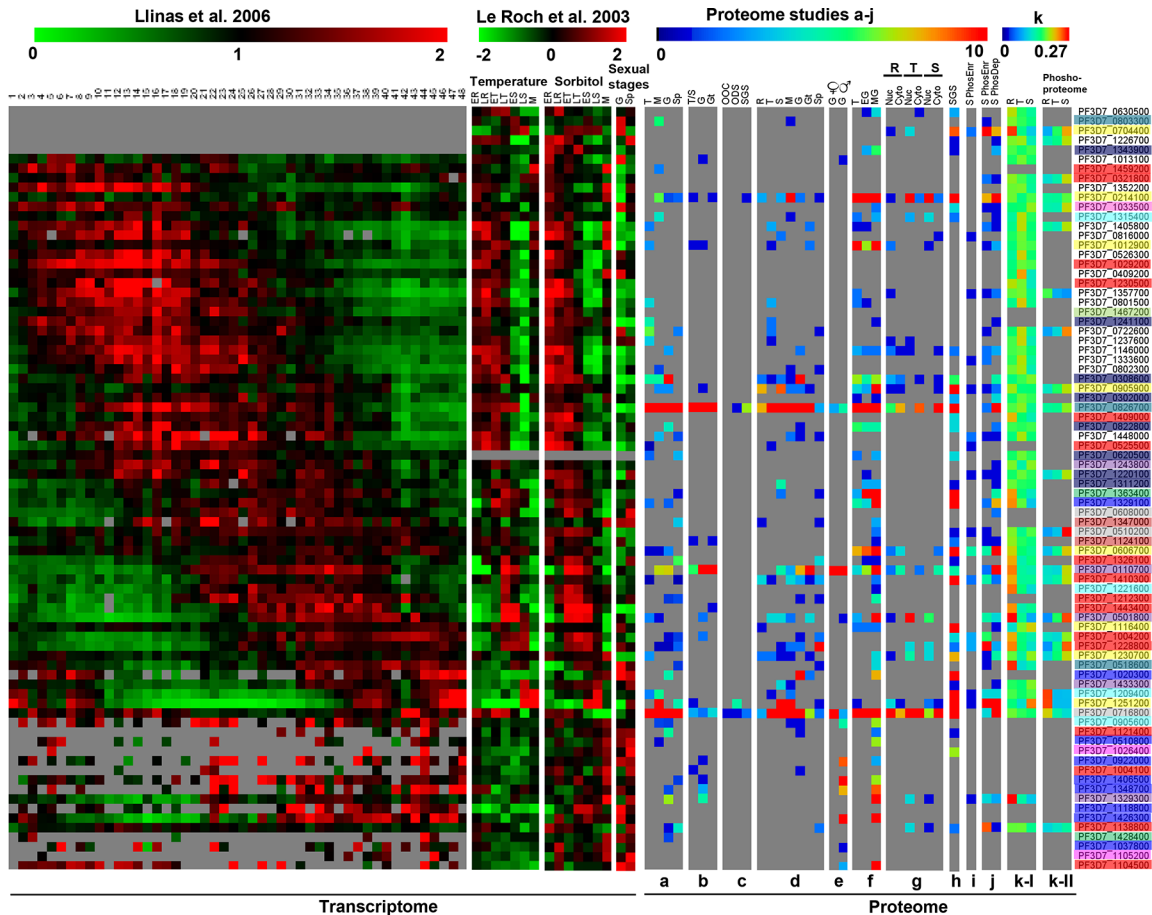


Fig 7. Expression patterns of the PfWDR genes during life cycle of the parasite. A PfWDRs phaseogram from microarray data of Linas/Derisi et al. [25] was generated covering IDC (1–48h) and compared with Le Roch/Winzeler et al. [26] data from two independently synchronized *P. falciparum* 3D7 cultures in the temperature and sorbitol covering IDC stages (R,T,S,M) as well as G and Sp. Colorimetric representation used for heat maps of transcriptome data is green-red (green, low expression; black, medium expression; red, high expression). Heat map panels at the right side with blue-red colour scale (blue, low expression; red, high expression) represent comparison of proteome and phosphoproteome data obtained from (a) Florens et al. [27], (b & c) Lasonder et al. [28,29], (d) Le Roch et al. [30], (e) Khan et al. [31], (f) Silvestrini et al. [32], (g) Oehring et al. [33], (h) Linder et al. [34], (i) Solyakov et al. [35], (j) Treeck et al. [36] and (k-I & k-II) Pease et al. [37]. Column to the right indicates PlasmoDB gene IDs of PfWDRs coloured according to the functional classification (see Fig 5). Different life cycle stages are abbreviated as: ER and LR, early and late rings; ET and LT, early and late trophozoites; ES and LS, early and late schizonts; M, merozoites; G, gametocytes; Sp, sporozoites; Gt, gamete; EG, early gametocyte; MG, mature gametocyte; OOC, oocyst; ODS, oocyst derived sporozoites; SGS, salivary gland sporozoites; phosEnr, phospho-enriched; phosDep, phospho-depleted; Nuc, nuclear; and cyto, cytoplasmic. Grey colour represents absence of detection.

doi:10.1371/journal.pone.0128507.g007

Accordingly, its protein was found in R, T, S, G, Sp. PF3D7_1026400-cell division cycle protein 20 (CDC 20) homolog has been shown to be highly upregulated in Sp at mRNA level and its protein has been detected in only Sp by report (h). However, Guttery et al. [38] showed that *P. berghei* CDC20 was highly expressed at mRNA and protein level in male G and also present throughout the life cycle of malarial parasite with no upregulation in Sp. PF3D7_1104500 showed upregulation of mRNA in G by Winzeler dataset and accordingly its protein was detected in male G by reports (e) & (f). The gene remained unannotated, however, BLAST analysis showed *HsPOC1* as a noteworthy hit (S3 Table) though coiled coil domain (characteristic feature of POCs) seems to be absent in *P. falciparum*. The POC1 proteins are the constituents of centriole having a major role in formation of cilia and flagella. Nevertheless, a detailed phylogenetic analysis of the Poc1 proteins showed absence of these proteins in *Plasmodium* [39].

The PfWDRs involved in rRNA and mRNA processing in accordance with their function were found to be present throughout IDC- R,T,S and some were also detected in G, Sp and M by various proteome studies. Transcripts of all rRNA processing genes were upregulated from R to T stage and the relative abundance of their proteins was higher in T except for PF3D7_0630500 as shown by study (k). As per study (g) out of 15 predicted rRNA processing PfWDRs, only three (PF3D7_1357700, PF3D7_1237600, PF3D7_1146000) were detected to be nuclear. This could be due to either less efficient methods of extraction of nuclear proteome or low abundance of protein to be detected by mass spectrometry.

Despite overall coordination, significant contradictions between different proteome studies have also been noted. PF3D7_0525500 protein was detected in T by studies a, d and in S by study i. However, other studies (b, g and k) including analysis of both these stages did not detect any peptide belonging to this protein. Similarly, mRNA level of PF3D7_0608000 was upregulated in T and Sp. However, its protein was detected in S only by study j, whereas, none of the other proteome study detected its presence, pointing towards lack of coordination between various proteome studies as well as nonlinear relationship between transcript and protein.

Only a few PfWDRs were stage specific i.e. confined to one or two stages (S2 Fig). Nine PfWDRs were present in only one stage and six PfWDRs were restricted to two stages. Thus, except few PfWDRs, all are resident of multiple life cycle stages highlighting their role in basic cellular and molecular processes throughout the life cycle of *Plasmodium*. PfWDRs restricted to single stage were either cytoskeleton regulatory proteins or conserved proteins with unknown function that may be playing parasite stage specific roles in cell division, invasion, motility etc. Out of 9 PfWDRs confined to a single stage, 7 PfWDRs transcripts were identified to be in coordination with proteome (S2 Fig). No correlation was seen between transcripts and proteome abundance in the PfWDRs restricted to two stages except PF3D7_1121400. In addition, some highly abundant PfWDRs transcripts showed high protein expression (PF3D7_0826700, PF3D7_0716800, PF3D7_0214100) (Fig 7, S1 Fig). However, it is not true for all and disagreement is quite common between level of transcript and protein abundance (PF3D7_1029200, PF3D7_1352200, PF3D7_1230700, PF3D7_0110700).

Thus, here we present detailed expression analyses of the PfWDRs revealing a mixture of linear and nonlinear relationships between the transcriptome and proteome. Nonlinear relationships highlight presence of regulatory mechanisms at transcript as well as protein level. Literature revealed only a handful of studies so far that draw relationships between global or gene family transcriptome and proteome in *Plasmodium* [40–43]. Our efforts to illustrate the degree of correlation between the PfWDRs transcriptome and proteome are significant. The study highlights importance of future experimental efforts that aim at dissecting regulation mechanisms in *Plasmodium*.

Genomic localization of PfWDRs

The physical map positions of 80 PfWDR genes on 14 chromosomes of *P. falciparum* were identified using genome browser at PlasmoDB and accordingly chromosomal localization map was constructed (Fig 8). The PfWDRs were found to be widely distributed over 14 chromosomes. No gene was found to be localized at the end of chromosomes i.e. telomeric positions, where only highly variable gene families (var, rif and stevor) are known to be localized [18]. Most of the PfWDR genes were located over chromosome 5 to 14. Few chromosomes had a relatively high density of PfWDR genes i.e. a maximum of 12 genes were present on chromosome 12 and 13, followed by 11 genes on the longest chromosome 14.

Genes having coordinated expression and/or similar function have been shown to cluster in the *Plasmodium* genome [27]. Genomic clustering of common function genes can facilitate

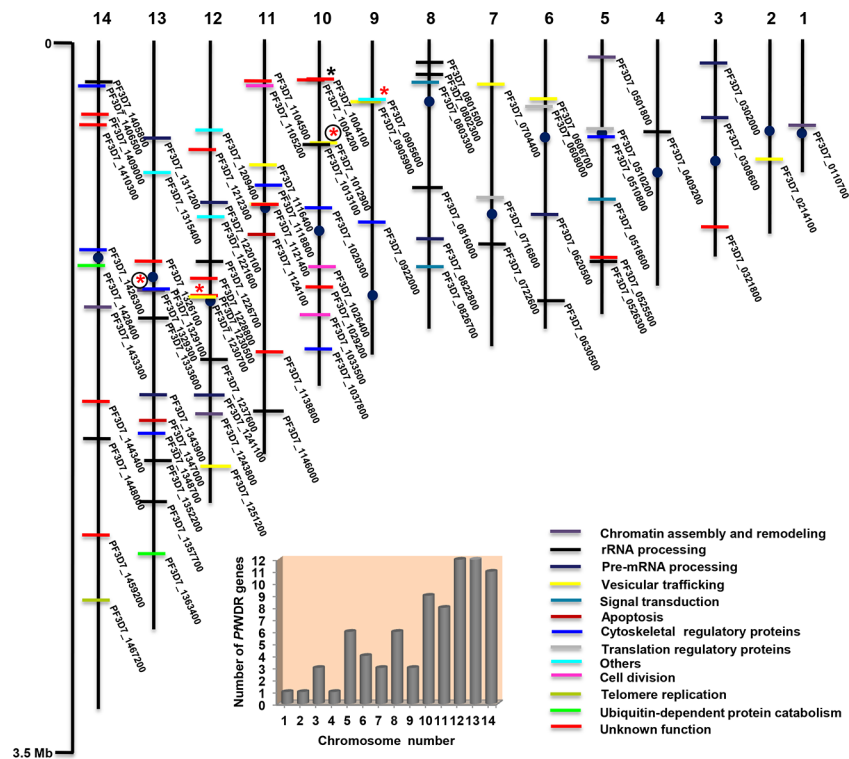


Fig 8. Physical mapping of PfWDR genes depicting their genomic localization onto 14 chromosomes of *P. falciparum*. Positions of centromeres are represented by filled circles on the chromosomes (vertical bars). Integers at the top of each bar indicate chromosome number. Grouped genes, adjacent genes and genes leaving one or two gene positions in between; are highlighted with black and red asterisks, respectively. Further, asterisks for clusters having co-expressed genes are encircled. Genes on chromosomes are colour coded as per their functional classification (see Fig 5). The scale on the left is in megabases (Mb). Number of PfWDR genes per chromosome is also shown in the graph.

doi:10.1371/journal.pone.0128507.g008

regulation of several genes simultaneously [44,45]. The chromosome localization map was further analyzed to determine whether genomic organization of PfWDR genes were in accordance with their functions. However, no real clusters of similar function PfWDRs were observed except one cluster of two genes (PF3D7_1004100, PF3D7_1004200) on chromosome number 10 (Fig 8). Further, four clusters of PfWDRs leaving one/two positions in between (marked by red asterisks, Fig 8) were also present. Out of these, two were coexpressed with their proteins detected in similar stages. All PfWDRs were found to be syntenic with other *Plasmodium* species except two (PF3D7_1004100, PF3D7_1428400) whose orthologs could not be detected in other *Plasmodium* species except *P. reichenowi* indicating towards intrasyntenic indels.

Phylogenetic analysis

To explore the evolutionary relationships of the PfWDR genes, an un-rooted neighbor-joining (NJ) phylogenetic tree was generated from the alignments of full-length protein sequences as mentioned in the methods. For statistical reliability, we conducted bootstrap analysis with 100 replicates. Nevertheless, phylogenetic relationships were not much clear with very poor bootstrap values especially in internal nodes (Fig 9).

Low boot strap values were also evident in the phylogenetic analysis of foxtail millet [13]. This may be because of divergence in the WDRs resulting from variations in length, number and position of repeats, sequence conservation and domain composition; thus making it

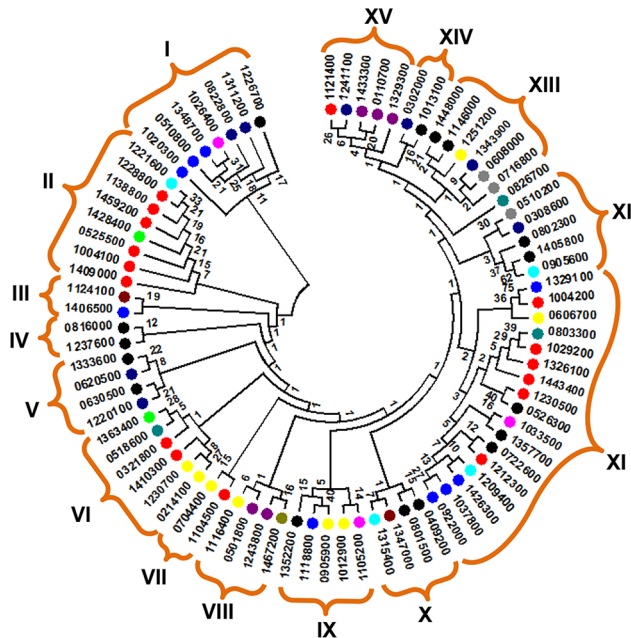


Fig 9. Phylogenetic relationships of the PfWDRs. An un-rooted NJ tree was built using Phylip with 100 bootstrap replicates and visualized by MEGA5.2. Number at the nodes represents bootstrap values. PlasmoDB gene IDs are shown by last seven digits only along with distinctly coloured circles representing functional categories of each PfWDR (see Fig 5).

doi:10.1371/journal.pone.0128507.g009

difficult to draw phylogenetic relationships. However, in outer nodes WDRs have better resolution. Accordingly, we divided the PfWDRs in 15 distinct groups. Cluster XI was the largest with 17 members. Phylogenetic analysis of the WDRs of rice, *Arabidopsis* and foxtail millet divided these proteins in 5 clusters [3,12,13]. As depicted in the Fig 9, PfWDRs revealed some clustering of similar function genes. For example, cluster I was mainly composed of RNA processing and cytoskeleton regulatory proteins of approximately similar length and number of WD40 repeats. Cluster II contains 6 *Plasmodium* specific WDRs with no assigned orthologs. Three chromatin assembly factors were grouped together in cluster XV. Cluster IV and V showed grouping of PfWDRs involved in RNA processing. While most of the cytoskeleton regulatory PfWDRs were clustered in groups I and XI.

Homology modeling

We employed the homology modeling approach to predict 3D structures for all PfWDRs by Phyre2 or Swiss-model. Structures for only 23 PfWDRs could be modeled at >90% confidence covering 80–100% residues (Fig 10, S3 Fig, S6 Table). Stereo-chemical qualities of the generated protein models were evaluated using RAMPAGE showing ~84–98% residues in allowed regions of the Ramachandran plot. As expected, 3D structures of most of the PfWDRs primarily have β -propeller structures composed of β -sheets (Fig 10). Importantly, modeled structures revealed more WD40 repeats as compared to the number of WD40 repeats predicted either by SMART, Pfam or HMM. This trend was also reported by the solved crystal structures of WDR proteins [1,8] revealing more WD40 repeats as compared to WD40 repeats identified by sequence-based algorithms which may be due to the poor level of sequence conservation. Further, we could observe insertions in the PfWDRs as compared to their human orthologs. However these insertions, as observed in PfCAF-1 (PF3D7_0110700), PfRACK

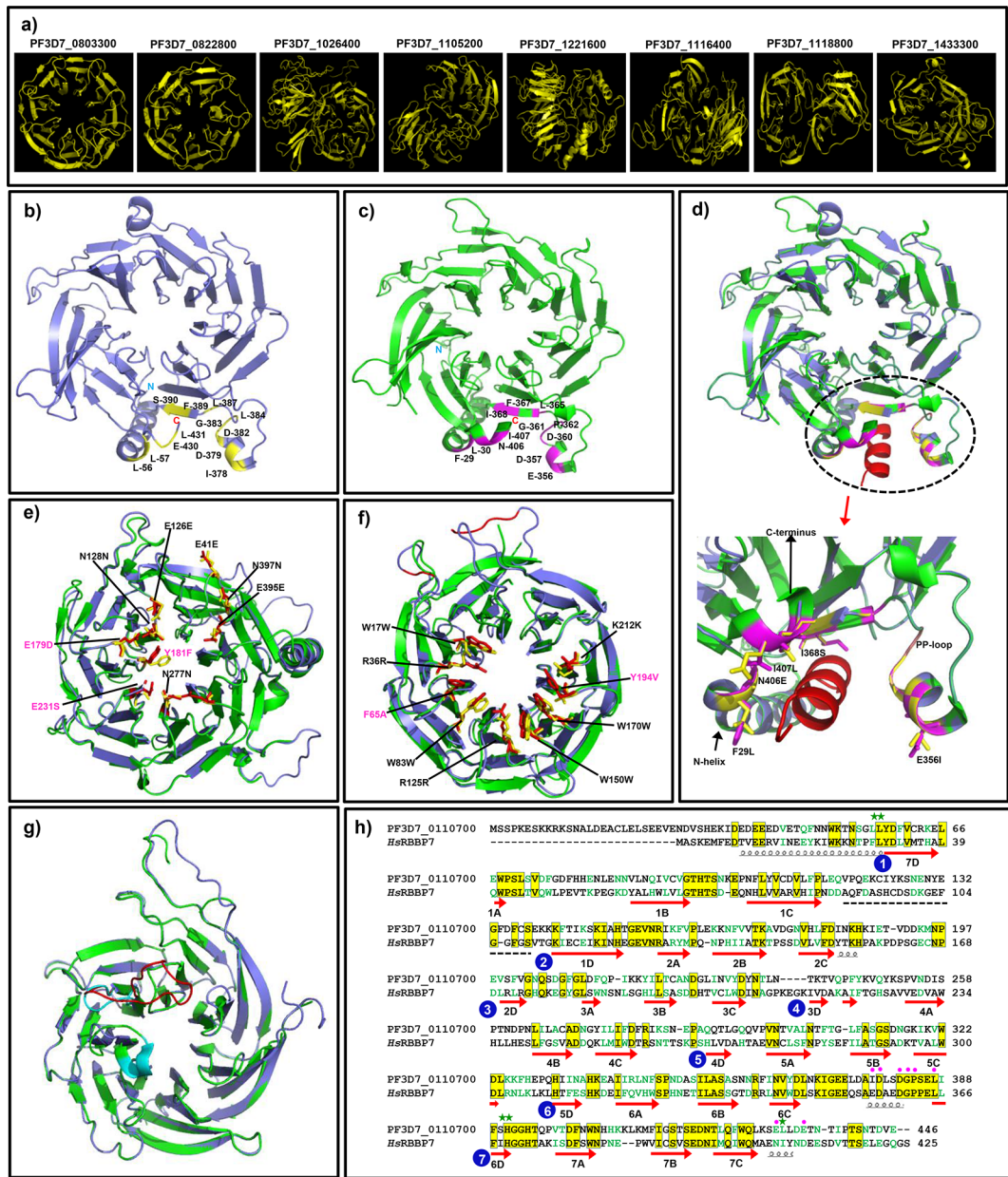


Fig 10. Structural analysis of the PfWDRs. a) Predicted structures of the 8 PfWDRs by homology modeling with >90% confidence level and $\geq 95\%$ residues in the allowed region of Ramachandran plot b) Predicted 3D structure of PfCAF-1 subunit (PF3D7_0110700) depicting histone H4 binding residues (yellow) as inferred from its human homolog RBBP4/RBBP7. c) Structure of HsRBBP7 [5] [PDB: 3CFV, green] highlighting H4 binding residues (magenta). d) Superimposition of structures of PfCAF-1 subunit and HsRBBP7 with histone H4 peptide (red) clearly showing overlapping histone H4 binding pockets (highlighted in dotted circle). Close-up view of overlapped histone binding pockets is also shown depicting variant residues (highlighted as sticks) of PfCAF-1 in comparison to HsRBBP7. Residues position in the figure are according to HsRBBP7 e.g. F29L represents Phe at 29th position of HsRBBP7 is replaced by Leu in PfCAF-1. e) Overlay of PfCAF-1 model (light blue) and HsRBBP4 [6] crystal structure [PDB: 2XU7, green] highlighting FOG-1 binding residues as yellow and red sticks respectively. Residues position scheme as mentioned above. f) Structural alignment of 3D model of PfRACK (PF3D7_0826700-light blue) and HsRACK1 [7] [PDB: 4AOW, green]. The residues of hydrophobic ring important in binding to protein ligands at the top surface of propeller structure are shown as yellow and red sticks for PfRACK and HsRACK1 [7], respectively. Insertions in PfRACK are highlighted in red that mainly lie in the loop regions. g) Overlay of predicted model of PfWDR92 (PF3D7_1347000-light blue) with the crystal structure of HsWDR92/Monad [PDB: 3I2N, green] comparing loops with insertion in *P. falciparum* i.e. Pf long loop (red) and Hs short loop (cyan). h) A structure based sequence alignment between PfCAF-1 and HsRBBP7. Secondary structure elements of HsRBBP7 are shown below the alignment indicated by coils, arrows and gaps for helices, β -strands and loops, respectively [5]. Green star and magenta boxes above the alignment indicate key residues involved in hydrophobic and hydrophilic interactions with histone H4, respectively [5]. Conserved residues are highlighted in yellow boxes while similar residues are highlighted in green text. Black dotted line below alignment indicates sequence part for which no structure is available.

doi:10.1371/journal.pone.0128507.g010

(PF3D7_0826700) and *PfWDR92* (PF3D7_1347000) were either at N-terminus, C-terminus and loop regions that do not disturb the overall secondary structures (Fig 10D–10H).

The modeled structure of *PfCAF-1* was compared to its human homolog RBBP7 (RbAp46)/RBBP4 (RbAp48) (Fig 10B–10E). The structure of *PfCAF-1* (Fig 10B) showed seven bladed β -propeller conformation in agreement with *HsRBBP7* (Fig 10C); however, β strands 1A and 7A in blades 1 and 7 were found missing. Residues responsible for binding to the histone H4 in *PfCAF1* and *HsRBBP7* are shown in the Fig 10B and 10C. As illustrated in the Fig 10D, model of *PfCAF-1* overlaps with the structure of *HsRBBP7* with conservation in the histone binding pocket present between N-terminal α -helix and PP loop (negatively charged loop which terminates in two proline) [5]. The subtle differences (F29L, E356I, I368S, I407L and N406E) in the conserved histone binding pocket are highlighted in Fig 10D. Further, FOG-1 transcription factor binding site in *HsRBBP4* [6] was also found to be conserved in *PfCAF-1* except some differences (E179D, Y181F and E231S) (Fig 10E).

Likewise, we also compared modeled structure of *PfRACK* with the crystal structure of *HsRACK1* [7]. The presence of hydrophobic ring (important in binding to protein ligands) on the top surface of *PfRACK* showed good agreement with *HsRACK1* but few differences were also evident (Fig 10F). The superimposed structures of *PfWDR92* and *HsWDR92* revealed insertions in the loop region between blades (Fig 10G). Thus, the comparison of *PfWDRs* and *HsWDRs* revealed subtle parasite-specific structural features apparent as partial conservation of important binding residues and insertions in loop regions in *PfWDRs*.

Protein-protein interactions

The unifying role of most WDRs is simultaneous or sequential binding to other proteins. We mined the PPIs for the 80 *PfWDR* proteins on the basis of experiment, text mining and database evidences available at the STRING database, Y2H datasets available at PlasmoDB [46] and individual experimental interaction evidence. Sixty out of 80 *PfWDR* proteins were associated with at least one other protein comprising a total of 1928 PPIs depicted in Fig 11A (S7 Table). Out of these, 1847 PPIs were derived from STRING, 60 by Y2H and 25 from co-immunoprecipitation (co-IP) of HA tagged *PfSec13* [16] with four overlaps i.e. one common between STRING and Y2H and three common between STRING and co-IP. The extent of connectivity differs among the 60 *PfWDRs* with PPIs ranging from 1 (PF3D7_1347000, PF3D7_1221600) to 110 (PF3D7_0308600) (S7 Table). Five *PfWDRs* have less than 5 interacting partners, 12 have 5–10 partners and 43 proteins were highly connected with more than 11 partners, suggesting that these proteins were involved in complex cellular networks. The confidence score of interactions derived from STRING database ranged from 0.4 (medium) to 0.9 (high) with 653 associations (35.4%) having high confidence scores ($S > 0.7$) and 1194 associations (64.6%) having medium confidence scores ($0.4 \leq S < 0.7$). As per the Y2H data, only 18 *PfWDRs* were involved in interactions resulting in 60 PPIs ranging from 1 to 17. Nevertheless, interactions shown by Y2H do not overlap with the interactions predicted by STRING except one.

Statistical GO enrichment of the *PfWDRs* associations using BiNGO, revealed over-representation of 138 GO ontology terms ($p < 0.05$) (S4 Fig, S8 Table) emphasizing the involvement of *PfWDRs* in many basic cellular, molecular and biological processes. Fig 11B shows the PPI network of 5 *PfWDRs* involved in chromatin remodeling/ chromatin associated processes as derived from STRING with a confidence score > 0.3 . The interaction network was composed of 203 interactions. Amongst these, PF3D7_0501800 has maximum of 83 interactions. As anticipated, the highlight of this sub-network is involvement of 129 interactions with chromatin/transcription factors/DNA binding proteins. All the 5 *PfWDRs* were found to interact with one or more histones which further authenticate their role in chromatin associated processes.

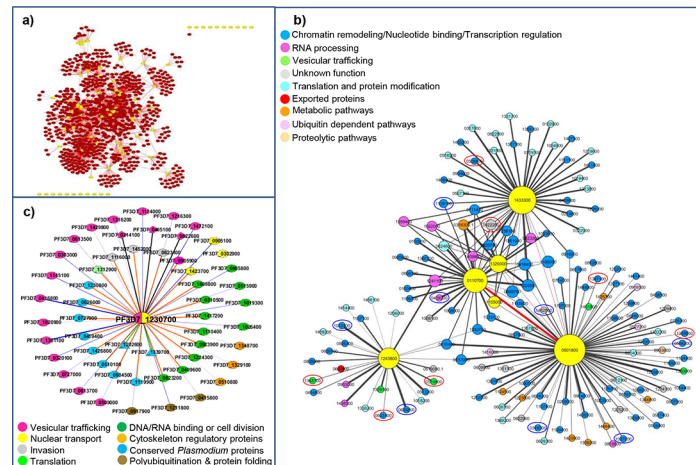


Fig 11. Protein-protein interactions (PPIs) network analysis for the PfWDRs. a) PPIs network of all the 80 PfWDRs (yellow nodes). b) PPIs network of the PfWDRs predicted to be engaged in chromatin assembly and remodeling (yellow nodes). Node size is proportional to the degree of node. Nodes are coloured according to their functional classification based on PlasmoDB/human homologs annotations. Edge width is proportional to the confidence score from STRING for each interaction. Interactions among PfWDR proteins are highlighted with red edges. Nodes not coexpressed even at a single stage with the PfWDRs are encircled in red. The nodes for which no protein expression data was available at PlasmoDB are encircled in blue colour. c) PPIs network of PfSec13 (PF3D7_1230700) (yellow and magenta node) derived from STRING (outer ring with blue edges), co-IP [16] (inner ring with orange edges) and Y2H (triangles with green edges). Interactions common between STRING and co-IP are indicated by diamond shapes and black edges. Nodes are colour coded as per their functions.

doi:10.1371/journal.pone.0128507.g011

PPIs for PfSec13, a protein identified as component of nuclear pore complex and COPII coat, are depicted in Fig 11C. Interaction network was composed of 54 PPIs derived from STRING (30 PPIs), Y2H (2PPIs) and HA tag co-IP (25PPIs) as performed by Dahan-Pasternak et al. [16]. In this network only 3 PPIs overlapped between STRING and co-IP datasets; questioning the reliability of predictions of the STRING database in case of *Plasmodium*. Y2H dataset detected only 2 PPIs that too non-overlapping with other two datasets. This points towards Y2H assays drawbacks in terms of sensitivity and specificity. STRING predicted 7 high scoring interactions ($S > 0.7$) for PfSec13 out of these 3 overlaps with co-IP. These 3 overlapping interacting proteins belong to coat complex of COPII vesicles i.e. sec23, sec24-like and sec31. A functional comparison of STRING PPIs and co-IP PPIs of PfSec13 is shown in Fig 11C.

Conclusions

In the present study, an extensive analysis of the PfWDRs in terms of their domain attributes, functional classification, genomic and subcellular localization, transcript profiles, protein expression, evolutionary relations, sequence features, homology modeling and interaction networks was performed. This study led to the identification of 80 putative PfWDRs. Analysis of the Pfam and *P. falciparum* WDR logos highlighted poor sequence conservation of the WD40 motif. A larger fraction of *P. falciparum* proteome is devoted to WDRs as compared to humans. Of note, we have identified 5 distinct PfWDRs with no clear human counterparts in terms of their domain structures. Importantly, assignment of orthologs to the PfWDRs helped us to annotate and predict potential functions of several PfWDRs. Interestingly, most of the assigned human orthologs were of shorter length as compared to their respective PfWDRs underscoring the presence of residue insertions in *Plasmodium* proteins. The phylogenetic analysis hints at divergence in PfWDRs as no clear evolutionary relationships could be drawn within

the PfWDR superfamily. Proteome profiling revealed presence of most of the PfWDRs in multiple stages of *Plasmodium* life cycle except for 9 PfWDRs that are restricted to single stage either of S, M, G, Sp. Expression profiling disclosed a blend of linear and nonlinear correlations between mRNA and protein existence as well as abundance between various *Plasmodium* life cycle stages. Our efforts to draw relationships between transcriptome and proteome profiles of the PfWDRs are an important addition to a handful of similar studies [40–43]. Our analyses of the modeled PfWDR 3D structures highlighted slight deviations in highly conserved binding sites and presence of insertions mainly in loop regions. PPI network analysis suggested the involvement of PfWDRs in a large number of interactions. In summary, the present efforts to identify and describe key attributes of this uncharacterized WDR family in *P. falciparum* provide a foundation for dissection of their regulatory roles in parasite biology.

Methods

Extraction of putative PfWDR genes and domain composition analysis

Two approaches were used for the mining of *Plasmodium* genomic resource database PlasmoDBv9.0 (<http://plasmodb.org/plasmo/>) [47] to obtain all the putative PfWDR genes. Firstly, PlasmoDB was explored by text search using keyword 'WD40' and InterPro domain search 'IPR001680: WD40_repeat, IPR017986: WD40_repeat_dom' [48]. Secondly, HMM search was performed with the WD40 domain HMM seed (PF00400) downloaded from Pfam [49] against *P. falciparum* protein database using HMMER3.0 program [50]. To confirm the presence of WD40 domain in each of the predicted PfWDR, Pfam and SMART databases [19] were explored. Domain architecture for the PfWDRs was drawn manually as identified by Pfam, SMART and InterPro. Sequence logos were generated using Skylign tool [51].

Ortholog and functional assignments

Each PfWDR protein sequence was queried against UniProtKB or non redundant protein sequence databases using NCBI BLAST/PSI-BLAST as well as against organism specific databases. The hits were explored manually for the assignment of putative orthologs based on a number of parameters i.e. e-value and score, sequence coverage, percentage identity, the length of the hit as compared to the query, domain composition, UniProt annotation for the protein, any functional knowledge and available literature review. Additionally, OrthoMCL [52] and PhylomeDB [53] were also queried for the identification of putative orthologs.

Functional classification of the PfWDRs was done manually based on their annotation in PlasmoDB and conservation of this annotation in various *Plasmodium* species and/or annotation of their orthologs at UniProt or their respective databases, related literature either for PfWDR or its homolog or for a particular functional annotation/category with a further assistance from Gene Ontology database (www.geneontology.org).

Subcellular localization and expression profiling of PfWDRs

To predict subcellular location of the PfWDRs, firstly, experimental evidences either for PfWDR or its ortholog through ApiLoc v3 (<http://apiloc.biochem.unimelb.edu.au>) and published articles were searched. Additionally, various online servers like MitoProt [54], Euk-mPloc 2.0 server [55], PATS [56], PlasmoAP [57], PSORT II [58] and NetNES [59] were also explored. Evaluation of the transcriptomic data [25,26] and proteomic data [27–37] of PfWDRs obtained from literature was done using MeV (version 4.9) software. For Le Roch et al. [26] transcriptome data, K-means clustering was performed with MeV using 4 classes and Euclidian distance of genes.

Phylogenetic analysis and homology modeling of PfWDRs

Multiple sequence alignment of the full length PfWDRs was carried out using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). Subsequently, an un-rooted NJ tree was constructed using the program Phylip v3.695 [60] with standard parameters and visualized by MEGA v5.2 [61].

3D structures of PfWDRs were predicted by homology modeling servers Phyre2 [62] and Swiss-model [63]. Structures were visualized with PYMOL [64]. Structural validation of the protein models was done using RAMPAGE [65] and QMEAN server available at Swiss- model [63].

Network data analysis

The entire set of protein-protein interactions for 80 PfWDRs was extracted from the STRING database [66] having a confidence score (S) in a range of 0.4 to 0.999 based on experiment, text mining and database evidences. In addition, Y2H datasets [46] and pull downs and any other experimental interaction data of PfWDRs through literature were also explored. Undirected weighted graph with a single edge between any pair of proteins weighed by the S value was generated in Cytoscape [67]. BinGO application within Cytoscape was employed to identify enriched GO terms using hypergeometric test with Benjamini and Hochberg FDR correction at 0.05 significance level. Y2H interaction data was obtained from PlasmoDB.

Supporting Information

S1 Fig. Clustering of Le Roch et al. [26] data. The data was clustered in four groups from low to high expression (blue-yellow colorimetric representation) using Mev 4.9. Different stages of IDC are as in Fig 7. Gene IDs on right side are coloured according to their functional classification (see Fig 5).

(TIFF)

S2 Fig. Expression patterns of stage specific PfWDR genes confined to one or two stages. Figure legend is same as of Fig 7. Genes having coordination between transcriptome and proteome are marked with asterisk.

(TIF)

S3 Fig. Predicted structures of the 12 PfWDRs by homology modeling with >90% confidence level and 85–94% residues in the allowed region of Ramachandran plot.

(TIF)

S4 Fig. Graphical representation of Gene Ontology analysis based on molecular, cellular and biological processes using BiNGO. The size of node is proportional to the number of proteins represented by GO term. The colour represents the enrichment significance (p-value) for each GO term while white nodes are not enriched and represents the hierarchical relationship among enriched members.

(TIFF)

S1 File. PfWDRs sequences alignment (.sto file) (File A) and seed HMM file (.hmm file) (File B).

(ZIP)

S1 Table. List of putative PfWDR genes. List of 80 putative PfWDR genes confirmed by SMART and/or Pfam databases (Table A). List of 12 PfWDR genes in which WD40 domain is confirmed by Superfamily database only (Table B). Information regarding product description,

gene location, nucleotide sequence length, number of introns, isoelectric point, molecular weight and amino acid sequence length as extracted from PlasmoDB.

(XLSX)

S2 Table. Data file for the graph in Fig 1C showing the predicted percentage of proteome of eukaryotic organisms devoted for the WDR proteins.

(DOCX)

S3 Table. Functional classification and human orthologs of the PfWDRs. Domains specific to *P. falciparum* and *H. sapiens* are indicated with superscript 'Pf' and 'Hs'. Orthologs in other organisms are given in '['] where human homologs were traced based on these.

(DOC)

S4 Table. Orthologs of the PfWDRs. List of *B. bovis* and *T. gondii* orthologs of 61 PfWDRs (with assigned human homologs) (Table A). Orthologs of 19 PfWDRs (with no assigned human homolog) in various organisms i.e. *Plasmodium vivax*, *P. berghei*, *Tetrahymena thermophila*, *B. bovis*, *T. gondii*, *Cryptosporidium parvum*, *Theileria annulata*, *S. cerevisiae*, *Drosophila melanogaster*, *A. thaliana*, *H. sapiens* (Table B). Doubtful orthologs are given in grey text.

(XLSX)

S5 Table. Predicted subcellular localization of the PfWDRs derived from various online programs and published sources (either for PfWDRs or assigned orthologs). Localization of gene IDs marked with asterisk was predicted *in silico*.

(XLSX)

S6 Table. Characteristics of the homology modeled 3D structures of 23 PfWDRs.

(DOCX)

S7 Table. PfWDRs protein-protein associations derived from STRING database, Y2H interaction data and literature.

(XLSX)

S8 Table. BinGO results for over-represented GO functional categories.

(XLSX)

Acknowledgments

The authors acknowledge the help of Dr. M. Jain, NIPGR, New Delhi and Dr. H. Dhekne, Vanderbilt University Medical Center, Nashville, Tennessee, USA for critical reading of the MS.

Author Contributions

Conceived and designed the experiments: RG AS SSG NG. Performed the experiments: RG PC MK MD. Analyzed the data: RG PC MK MD. Wrote the paper: RG PC AS SSG SKG.

References

1. Stirnimann CU, Petsalaki E, Russell RB, Muller CW (2010) WD40 proteins propel cellular networks. *Trends Biochem Sci* 35: 565–574. doi: [10.1016/j.tibs.2010.04.003](https://doi.org/10.1016/j.tibs.2010.04.003) PMID: [20451393](https://pubmed.ncbi.nlm.nih.gov/20451393/)
2. Neer EJ, Schmidt CJ, Nambudripad R, Smith TF (1994) The ancient regulatory-protein family of WD-repeat proteins. *Nature* 371: 297–300. PMID: [8090199](https://pubmed.ncbi.nlm.nih.gov/8090199/)
3. van Nocker S, Ludwig P (2003) The WD-repeat protein superfamily in Arabidopsis: conservation and divergence in structure and function. *BMC Genomics* 4: 50. PMID: [14672542](https://pubmed.ncbi.nlm.nih.gov/14672542/)
4. Xu C, Min J (2011) Structure and function of WD40 domain proteins. *Protein Cell* 2: 202–214. doi: [10.1007/s13238-011-1018-1](https://doi.org/10.1007/s13238-011-1018-1) PMID: [21468892](https://pubmed.ncbi.nlm.nih.gov/21468892/)

5. Murzina NV, Pei XY, Zhang W, Sparkes M, Vicente-Garcia J, Pratap JV, et al. (2008) Structural basis for the recognition of histone H4 by the histone-chaperone RbAp46. *Structure* 16: 1077–1085. doi: [10.1016/j.str.2008.05.006](https://doi.org/10.1016/j.str.2008.05.006) PMID: [18571423](https://pubmed.ncbi.nlm.nih.gov/18571423/)
6. Lejon S, Thong SY, Murthy A, AlQarni S, Murzina NV, Blobel GA, et al. (2011) Insights into association of the NuRD complex with FOG-1 from the crystal structure of an RbAp48- FOG-1 complex. *J Biol Chem* 286: 1196–1203. doi: [10.1074/jbc.M110.195842](https://doi.org/10.1074/jbc.M110.195842) PMID: [21047798](https://pubmed.ncbi.nlm.nih.gov/21047798/)
7. Ruiz Carrillo D, Chandrasekaran R, Nilsson M, Comvik T, Liew CW, Tan SM, et al. (2012) Structure of human Rack1 protein at a resolution of 2.45 Å. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 68: 867–872. doi: [10.1107/S1744309112027480](https://doi.org/10.1107/S1744309112027480) PMID: [22869111](https://pubmed.ncbi.nlm.nih.gov/22869111/)
8. Hudson AM, Cooley L (2008) Phylogenetic, structural and functional relationships between WD- and Kelch-repeat proteins. *Subcell Biochem* 48: 6–19. doi: [10.1007/978-0-387-09595-0_2](https://doi.org/10.1007/978-0-387-09595-0_2) PMID: [18925367](https://pubmed.ncbi.nlm.nih.gov/18925367/)
9. Smith TF, Gaitatzes C, Saxena K, Neer EJ (1999) The WD repeat: a common architecture for diverse functions. *Trends Biochem Sci* 24: 181–185. PMID: [10322433](https://pubmed.ncbi.nlm.nih.gov/10322433/)
10. Li D, Roberts R (2001) WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. *Cell Mol Life Sci* 58: 2085–97. PMID: [11814058](https://pubmed.ncbi.nlm.nih.gov/11814058/)
11. Chen CK, Chan NL, Wang AH (2011) The many blades of the beta-propeller proteins: Conserved but versatile. *Trends Biochem Sci* 36: 553–561. doi: [10.1016/j.tibs.2011.07.004](https://doi.org/10.1016/j.tibs.2011.07.004) PMID: [21924917](https://pubmed.ncbi.nlm.nih.gov/21924917/)
12. Ouyang Y, Huang X, Lu Z, Yao J (2012) Genomic survey, expression profile and co-expression network analysis of OsWD40 family in rice. *BMC Genomics* 13: 100. doi: [10.1186/1471-2164-13-100](https://doi.org/10.1186/1471-2164-13-100) PMID: [22429805](https://pubmed.ncbi.nlm.nih.gov/22429805/)
13. Mishra AK, Muthamilarasan M, Khan Y, Parida SK, Prasad M (2014) Genome-wide investigation and expression analyses of WD40 protein family in the model plant foxtail millet (*Setaria italica* L.). *PLoS One* 9: e86852. doi: [10.1371/journal.pone.0086852](https://doi.org/10.1371/journal.pone.0086852) PMID: [24466268](https://pubmed.ncbi.nlm.nih.gov/24466268/)
14. World Health Organization (2013) World malaria report 2013. Geneva: World Health Organization.
15. Adisa A, Albano FR, Reeder J, Foley M, Tilley L (2001) Evidence for a role for a *Plasmodium falciparum* homologue of Sec31p in the export of proteins to the surface of malaria parasite-infected erythrocytes. *J Cell Sci* 114: 3377–3386. PMID: [11591825](https://pubmed.ncbi.nlm.nih.gov/11591825/)
16. Dahan-Pasternak N, Nasereddin A, Kolevzon N, Peer M, Wong W, Shinder V, et al. (2013) PfSec13 is an unusual chromatin-associated nucleoporin of *Plasmodium falciparum* that is essential for parasite proliferation in human erythrocytes. *J Cell Sci* 126: 3055–3069. doi: [10.1242/jcs.122119](https://doi.org/10.1242/jcs.122119) PMID: [23687383](https://pubmed.ncbi.nlm.nih.gov/23687383/)
17. Sartorelloa R, Amayab MJ, Nathansonb MH, Garciac CRS (2009) The *Plasmodium* receptor for activated C kinase protein inhibits Ca²⁺ signaling in mammalian cells. *Biochem Biophys Res Commun* 389: 586–592. doi: [10.1016/j.bbrc.2009.09.025](https://doi.org/10.1016/j.bbrc.2009.09.025) PMID: [19748487](https://pubmed.ncbi.nlm.nih.gov/19748487/)
18. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, et al. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419: 498–511. PMID: [12368864](https://pubmed.ncbi.nlm.nih.gov/12368864/)
19. Letunic I, Doerks T, Bork P (2012) SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res* 40: D302–D305. doi: [10.1093/nar/gkr931](https://doi.org/10.1093/nar/gkr931) PMID: [22053084](https://pubmed.ncbi.nlm.nih.gov/22053084/)
20. Rorick NK, Kinoshita A, Weirather JL, Peyrard-Janvid M, de Lima RLLF, Dunwald M, et al. (2011) Genomic strategy identifies a missense mutation in WD-repeat domain 65 (WDR65) in an individual with Van der Woude syndrome. *Am J Med Genet Part A* 155A: 1314–1321. doi: [10.1002/ajmg.a.33980](https://doi.org/10.1002/ajmg.a.33980) PMID: [21574244](https://pubmed.ncbi.nlm.nih.gov/21574244/)
21. Eckert C, Hammesfahr B, Kollmar M (2011) A holistic phylogeny of the coronin gene family reveals an ancient origin of the tandem-coronin, defines a new subfamily, and predicts protein function. *BMC Evol Biol* 11: 268. doi: [10.1186/1471-2148-11-268](https://doi.org/10.1186/1471-2148-11-268) PMID: [21943019](https://pubmed.ncbi.nlm.nih.gov/21943019/)
22. Foth BJ, Goedecke MC, Soldati D (2006) New insights into myosin evolution and classification. *Proc Natl Acad Sci USA* 103: 3681–3686. PMID: [16505385](https://pubmed.ncbi.nlm.nih.gov/16505385/)
23. Singh GP, Chandra BR, Bhattacharya A, Akhourri RR, Singh SK, Sharma A (2004) Hyper-expansion of asparagines correlates with an abundance of proteins with prion-like domains in *Plasmodium falciparum*. *Mol Biochem Parasitol* 137: 307–19. PMID: [15383301](https://pubmed.ncbi.nlm.nih.gov/15383301/)
24. Bréhélin L, Florent I, Gascuel O, Maréchal E (2010) Assessing functional annotation transfers with inter-species conserved coexpression: application to *Plasmodium falciparum*. *BMC Genomics* 11: 35. doi: [10.1186/1471-2164-11-35](https://doi.org/10.1186/1471-2164-11-35) PMID: [20078859](https://pubmed.ncbi.nlm.nih.gov/20078859/)
25. Llinas M, Bozdech Z, Wong ED, Adai AT, DeRisi JL (2006) Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. *Nucleic Acids Res* 34: 1166–1173. PMID: [16493140](https://pubmed.ncbi.nlm.nih.gov/16493140/)
26. Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, et al. (2003) Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* 301: 1503–1508. PMID: [12893887](https://pubmed.ncbi.nlm.nih.gov/12893887/)

27. Florens L, Washburn MP, Raine JD, Anthony RM, Grainger M, Haynes JD, et al. (2002) A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* 419: 520–526. PMID: [12368866](#)
28. Lasonder E, Ishihama Y, Andersen JS, Vermunt AM, Pain A, Sauerwein RW, et al. (2002) Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* 419: 537–542. PMID: [12368870](#)
29. Lasonder E, Janse CJ, van Gemert GJ, Mair GR, Vermunt AM, Douradinha BG, et al. (2008) Proteomic profiling of *Plasmodium* sporozoite maturation identifies new proteins essential for parasite development and infectivity. *PLoS Pathog* 4: e1000195. doi: [10.1371/journal.ppat.1000195](#) PMID: [18974882](#)
30. Le Roch KG, Johnson JR, Florens L, Zhou Y, Santrosyan A, Grainger M, et al. (2004) Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle. *Genome Res* 14: 2308–2318. PMID: [15520293](#)
31. Khan SM, Franke-Fayard B, Mair GR, Lasonder E, Janse CJ, Mann M, et al. (2005) Proteome analysis of separated male and female gametocytes reveals novel sex-specific *Plasmodium* biology. *Cell* 121: 675–687. PMID: [15935755](#)
32. Silvestrini F, Lasonder E, Olivieri A, Camarda G, Schaijk BV, Sanchez M, et al. (2010) Protein export marks the early phase of gametocytogenesis of the human malaria parasite *Plasmodium falciparum*. *Mol Cell Proteomics* 9: 1437–1448. doi: [10.1074/mcp.M900479-MCP200](#) PMID: [20332084](#)
33. Oehring SC, Woodcroft BJ, Moes S, Wetzel J, Dietz O, Pulfer A, et al. (2012) Organellar proteomics reveals hundreds of novel nuclear proteins in the malaria parasite *Plasmodium falciparum*. *Genome Biol* 13: R108. doi: [10.1186/gb-2012-13-11-r108](#) PMID: [23181666](#)
34. Lindner SE, Swearingen KE, Harupa A, Vaughan AM, Sinnis P, Moritz RL, et al. (2013) Total and putative surface proteomics of malaria parasite salivary gland sporozoites. *Mol Cell Proteomics* 12: 1127–1143. doi: [10.1074/mcp.M112.024505](#) PMID: [23325771](#)
35. Solyakov L, Halbert J, Alam MM, Semblat JP, Dorin-Semblat D, Reininger L, et al. (2011) Global kinomic and phospho-proteomic analyses of the human malaria parasite *Plasmodium falciparum*. *Nat Commun* 2: 565. doi: [10.1038/ncomms1558](#) PMID: [22127061](#)
36. Treeck M, Sanders JL, Elias JE, Boothroyd JC (2011) The Phosphoproteomes of *Plasmodium falciparum* and *Toxoplasma gondii* reveal unusual adaptations within and beyond the parasites' boundaries. *Cell Host Microbe* 10: 410–419. doi: [10.1016/j.chom.2011.09.004](#) PMID: [22018241](#)
37. Pease BN, Huttlin EL, Jedrychowski MP, Talevich E, Harmon J, Dillman T, et al. (2013) Global analysis of protein expression and phosphorylation of three stages of *Plasmodium falciparum* intraerythrocytic development. *J Proteome Res* 12: 4028–4045. doi: [10.1021/pr400394g](#) PMID: [23914800](#)
38. Guttery DS, Ferguson DJP, Poulin B, Xu Z, Straschil U, Klop O, et al. (2012) A putative homologue of CDC20/CDH1 in the malaria parasite is essential for male gamete development. *PLoS Pathog* 8: e1002554. doi: [10.1371/journal.ppat.1002554](#) PMID: [22383885](#)
39. Woodland HR, Fry AM (2008) Pix Proteins and the Evolution of Centrioles. *PLoS One* 3: e3778. doi: [10.1371/journal.pone.0003778](#) PMID: [19020665](#)
40. Hall N, Karras M, Raine JD, Carlton JM, Kooij TW, Berriman M, et al. (2005) A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses. *Science* 307: 82–86. PMID: [15637271](#)
41. Tarun AS, Peng X, Dumpit RF, Ogata Y, Silva-Rivera H, Camargo N, et al. (2008) A combined transcriptome and proteome survey of malaria parasite liver stages. *Proc Natl Acad Sci USA* 105: 305–310. doi: [10.1073/pnas.0710780104](#) PMID: [18172196](#)
42. Bischoff E, Vaquero C (2010) In silico and biological survey of transcription-associated proteins implicated in the transcriptional machinery during the erythrocytic development of *Plasmodium falciparum*. *BMC Genomics* 11: 34. doi: [10.1186/1471-2164-11-34](#) PMID: [20078850](#)
43. Hain AU, Bosch J (2013) Autophagy in *Plasmodium*, a multifunctional pathway. *Comput Struct Biotechnol J* 8: e201308002. doi: [10.5936/csbj.201308002](#) PMID: [24688742](#)
44. Hurst LD, Pál C, Lercher MJ (2004) The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet* 5: 299–310. PMID: [15131653](#)
45. Thévenin A, Ein-Dor L, Ozery-Flato M, Shamir R (2014) Functional gene groups are concentrated within chromosomes, among chromosomes and in the nuclear space of the human genome. *Nucleic Acids Res* 42: 9854–61. doi: [10.1093/nar/gku667](#) PMID: [25056310](#)
46. LaCount DJ, Vignali M, Chettier R, Phansalkar A, Bell R, Hesselberth JR, et al. (2005) A protein interaction network of the malaria parasite *Plasmodium falciparum*. *Nature* 438: 103–107. PMID: [16267556](#)
47. Aurrecochea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, et al. (2009) PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res* 37(Database issue): D539–43. doi: [10.1093/nar/gkn814](#) PMID: [18957442](#)

48. Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, et al. (2011) InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res* 40: D306–D312. doi: [10.1093/nar/gkr948](https://doi.org/10.1093/nar/gkr948) PMID: [22096229](https://pubmed.ncbi.nlm.nih.gov/22096229/)
49. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. (2012) The Pfam protein families database. *Nucleic Acids Res* 40: D290–D301. doi: [10.1093/nar/gkr1065](https://doi.org/10.1093/nar/gkr1065) PMID: [22127870](https://pubmed.ncbi.nlm.nih.gov/22127870/)
50. Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14: 755–763. PMID: [9918945](https://pubmed.ncbi.nlm.nih.gov/9918945/)
51. Wheeler TJ, Clements J, Finn RD (2014) Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. *BMC Bioinformatics* 15: 7. doi: [10.1186/1471-2105-15-7](https://doi.org/10.1186/1471-2105-15-7) PMID: [24410852](https://pubmed.ncbi.nlm.nih.gov/24410852/)
52. Chen F, Mackey AJ, Stoekert CJ Jr, Roos DS (2006) OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res Database*: D363–368. PMID: [16381887](https://pubmed.ncbi.nlm.nih.gov/16381887/)
53. Huerta-Cepas J, Bueno A, Dopazo J, Gabaldón T (2008) PhylomeDB: a database for genome-wide collections of gene phylogenies. *Nucleic Acids Res* 36: D491–496. PMID: [17962297](https://pubmed.ncbi.nlm.nih.gov/17962297/)
54. Claros MG, Vincens P (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur J Biochem* 241: 779–786. PMID: [8944766](https://pubmed.ncbi.nlm.nih.gov/8944766/)
55. Chou KC, Shen HB (2010) A New method for predicting the sub-cellular localization of eukaryotic proteins with both single and multiple sites: Euk-mPLoc 2.0. *PLoS One* 5: e9931. doi: [10.1371/journal.pone.0009931](https://doi.org/10.1371/journal.pone.0009931) PMID: [20368981](https://pubmed.ncbi.nlm.nih.gov/20368981/)
56. Zuegge J, Ralph S, Schmuker M, McFadden GI, Schneider G (2001) Deciphering apicoplast targeting signals—feature extraction from nuclear-encoded precursors of *Plasmodium falciparum* apicoplast proteins. *Gene* 280: 19–26. PMID: [11738814](https://pubmed.ncbi.nlm.nih.gov/11738814/)
57. Foth BJ, Ralph SA, Tonkin CJ, Struck NS, Fraunholz M, Roos DS, et al. (2003) Dissecting Apicoplast Targeting in the Malaria Parasite *Plasmodium falciparum*. *Science* 299: 705–708. PMID: [12560551](https://pubmed.ncbi.nlm.nih.gov/12560551/)
58. Nakai K, Horton P (1999) PSORT: a program for detecting the sorting signals of proteins and predicting their subcellular localization. *Trends Biochem Sci* 24: 34–35. PMID: [10087920](https://pubmed.ncbi.nlm.nih.gov/10087920/)
59. la Cour T, Kierner L, Mølgaard A, Gupta R, Skriver K, Brunak S (2004) Analysis and prediction of leucine-rich nuclear export signals. *Protein Eng Des Sel* 17: 527–536. PMID: [15314210](https://pubmed.ncbi.nlm.nih.gov/15314210/)
60. Felsenstein J (2009) PHYLIP—(Phylogeny Inference Package) Version 3.69. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.
61. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731–2739. doi: [10.1093/molbev/msr121](https://doi.org/10.1093/molbev/msr121) PMID: [21546353](https://pubmed.ncbi.nlm.nih.gov/21546353/)
62. Kelley LA, Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* 4: 363–371. doi: [10.1038/nprot.2009.2](https://doi.org/10.1038/nprot.2009.2) PMID: [19247286](https://pubmed.ncbi.nlm.nih.gov/19247286/)
63. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: A web-based environment for protein structure homology modeling. *Bioinformatics* 22: 195–201. PMID: [16301204](https://pubmed.ncbi.nlm.nih.gov/16301204/)
64. DeLano WL (2008) The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA <http://www.pymol.org>.
65. Lovell SC, Davis IW, Arendall WB III, de Bakker PIW, Word JM, Prisant MG, et al. (2002) Structure validation by C α geometry: phi, psi and C β deviation. *Proteins* 50: 437–450.
66. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* 41: D808–D815. doi: [10.1093/nar/gks1094](https://doi.org/10.1093/nar/gks1094) PMID: [23203871](https://pubmed.ncbi.nlm.nih.gov/23203871/)
67. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498–2504. PMID: [14597658](https://pubmed.ncbi.nlm.nih.gov/14597658/)
68. Ochoa A, Llinás M, Singh M (2011) Using context to improve protein domain identification. *BMC Bioinformatics* 12: 90. doi: [10.1186/1471-2105-12-90](https://doi.org/10.1186/1471-2105-12-90) PMID: [21453511](https://pubmed.ncbi.nlm.nih.gov/21453511/)