Landscape of the Plasmodium interactome

Charles Hillier, Mercedes Pardo, Lu Yu, Ellen Bushell, Theo Sanderson, Tom Metcalf, Colin Herd, Burcu Anar, Julian C. Rayner, Oliver Billker, Jyoti S. Choudhary
Title: Landscape of the Plasmodium interactome

Authors: Charles Hillier†, Mercedes Pardo†, Lu Yu, Ellen Bushell, Theo Sanderson, Tom Metcalf, Colin Herd, Burcu Anar, Julian C. Rayner, Oliver Billker, Jyoti S. Choudhary

Affiliations:
1 Developmental Biology Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany
2 Functional Proteomics, The Institute of Cancer Research, London SW7 3RP, UK
3 Department of Molecular Biology, The Laboratory for Molecular Infection Medicine Sweden, Umeå University, SE-901 87 Umeå, Sweden
4 Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge CB10 1SA, UK

† Contributed equally to this work

* Correspondence: Mercedes Pardo (mercedes.pardo@icr.ac.uk), Oliver Billker (oliver.billker@umu.se), Jyoti Choudhary (jyoti.choudhary@icr.ac.uk)

Lead contact: Mercedes Pardo

SUMMARY

Malaria represents a major global health issue, and the identification and prioritisation of new intervention targets remains an urgent priority. This search is significantly hampered by the fact that over a third of the genes of the malaria-causing Plasmodium parasites remain uncharacterized. We report the first large-scale interaction map of de novo protein assemblies in Plasmodium parasites, generated by combining blue native polyacrylamide electrophoresis with quantitative mass spectrometry and machine learning. Our integrative approach, which spans three different Plasmodium species, identified over twenty thousand putative protein interactions in Plasmodium schizonts, the majority of which are novel, and organized them into 600 complexes. We validate selected novel interactions, thus assigning putative functions in chromatin regulation to previously unannotated proteins and suggesting a role for an EELM2 domain-containing protein and a putative microrchidia protein as mechanistic links between complexes of AP2-domain transcription factors and epigenetic regulation. This represents the first high-confidence map of the native organisation of core conserved cellular processes in Plasmodium parasites. The network reveals putative functions of currently uncharacterized malaria proteins, provides mechanistic and structural insight, and reveals potential new therapeutic targets by grouping them into complexes with known targets.
INTRODUCTION

*Plasmodium* parasites caused 216 million new cases of malaria in 2016 and nearly half a million deaths (WHO World malaria report 2017). The lack of an effective vaccine and spread of resistance to frontline antimalarial treatments makes the search for new intervention targets a repeated and important research priority (Cowman et al., 2016). Target identification requires better understanding of parasite biology and underlying molecular mechanisms of parasite development and pathogenesis. However, despite being the object of intense research efforts, over a third of *Plasmodium* genes still lack functional annotation, in large part because they lack direct orthologues outside closely related parasite species. While systematic genetic screening efforts are filling some of this gap (Bushell et al., 2017; Zhang et al., 2018), genetic data alone is often not enough to provide details of the function of the encoded proteins.

Physical associations between proteins, either transient or stable, in the form of protein complexes, are central to cellular processes. Elucidating protein-protein interactions (PPIs) can therefore assist in placing proteins in cellular pathways or biological processes and ascribing function to poorly characterized proteins. Attempts to elucidate PPI networks in *Plasmodium* parasites have so far relied primarily on *in silico* analysis of functional association information, including gene expression and homology (Ramaprasad et al., 2012). Large-scale experimental evidence has been limited to yeast two-hybrid studies, covering only ~22% of the *P. falciparum* proteome and reliant on expression of peptides in a non-native context (LaCount et al., 2005). In other organisms, systematic pull-down studies have been effective in elucidating PPI networks (Gavin et al., 2006; Gavin et al., 2002; Hein et al., 2015; Huttlin et al., 2017; Huttlin et al., 2015), but lack of proteome-scale panels of antibodies and limitations in genetic manipulation that restrict high-throughput protein tagging in *Plasmodium* make this costly and time-consuming approach unfeasible, particularly to apply in several species. High-throughput chromatographic fractionation combined with quantitative mass spectrometry has emerged recently as an alternative strategy to elucidate protein complexes at systems level, and has been applied in organisms ranging from bacteria to human (Crozier et al., 2017; Havugimana et al., 2012; Kastritis et al., 2017; Kirkwood et al., 2013; Kristensen et al., 2012; Wan et al., 2015). This approach provides a global analysis of the interactome and does not require any genetic manipulation or affinity reagents. Blue native polyacrylamide gel electrophoresis (BN-PAGE), which separates protein complexes in native conformation based on Coomassie Brilliant Blue binding, has higher resolution than gel filtration or sucrose density centrifugation, and has proven to be particularly useful to resolve membrane protein complexes (Bode et al., 2016; Heide et al., 2012; Schagger et al., 1994; Schagger and von Jagow, 1991). Here we used BN-PAGE fractionation coupled to quantitative tandem mass spectrometry in combination with supervised machine learning to build a high-confidence PPI network for *Plasmodium*. 
RESULTS AND DISCUSSION

*Plasmodium* complexome profiling using blue native PAGE coupled to tandem mass spectrometry

To produce a comprehensive and accurate protein interaction network in *Plasmodium* we applied a strategy based on high-throughput biochemical fractionation using BN-PAGE coupled to quantitative tandem mass spectrometry to three different species, namely *P. falciparum*, *P. knowlesi*, and *P. berghei*. *P. falciparum* is the cause of most human malaria mortality and is the dominant malaria parasite in Africa. *P. knowlesi* is a zoonotic pathogen that causes thousands of cases of malaria every year but also serves as a valuable *in vitro* model for *P. vivax*, the most common cause of malaria outside Africa, but which cannot currently be cultured *in vitro*, making it inaccessible for proteomics studies. *P. berghei* is a widely used rodent model for malaria, and has the twin advantages of being an *in vivo* model and one that is highly amenable to genetic manipulation. Because the clinical symptoms of malaria are caused by the asexual erythrocytic parasites, we focused on the schizont stage. In order to achieve broad cellular coverage, we used several detergent concentrations (0, 0.1 and 1% Nonidet P-40) during cell lysis, aiming to balance solubilisation in native conditions with protein recovery. We performed proteome fractionation of schizont lysates by BN-PAGE in duplicate for each experimental condition (*3 Plasmodium* species with 3 detergent concentrations), collected 48 fractions per experiment and analysed each fraction by quantitative liquid chromatography-tandem mass spectrometry (LC-MS/MS) (Fig. 1A). In total, we gathered 18 high-throughput fractionation datasets comprising 864 fractions, yielding a collection of 3,085,858 tandem MS spectra that were attributed to *Plasmodium* peptides (Table S1).

By mapping the gene identifications for each species to *P. falciparum*, we detected a total of 2894 unique proteins (Table S1), representing over 50% of the proteins encoded in the *Plasmodium* genome (Fig. S1A). Given that not all genes are expressed in any given parasite stage, this is likely to represent the majority of schizont proteins. Our dataset overlapped with 74% of proteins identified in the deepest malaria proteome to date, containing 1673 proteins identified by more than one peptide, also from schizonts (Treeck et al., 2011), and identified 537 additional proteins. A greater number of proteins were identified when using higher detergent concentrations, as expected (Fig. S1A). To compare the properties of detected proteins between our experimental datasets, we performed Gene Ontology (GO) enrichment analysis (Table S2). The GO term ‘integral component of membrane’ was under-represented in datasets without detergent (Fisher’s exact test with Bonferroni correction, p < 0.05), but not in the presence of detergent (Fig. S1B). In keeping with the increased solubilisation of membrane proteins with increasing detergent concentrations, we detected some membrane complexes only after detergent extraction (Fig. S1C). This illustrates the enhanced capability of BN-PAGE to resolve both soluble and membrane protein complexes when compared with other fractionation
systems (Havugimana et al., 2012; Kirkwood et al., 2013). More than 60% of proteins ran significantly above their predicted molecular weight, indicating that the majority of proteins migrated as part of higher order assemblies (Fig. S1D).

Next, we quantified protein abundances using extracted MS1-based intensities across the 48 fractions in each experimental condition to derive protein migration profiles for each of the proteins identified (Fig. 1A). To assess the faithfulness of the biochemical fractionation we performed hierarchical clustering of protein migration profiles (Fig. 1B). The clustered profiles of all three Plasmodium species showed remarkable similarity, suggesting a conserved core network of functional units. Subunits of well-characterised conserved housekeeping complexes with a wide range of sizes, such as proteasome, ribosome, nucleosome, chaperonin containing TCP1 (CCT) complex, elongation factor 1 and MCM complex, showed highly correlated co-migration profiles that clustered together (Fig. 1, C and D). We also observed malaria-specific complexes such as the LCCL-lectin adhesive-like protein (LAP) core complex, which is involved in parasite development and infectivity (Saeed et al., 2010; Simon et al., 2009). Strikingly, this complex was only identified in P. berghei fractionations, despite being conserved across Plasmodium (Pradel et al., 2004). This likely reflects the presence of a small number of LCCL-expressing gametocytes in the P. berghei schizont preparation, highlighting that our strategy could be useful in resolving life cycle stage-specific protein complexes. For each pair of proteins we generated a co-migration score by correlating their migration profiles, and this score was later used for assembling co-migration networks through machine learning. To assess reproducibility between experimental profiles, we compared co-migration scores for pairs of proteins by plotting the distribution of absolute differences between common protein pairs. Over 50% of protein pairs displayed co-migration score differences smaller than 0.05 between biological replicates (Fig. S2A), and a similar degree of reproducibility was observed between Plasmodium species datasets (Fig. S2B), allowing data integration in subsequent analyses. Our data illustrates the validity of the BN-PAGE fractionation and protein correlation profiling strategy and overall procedure in recapitulating bona fide protein complexes.

Generation of a Plasmodium protein interaction network

In order to generate a high confidence PPI network, we employed a random forest machine learning approach (Havugimana et al., 2012) that integrated the co-migration scores derived from the biochemical fractionation with additional information supporting physical association (Fig. 2A). The rationale behind this is that physically interacting proteins carry out related biological functions, are co-expressed, and often have similar evolutionary conservation, and these features can be used to enhance the confidence when deriving interactions from the experimental migration profiles. Only protein pairs with strong biochemical evidence from the fractionation datasets (correlation score of at least 0.4) were used, and additional supporting features were applied to this subset. We included two
additional measures derived from biochemical fractionation data reflecting reproducibility, namely the number of fractionation experiments in which the protein pair had a co-migration score of at least 0.4, and the number of fractionation experiments in which the maximal peak in the migration profile overlapped for each protein pair. Other evidence supportive of functional association included gene co-expression (Bozdech et al., 2003; Hu et al., 2010; Modrzynska et al., 2017), interacting domains (Raghavachari et al., 2008; Yellaboina et al., 2011), co-evolution (Juan et al., 2008; Ochoa et al., 2015) and phenotypic data (Bushell et al., 2017) (Table S3). The machine learning classifier was trained and tested with a gold standard set comprising only experimentally-determined Plasmodium interactions annotated in the STRING database (Szklarczyk et al., 2017) (see Methods for details).

Assessment of the relative contribution of each feature to the prediction of PPIs (as measured by the Gini score) confirmed that the biochemical evidence collectively had the biggest impact on the classification (Table S3), reflecting the superior power of co-migration compared to other functional association information in predicting interactions. To measure the overall performance of the classifier, we performed receiver operator characteristic analysis of the high-confidence PPIs against the gold standard test set. This revealed a significant improvement in recalling true interactions for the classifier compared to the co-fractionation data alone (Fig. 2C). After applying a random forest (RF) score threshold of 0.9, based on maximal overlap of protein clusters with annotated protein complexes (see below), the machine learning analysis yielded a network of 1,761 Plasmodium proteins and 26,060 interactions (Fig. 2B, Tables S4, S5). More stringent filtering such as increasing the random forest score threshold or introducing additional criteria (like requiring a co-migration score of at least 0.4 in at least two experiments) did not result in significant increase in true positive recall but lead to loss of interactions that were subsequently validated experimentally (see below).

We also tested performance against other interaction datasets not used in the training (Fig. S3). The PPI network recapitulated 16% of Plasmodium experimentally-determined interactions annotated in STRING. Only 1% of the interactions for P. falciparum in BioGrid, comprised mainly of yeast two-hybrid data (LaCount et al., 2005), were captured, indicating a significant divergence between our native PPI network and that inferred from heterologous expression of peptides. It is worth noting that this yeast-two-hybrid interactome, the sole experimental systems level Plasmodium interactome published to date, has low coverage and is based on expression of small-sized AT-rich Plasmodium protein fragments, and thus the disagreement in the two sets is hardly surprising. Our network recapitulated 33% of the interactions derived from REACTOME, and replicated interactions previously described in other eukaryotic organisms. We found 3632 PPIs that have previously been reported for orthologs in yeast, worm, fly or human (as annotated in STRING, experimental interactions only, minimum of 0.4 evidence score), with a third of these reported in yeast but not in metazoans. Less than 200 PPIs have been reported in metazoans but not in yeast. For instance, we detected an interaction between Plasmodium Rab GDP dissociation inhibitor (GDI) and a
farnesyltransferase; this interaction is annotated for *Plasmodium* in STRING based on experimental evidence of putative orthologs interacting in *Saccharomyces cerevisiae*. Rab GDIs bind to prenylated Rab proteins that regulate vesicular trafficking and delimit membrane structures, delivering them to and retrieving them from their membrane-bound compartment (Pfeffer et al., 1995). Our data demonstrates that this interaction does indeed occur in *Plasmodium*. Overall, the majority of PPIs described here are novel (over 21,000) and represent a substantial increase in the number of apicomplexan interactions reported to date. The interaction dataset is available in *PlasmoGEM* (www.plasmogem.sanger.ac.uk) as a searchable database allowing easy access. To facilitate the use and navigation of the interaction network, we also provide all the underlying data (including random forest analysis score, correlation scores, supporting features and proteins’ attributes) in consolidated files that are compatible with network visualization programs, allowing interactive display and exploration of the malaria interactome (Tables S4, S5). These tables enable review of scores and supporting features for protein pairs of interest used for determining an interaction, and the attributes of proteins involved, and can assist with the prioritisation of candidate interactions for further studies.

**Inferring putative protein function for *Plasmodium* proteins from the interaction network**

To delineate distinct functional units from the network we performed cluster analysis using the ClusterONE algorithm (Nepusz et al., 2012). The resulting 593 clusters, comprising 1259 unique proteins, represent putative protein complexes ranging in size between two and 74 nodes (Fig. 2D, Table S6). The clusters recapitulated 67 complexes previously reported in STRING and 89 complexes previously reported in CORUM (Table S7), where at least 50% of the cluster nodes belong to a STRING or CORUM protein complex. The clusters are involved in a wide variety of biological processes, and 59 of them showed enrichment of GO biological function terms (Table S7). Hence, our PPI network is able to provide high resolution information on malaria protein complexes and is rich in novel functional protein assemblies.

We identified several clusters representing well-characterised malaria-specific protein complexes (Fig. 3A), including the PTEX translocon, responsible for protein export from the parasitophorous vacuole (de Koning-Ward et al., 2009), the adaptor protein 1 (AP-1) complex, which has a role in protein trafficking to rhoptry organelles (Kaderi Kibria et al., 2015), and the glideosome, a structure involved in merozoite entry into the host erythrocyte during invasion (Soldati-Favre, 2008). These are all examples of membrane protein complexes, which further highlight the utility of our BN-PAGE approach in resolving these less soluble assemblies that would be missed by alternative chromatographic fractionation approaches. Previous interactome studies based on chromatographic fractionation have focused mainly on soluble complexes (Havugimana et al., 2012) and they could be complemented by BN-PAGE fractionation studies.
To explore malaria-specific interactions in a systematic manner we looked into conserved *Plasmodium* proteins of unknown function. The clustered network encompassed 247 of these, distributed among 184 clusters (eleven clusters corresponded to seven CORUM overlapping complexes), potentially representing malaria-specific subunits and/or complexes (Table S6; Fig. 2D). Fifteen conserved *Plasmodium* proteins belonged to clusters enriched for certain biological process GO terms, and eleven belonged to CORUM clusters, and this data could inform their function through “guilt by association”. The high-confidence PPI network can indeed predict function of poorly characterized or unannotated malaria proteins, associate them with specific aspects of parasite biology, or shed light on mechanism of action (Table S8, Fig. 3B). A conserved *Plasmodium* protein of unknown function (PF3D7_1419700) is part of a cluster with multiple subunits of translation initiation factor 3 (eIF3), and is annotated in STRING, but not in PlasmoDB, as a putative ortholog of human EIF3H. Our results demonstrate that PF3D7_1419700 indeed interacts with the eIF3 translation initiation complex. Another conserved *Plasmodium* protein (PF3D7_1457300) interacts with different eIFs and contains overlapping multiple copies of the MA3 domain, a protein-protein interaction domain present in eIF4G, suggesting that PF3D7_1457300 could be a scaffold for eIF4 and/or other translation initiation factors. A third conserved *Plasmodium* protein of unknown function (PF3D7_1225200), containing Myb-binding, SWIRM and SANT domains, belongs to cluster 57, which also includes other DNA-binding proteins, namely transcriptional coactivator ADA2, a SET domain protein, a WD repeat protein, a protein containing a bromodomain, which binds acetylated histones (Owen et al., 2000), and a Snf2h-related CBP activator, and also has a link with histone deubiquitinase GCN5. The domains in PF3D7_1225200 are also present in mammalian histone H2A deubiquitinase Mysm1, which interacts with histone acetyltransferase p/CAF (Zhu et al., 2007). The links between the WD repeat, bromodomain and SET domain proteins are reminiscent of interactions reported in mammalian cells (Bode et al., 2016; Dharmarajan et al., 2012; Dou et al., 2005; Revenko et al., 2010).

Sequence-specific regulators of transcription in Apicomplexa are characterized by functional *apetala2* (AP2) DNA binding domains (Balaji et al., 2005). Mechanisms through which these ApiAP2 proteins regulate transcription have remained elusive, and it is therefore remarkable that we observed a complex of three likely essential ApiAP2 proteins (Table S6, cluster 454; Fig. 3C), two of which (PBANKA_1453700/PF3D7_1239200 and PBANKA_0939100/PF3D7_1107800) mark each of two major gene expression clusters in the second half of the intraerythrocytic developmental cycle in both *P. berghei* and *P. falciparum* (Reid et al., 2018). Amongst ApiAP2 TF second neighbours we identified several DNA-binding proteins including DNA polymerase I, DNA damage proteins and putative epigenetic regulators (Table S6). Other interactors, such as a protein kinase, could be involved in regulation of transcription factor activity and represent potential novel therapeutic targets.
Regulatory interactions such as those between enzyme and substrate might not be encompassed into clusters, yet are still found in the network. For example, we detected an interaction between merozoite surface protein 1 (MSP1) and putative protein arginine N-methyltransferase 5. MSP1 has previously been identified in a pull-down study of arginine-methylated proteins (Zeeshan et al., 2017); our data suggests that it could be a target of PRMT5. Arginine methylation may regulate a broad array of Plasmodium physiological processes, and in fact three other interactors of PRMT5 identified in our study were also previously found to be arginine-methylated (Zeeshan et al., 2017). Overall, these examples illustrate the potential of our PPI network as a source of novel insight into individual complexes representing a wide array of parasite pathophysiology.

Validation of protein complexes from the Plasmodium protein interaction network

To assess the precision of our study, we chose candidate interactions for validation based on availability of reagents, similarity of GO terms with putative interacting proteins in the same cluster and essential phenotype. We performed affinity purification coupled to mass spectrometry experiments from P. berghei schizonts expressing endogenous HA-tagged versions of selected candidates and validated several interactions from the network, in addition to discovering other interactions (Fig. 3D, E). Bearing in mind that the selection of candidates was done on a predetermined list of available tagged proteins, not necessarily from the most confident interactions in the network, and also that the two interaction mapping methods, blue native co-migration and affinity purification, are based on different biochemical attributes and may not recapitulate exactly, these results are noteworthy.

To validate the proposed complex of ApiAP2 proteins, epitope-tagged PBANKA_0939100 was immunoprecipitated from lysates of P. berghei schizonts, and mass spectrometry identification of co-purifying proteins confirmed the interaction with PBANKA_0112100 (Table 1, Table S9). We additionally confirmed an association with two conserved Plasmodium proteins of unknown function. The first is a nuclear protein with an extended Egl-27 and MTA1 homology 2 (EELM2) domain (Oehring et al., 2012). The second combines an array of Kelch motifs with a GHKL (Gyrase, Hsp90, Histidine kinase, MutL)-type ATPase domain, which is a hallmark of microrchidia (MORC) proteins (Iyer et al., 2008). ELM2 and MORC proteins are typically found in a chromatin regulatory complex with histone deacetylase (HDAC) and nucleosome remodelling activities (Solari et al., 1999). Consistently, the Toxoplasma gondii ortholog of Plasmodium MORC forms part of a co-repressor complex with TgHDAC3 (Saksouk et al., 2005). In our interactome, MORC has a direct link with HDAC1, with a second EELM2 protein and with two other ApiAP2 proteins (Fig. 3D, Table S5). These data lead us to hypothesise that Plasmodium MORC and EELM2 proteins form a scaffold connecting transcription factor complexes with epigenetic regulators to effect nucleosome reorganisation and regulate gene expression. In support of such a mechanism, a DNA motif
recognised by PF3D7_1107800, the *P. falciparum* orthologue of PBANKA_0939100, mirrors nucleosome spacing around transcriptional start sites (Kensche et al., 2016). Furthermore, the same ApiAP2 protein targets a genetic element in the introns of *P. falciparum* var genes that is important for tethering members of this multigene family to the nuclear periphery as part of their epigenetic silencing (Zhang et al., 2011).

Affinity purification followed by mass spectrometry analysis of HA-tagged SNF2L, a putative chromatin remodelling protein, identified uncharacterised protein PBANKA_1461600, confirming a physical link between these two proteins (Fig. 3C, D), which are connected in the network through the eIF3 complex. We also confirmed a physical association between putative cohesin subunits SMC3 (PBANKA_0716000) and SMC1 (PBANKA_0917500) and identified a further SMC3-associated protein, uncharacterized protein PBANKA_1304000, which in the network is connected to both SMC1 and SMC3 through another uncharacterised protein that we did not detect in the pull-downs. PBANKA_1304000 contains a Rad21/Rec8-like N-terminal domain, which is a conserved N-terminal region present in eukaryotic cohesins. Our results indicate that PBANKA_1304000 is indeed a subunit of the cohesin complex. Altogether, these data demonstrate the utility of our interaction network in informing novel biological function of uncharacterized *Plasmodium* proteins and validate the use of the blue native PAGE-based approach to identify protein interactions at proteome scale.

**Species-specific interactions and a core conserved Plasmodium interactome**

We took advantage of the three *Plasmodium* species datasets and carried out cross-species analysis, focusing on proteins specific to only one species or to a subset of *Plasmodium* species, which possibly affect important differential phenotypes between malaria parasites, such as their ability to cause human disease (Frech and Chen, 2011). Primate parasite proteins that are absent in rodent parasites could represent pathogenicity genes required for human infection, which potentially makes them interesting targets for the development of new antimalarials (Frech and Chen, 2011; Muller and Kappes, 2007; Pessi et al., 2004). We found six examples of such proteins distributed in nine clusters (three of them overlapping). In agreement with their parasite specificity, none of their interactions were observed in *P. berghei* samples (Table S6). Three of these human *Plasmodium* specific proteins are discussed further below.

Cluster 7 contained phosphomethylpyrimidine kinase, a key enzyme of the vitamin B1 biosynthesis pathway. Rodent parasites lack three of the enzymes of the pathway, suggesting they are incapable of *de novo* thiamine synthesis (Frech and Chen, 2011; Muller and Kappes, 2007). Indeed, thiamine metabolism has been proposed as antimalarial drug target (Chan et al., 2013). Intriguingly, cluster 7 also contained two uncharacterised *P. knowlesi* proteins without orthologs in other species, PKNH_1402300 and PKNH_1473300. The latter is a *Plasmodium* exported protein of the PHIST (*Plasmodium* helical interspersed subtelomeric family) family, whose members are present in many
copies in human *Plasmodium* but only one in rodent *Plasmodium*, suggesting that cluster 7 is a primate-specific cluster potentially involved in the biosynthesis of thiamine.

Cluster 32 included primate-specific uncharacterised protein PF3D7_1118000 (Frech and Chen, 2011), *P. knowlesi* PIR (protein with interspersed repeats) family protein PKNH_1000600, and signal recognition particle (SRP) subunit SRP19, which is involved in SRP-dependent co-translational protein targeting to membranes. PIR protein PKNH_1000600 is part of the vir family. VIR proteins can localise to the membrane of infected erythrocytes and have been linked to immune evasion (Ben Mamoun et al., 2010; Bernabeu et al., 2012; del Portillo et al., 2001; Kilian et al., 2018). In light of its interactions, we hypothesise that PF3D7_1118000 may have a function related to endothelial adherence specifically in the human host.

Cluster 223 included phosphoethanolamine N-methyltransferase (PMT), an enzyme involved in polyamine and phospholipid metabolism that is absent in rodent *Plasmodium*, leading to marked differences in phospholipid biosynthetic pathways between human and rodent malaria parasites (Dechamps et al., 2010; Frech and Chen, 2011). Other proteins in the cluster were a putative hydrolase, a coatomer subunit, Fe-S cluster assembly factor NBP35, and PF3D7_1441100, a functionally uncharacterised protein annotated to the mitochondrial disulfide relay system metabolic pathway (PlasmoDB). Isoprenoid biosynthesis is predicted to rely on Fe-S cluster cofactors, but little is known about Fe-S cluster synthesis or the roles that Fe-S cluster proteins play in *Plasmodium* biology (Gisselberg et al., 2013). *P. falciparum* PMT is important for membrane biogenesis and parasite development, survival and propagation, therefore this and other enzymes in the phosphatidylcholine biosynthetic pathway are attractive targets for antimalarial therapy (Ben Mamoun et al., 2010; Kilian et al., 2018). Indeed, all proteins in cluster 223 are deemed to be essential (Zhang et al., 2018).

To define a set of interactions conserved across the full range of the *Plasmodium* genus we segregated edges that were common to protein families across the three species (7117) and performed cluster analysis. This identified 158 clusters conserved across all three *Plasmodium* species (Fig. 4A, 4B; Table S10). These conserved clusters represent 50 of the 89 CORUM clusters detected in the high confidence PPI network. This represents the most confident subset of interactions, since they were detected in biochemical fractionations in all three different *Plasmodium* species. We observed positive enrichment for “essential” and negative enrichment for “dispensable” proteins in this group (Fisher’s exact test with Bonferroni correction, $p < 0.05$), using growth phenotype data from a recent large-scale *P. berghei* knockout screen (Bushell et al., 2017), indicating these proteins are involved in core biological processes. To confirm our suggestion that these interactions are conserved, we compared the distribution of protein family relative evolutionary rates. We observed similar distributions for the protein families in each network, but slower relative evolutionary rates in the interactions conserved
across species (Fig. 4C). This confirms that the common protein interaction network embodies an evolutionarily conserved core set of protein interactions representing the essential backbone of *Plasmodium* schizont biology.

Summing up, the data presented here demonstrate that BN-PAGE fractionation studies provide a valuable complement to previous interactome studies based on chromatographic fractionation, which have mainly focused on soluble complexes. Our integrative interaction proteomics approach has resulted in the identification of around 600 novel malaria protein complexes encompassing 1785 proteins (approximately one third of the malaria proteome) and represents the broadest experimentally-determined malaria interactome available. Information derived from our PPI network will be useful to assign function to many currently uncharacterized *Plasmodium* proteins, provide mechanistic insight into proteins’ roles, guide directed studies on malaria proteins’ function, structure and dynamics, and help characterize potential therapeutic targets. Addition of new high throughput data such as localisation information onto this model could be helpful in further refining the network. This first *Plasmodium* schizont interactome paves the way for investigating changes in protein complexes during the different phases of the malaria parasite life cycle at systems level.

**ACKNOWLEDGEMENTS**

This work was supported by the Wellcome Trust [grant number 2016194/Z/17/Z].

**AUTHOR CONTRIBUTIONS**

Conceptualization: J.S.C., J.R., M.P., O.B. Formal analysis: C.Hi., M.P., O.B. Investigation: C.Hi., C.He., E.B., M.P., T.M., T.S. Resources: B.A. Writing - original draft: C. Hi., M.P. Writing - review and editing: M.P., C. Hi., E.B., J.S.C., J.R., O.B., T.S. Visualization: C.Hi., M.P., O.B. Supervision: E.B., J.S.C., J.R., M.P., O.B.

**DECLARATION OF INTERESTS**

The authors declare no competing interests.

**REFERENCES**

Aurrecoechea, C., Brestelli, J., Brunk, B.P., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O.S., et al. (2009). PlasmoDB: a functional genomic database for malaria parasites. Nucleic Acids Res 37, D539-543.

Balaji, S., Babu, M.M., Iyer, L.M., and Aravind, L. (2005). Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. Nucleic Acids Res 33, 3994-4006.

Ben Mamoun, C., Prigge, S.T., and Vial, H. (2010). Targeting the Lipid Metabolic Pathways for the Treatment of Malaria. Drug Dev Res 71, 44-55.
Bernabeu, M., Lopez, F.J., Ferrer, M., Martin-Jaujar, L., Razaname, A., Corradin, G., Maier, A.G., Del Portillo, H.A., and Fernandez-Becerra, C. (2012). Functional analysis of Plasmodium vivax VIR proteins reveals different subcellular localizations and cytoadherence to the ICAM-1 endothelial receptor. Cell Microbiol 14, 386-400.

Bode, D., Yu, L., Tate, P., Pardo, M., and Choudhary, J. (2016). Characterization of Two Distinct Nucleosome Remodeling and Deacetylase (NuRD) Complex Assemblies in Embryonic Stem Cells. Mol Cell Proteomics 15, 878-891.

Bozdech, Z., Llinas, M., Pulliam, B.L., Wong, E.D., Zhu, J., and DeRisi, J.L. (2003). The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. PLoS Biol 1, E5.

Bushell, E., Gomes, A.R., Sanderson, T., Anar, B., Girling, G., Herd, C., Metcalf, T., Modzynska, K., Schwach, F., Martin, R.E., et al. (2017). Functional Profiling of a Plasmodium Genome Reveals an Abundance of Essential Genes. Cell 170, 260-272 e268.

Chan, X.W., Wrenger, C., Stahl, K., Bergmann, B., Winterberg, M., Muller, I.B., and Saliba, K.J. (2013). Chemical and genetic validation of thiamine utilization as an antimalarial drug target. Nat Commun 4, 2060.

Cowman, A.F., Healer, J., Marapana, D., and Marsh, K. (2016). Malaria: Biology and Disease. Cell 167, 610-624.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26, 1367-1372.

Crozier, T.W.M., Tinti, M., Larance, M., Lamond, A.I., and Ferguson, M.A.J. (2017). Prediction of Protein Complexes in Trypanosoma brucei by Protein Correlation Profiling Mass Spectrometry and Machine Learning. Mol Cell Proteomics 16, 2254-2267.

de Koning-Ward, T.F., Gilson, P.R., Boddey, J.A., Rug, M., Smith, B.J., Papenfuss, A.T., Sanders, P.R., Lundie, R.J., Maier, A.G., Cowman, A.F., et al. (2009). A newly discovered protein export machine in malaria parasites. Nature 459, 945-949.

Dechamps, S., Maynadier, M., Wein, S., Gannoun-Zaki, L., Marechal, E., and Vial, H.J. (2010). Rodent and nonrodent malaria parasites differ in their phospholipid metabolic pathways. J Lipid Res 51, 81-96.

del Portillo, H.A., Fernandez-Becerra, C., Bowman, S., Oliver, K., Preuss, M., Sanchez, C.P., Schneider, N.K., Villalobos, J.M., Rajandream, M.A., Harris, D., et al. (2001). A superfamily of variant genes encoded in the subtelomeric region of Plasmodium vivax. Nature 410, 839-842.

Dharmarajan, V., Lee, J.H., Patel, A., Skalnik, D.G., and Cosgrove, M.S. (2012). Structural basis for WDR5 interaction (Win) motif recognition in human SET1 family histone methyltransferases. J Biol Chem 287, 27275-27289.

Dou, Y., Milne, T.A., Tackett, A.J., Smith, E.R., Fukuda, A., Wysocka, J., Allis, C.D., Chait, B.T., Hess, J.L., and Roeder, R.G. (2005). Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. Cell 121, 873-885.

Edgar, R.C. (2004a). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 113.

Edgar, R.C. (2004b). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32, 1792-1797.

Frech, C., and Chen, N. (2011). Genome comparison of human and non-human malaria parasites reveals species subset-specific genes potentially linked to human disease. PLoS Comput Biol 7, e1002320.

Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dunpelfeld, B., et al. (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature 440, 631-636.

Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415, 141-147.
Gisselberg, J.E., Dellibovi-Ragheb, T.A., Matthews, K.A., Bosch, G., and Prigge, S.T. (2013). The sulfiron-sulfur cluster synthesis pathway is required for apicoplast maintenance in malaria parasites. PLoS Pathog 9, e1003655.

Hall, N., Karras, M., Raine, J.D., Carlton, J.M., Kooij, T.W., Berriman, M., Flores, L., Janssen, C.S., Pain, A., Christophides, G.K., et al. (2005). A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and proteomic analyses. Science 307, 82-86.

Havugimana, P.C., Hart, G.T., Nepusz, T., Yang, H., Turinsky, A.L., Li, Z., Wang, P.I., Boutz, D.R., Fong, V., Phanse, S., et al. (2012). A census of human soluble protein complexes. Cell 150, 1068-1081.

Heide, H., Bleier, L., Steger, M., Ackermann, J., Drose, S., Schwamb, B., Zornig, M., Reichert, A.S., Koch, I., Wittig, I., et al. (2012). Complexome profiling identifies TMEM126B as a component of the mitochondrial complex I assembly complex. Cell Metab 16, 538-549.

Hein, M.Y., Hubner, N.C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Dak, I.A., Weisswange, I., Mansfeld, J., Buchholz, F., et al. (2015). A human interactome in three quantitative dimensions organized by stoichiometries and abundances. Cell 163, 712-723.

Hu, G., Cabrera, A., Kono, M., Mok, S., Chaal, B.K., Haase, S., Engelberg, K., Cheemadan, S., Spielmann, T., Preiser, P.R., et al. (2010). Transcriptional profiling of growth perturbations of the human parasite Plasmodium falciparum. Nat Biotechnol 28, 91-98.

Huttlin, E.L., Bruckner, R.J., Paulo, J.A., Cannon, J.R., Ting, L., Baltier, K., Colby, G., Gebreab, F., Gygi, M.P., Parzen, H., et al. (2017). Architecture of the human interactome defines protein communities and disease networks. Nature 545, 505-509.

Kensche, P.R., Hoeijmakers, W.A., Toenhake, C.G., Bras, M., Chappell, L., Berriman, M., and Bartfai, R. (2016). The nucleosome landscape of Plasmodium falciparum reveals chromatin architecture and dynamics of regulatory sequences. Nucleic Acids Res 44, 2110-2124.

Kristensen, A.R., Gspner, J., and Foster, L.J. (2012). A high-throughput approach for measuring temporal changes in the interactome. Nat Methods 9, 907-909.

LaCount, D.J., Vignali, M., Chettier, R., Phansalkar, A., Bell, R., Hesselberth, J.R., Schoenfeld, L.W., Ota, I., Sahasrabudhe, S., Kurschner, C., et al. (2005). A protein interaction network of the malaria parasite Plasmodium falciparum. Nature 438, 103-107.

Liaw, A., and Wiener, M. (2001). Classification and Regression by RandomForest, Vol 23.
Logan-Klumpler, F.J., De Silva, N., Boehme, U., Rogers, M.B., Velarde, G., McQuillan, J.A., Carver, T., Aslett, M., Olsen, C., Subramanian, S., et al. (2012). GeneDB—an annotation database for pathogens. Nucleic Acids Res 40, D98-108.

Modrzynska, K., Pfaender, C., Chappell, L., Yu, L., Suarez, C., Dundas, K., Gomes, A.R., Goulding, D., Rayner, J.C., Choudhary, J., et al. (2017). A Knockout Screen of ApiAP2 Genes Reveals Networks of Interacting Transcriptional Regulators Controlling the Plasmodium Life Cycle. Cell Host Microbe 21, 11-22.

Muller, S., and Kappes, B. (2007). Vitamin and cofactor biosynthesis pathways in Plasmodium and other apicomplexan parasites. Trends Parasitol 23, 112-121.

Nepusz, T., Yu, H., and Paccanaro, A. (2012). Detecting overlapping protein complexes in protein-protein interaction networks. Nat Methods 9, 471-472.

Ochoa, D., Juan, D., Valencia, A., and Pazos, F. (2015). Detection of significant protein coevolution. Bioinformatics 31, 2166-2173.

Oehring, S.C., Woodcroft, B.J., Moes, S., Wetzel, J., Dietz, O., Pulfer, A., Dekiwadia, C., Maeser, P., Flick, C., Witmer, K., et al. (2012). Organellar proteomics reveals hundreds of novel nuclear proteins in the malaria parasite Plasmodium falciparum. Genome Biol 13, R108.

Owen, D.J., Ornaghi, P., Yang, J.C., Lowe, N., Evans, P.R., Ballario, P., Neuhaus, D., Filetici, P., and Travers, A.A. (2000). The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. EMBO J 19, 6141-6149.

Pardo, M., Bode, D., Yu, L., and Choudhary, J.S. (2017). Resolving Affinity Purified Protein Complexes by Blue Native PAGE and Protein Correlation Profiling. J Vis Exp. 122, Apr 1, doi 10.3791/55498.

Pardo, M., Lang, B., Yu, L., Prosser, H., Bradley, A., Babu, M.M., and Choudhary, J. (2010). An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. Cell Stem Cell 6, 382-395.

Pessi, G., Kociubinski, G., and Mamoun, C.B. (2004). A pathway for phosphatidylcholine biosynthesis in Plasmodium falciparum involving phosphoethanolamine methylation. Proc Natl Acad Sci U S A 101, 6206-6211.

Pfeffer, S.R., Dirac-Svejstrup, A.B., and Soldati, T. (1995). Rab GDP dissociation inhibitor: putting rab GTPases in the right place. J Biol Chem 270, 17057-17059.

Pradel, G., Hayton, K., Aravind, L., Iyer, L.M., Abrahamsen, M.S., Bonawitz, A., Mejia, C., and Templeton, T.J. (2004). A multidomain adhesion protein family expressed in Plasmodium falciparum is essential for transmission to the mosquito. J Exp Med 199, 1533-1544.

Raghavachari, B., Tasneem, A., Przytycka, T.M., and Jothi, R. (2008). DOMINE: a database of protein domain interactions. Nucleic Acids Res 36, D656-661.

Ramaprasad, A., Pain, A., and Ravasi, T. (2012). Defining the protein interaction network of human malaria parasite Plasmodium falciparum. Genomics 99, 69-75.

Reid, A.J., Talman, A.M., Bennett, H.M., Gomes, A.R., Sanders, M.J., Illingworth, C.J.R., Billker, O., Berriman, M., and Lawniczak, M.K. (2018). Single-cell RNA-seq reveals hidden transcriptional variation in malaria parasites. Elife 7.

Revenko, A.S., Kalashnikova, E.V., Gemo, A.T., Zou, J.X., and Chen, H.W. (2010). Chromatin loading of E2F-MLL complex by cancer-associated coregulator ANCCA via reading a specific histone mark. Mol Cell Biol 30, 5260-5272.

Saeed, S., Carter, V., Trempl, A.Z., and Dessens, J.T. (2010). Plasmodium berghei crystalloids contain multiple LCCL proteins. Mol Biochem Parasitol 170, 49-53.

Saksouk, N., Bhattacharyya, R., Kieber, S., Smith, S.A., Musset, K., Garin, J., Sullivan, W.J., Jr., Cesbron-Delauw, M.F., and Hakimi, M.A. (2005). Histone-modifying complexes regulate gene expression pertinent to the differentiation of the protozoan parasite Toxoplasma gondii. Mol Cell Biol 25, 10301-10314.

Schagger, H., Cramer, W.A., and von Jagow, G. (1994). Analysis of molecular masses and oligomeric states of protein complexes by blue native electrophoresis and isolation of membrane protein complexes by two-dimensional native electrophoresis. Anal Biochem 217, 220-230.
Schagger, H., and von Jagow, G. (1991). Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. Anal Biochem 199, 223-231.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13, 2498-2504.

Simon, N., Scholz, S.M., Moreira, C.K., Templeton, T.J., Kuehn, A., Dude, M.A., and Pradel, G. (2009). Sexual stage adhesion proteins form multi-protein complexes in the malaria parasite Plasmodium falciparum. J Biol Chem 284, 14537-14546.

Soldati-Favre, D. (2008). Molecular dissection of host cell invasion by the apicomplexans: the glideosome. Parasite 15, 197-205.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods 13, 731-740.

Veerassamy, S., Smith, A., and Tillier, E.R. (2003). A transition probability model for amino acid substitutions from blocks. J Comput Biol 10, 997-1010.

Vizcaino, J.A., Csordas, A., Del-Toro, N., Diames, J.A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., et al. (2016). 2016 update of the PRIDE database and its related tools. Nucleic Acids Res 44, 11033.

Wan, C., Borgeson, B., Phanse, S., Tu, F., Drew, K., Clark, G., Xiong, X., Kagan, O., Kwan, J., Bezginov, A., et al. (2015). Panorama of ancient metazoan macromolecular complexes. Nature 525, 339-344.

Zhang, Q., Huang, Y., Zhang, Y., Fang, X., Claes, A., Duchateau, M., Namane, A., Lopez-Rubio, J.J., Pan, W., and Scherf, A. (2011). A critical role of perinuclear filamentous actin in spatial repositioning and mutually exclusive expression of virulence genes in malaria parasites. Cell Host Microbe 10, 451-463.

Zhu, P., Zhou, W., Wang, J., Puc, J., Ohgi, K.A., Erdjument-Bromage, H., Tempst, P., Glass, C.K., and Rosenfeld, M.G. (2007). A histone H2A deubiquitinase complex coordinating histone acetylation and H1 dissociation in transcriptional regulation. Mol Cell 27, 609-621.
FIGURES

Figure 1. Identification of *Plasmodium* protein complexes. (A) Schematic overview of the blue native polyacrylamide gel electrophoresis (BN-PAGE) strategy used to derive protein interaction network. *Plasmodium* schizont lysates were subjected to BN-PAGE and migration profiles generated for each protein using MS1 peak intensities measured by quantitative LC-MS/MS. Profiles were correlated to generate pairwise co-migration scores, which were then used to build a network. (B) Representative heatmaps derived from hierarchical clustering of protein migration profiles for *P. berghei* (green), *P. falciparum* (red) and *P. knowlesi* (purple). Dendrograms are shown on the left. Clustering was performed with Pearson correlation, complete linkage, and pre-processing with K-means. Protein markers (molecular weight) are shown at the top. Scale bars represent normalised intensity. The dotted box indicates the ribosome. (C) Migration profiles of select examples of known complexes. Protein descriptions from PlasmoDB are shown. (D) Network graphs of protein complexes shown in C. Coloured nodes represent co-migrating proteins identified here. White nodes represent other known interactors annotated in STRING. Solid edges are found in StringDB. Dashed edges are derived from co-migration only. See also Tables S1 and S2, Figures S1 and S2.
Fig. 1

A

B

C

D

P. berghei

P. falciparum

P. knowlesi

LC-MS/MS

Intensity

Pearson correlation

Parwise co-migration scores

B

C

D

Replication factor C subunit 4
Replication factor C subunit 2
Replication factor C subunit 1
Replication factor C subunit 3
T-complex protein 1 subunit gamma
T-complex protein 1 subunit theta
T-complex protein 1 subunit zeta
T-complex protein 1 subunit delta
T-complex protein 1 subunit epsilon
T-complex protein 1 subunit beta
T-complex protein 1 subunit eta
Eukaryotic translation initiation factor 3A
Eukaryotic translation initiation factor 3C
Eukaryotic translation initiation factor 3B
Eukaryotic translation initiation factor 3D
Eukaryotic translation initiation factor 3E
Eukaryotic translation initiation factor 3F
RNA polymerases I, II, III subunit RPB2
RNA polymerase II subunit RPA2
Proteasome subunit beta type-2
Proteasome subunit beta type-3
Proteasome subunit alpha type-1
Proteasome subunit alpha type-2
Proteasome subunit alpha type-3
Insulin-5-monophosphate dehydrogenase
Proteasome subunit alpha type-4
Proteasome subunit alpha type-5
Proteasome subunit alpha type-6
Proteasome subunit beta type-2
60Da-chaperonin
Coatomer epsilon subunit
Coatomer beta subunit
Fig. 2. Generation of a high-confidence *Plasmodium* protein interaction network. (A) Schematic of the machine learning pipeline applied to the BN-PAGE fractionation data. Pairwise co-migration scores were supported with functional association information using a random forest classifier trained with a gold standard set derived from STRING. (B) High-confidence *Plasmodium* PPI network based on BN-PAGE fractionation data and machine learning. (C) Receiver Operating Characteristic analysis of BN-PAGE fractionation experiments (brown, mean area under curve (auc) = 0.63, sd = 0.023) and the random forest classifier output (blue, auc = 0.94). Performance was assessed against a gold standard set derived from STRING. (D) Protein clusters representing putative protein complexes. Conserved *Plasmodium* proteins of unknown function are shown with a thick border. For C and D, examples of well-known complexes are coloured. Red edges represent interactions annotated in STRING. See also Tables S3, S4, S5, S6, S7 and S8, and Figure S3.
Fig. 3. Protein complex membership for predicting function of malaria uncharacterised proteins. (A) Examples of clusters representing known malaria-specific protein complexes (PTEX (de Koning-Ward et al., 2009), AP-1 (Kaderi Kibria et al., 2015), glideosome (Soldati-Favre, 2008)) found in this study. (B) Examples of clusters containing conserved *Plasmodium* proteins of unknown function (in orange). Red edges represent interactions annotated in STRING. (C) The ApiAP2 transcription factor interaction network. Relevant first order interactions involving ApiAP2 transcription factors were extracted from the PPI network. Nodes labelled cPpuF are conserved *Plasmodium* proteins of unknown function. Orange nodes are essential proteins, yellow nodes represent proteins whose mutation results in slow growth. (D) Validation of interactions by affinity purification-mass spectrometry from tagged *P. berghei* lines. Represented are subsections of the PPI network with the numeric part of *P. falciparum* gene names. Baits are surrounded by a circle. Proteins identified specifically in each bait’s immunoprecipitate are in green. The dotted line represents an indirect link in the network. See also Table S9.
Fig. 4. A conserved malaria interaction network. (A) Clusters were generated from the protein interactions common to all three Plasmodium species studied, visualised as network graphs. Essential proteins are shown as red nodes, blue nodes are proteins whose absence leads to slow growth, and green nodes are proteins which cause no growth phenotype when absent. (B) Examples of conserved clusters detected in all three Plasmodium species. Interactions between green nodes were detected in all species, interactions with blue nodes were not detected in P. knowlesi. Red edges represent interactions annotated in STRING. (C) Density distribution and box plots of relative evolutionary rates of OrthoMCL protein families to P. falciparum. Plots for the high-confidence PPI network and of protein families in the clusters from A are shown. Differences in distribution were assessed with a Wilcoxon-Rank-Sum test (***p-value < 2.9e-14). See also Table S9.
**TABLES**

**Table 1. Validation of protein interactions by AP-MS on tagged *P. berghei* lines.** Number of unique peptide sequences in replicate AP-MS experiments on indicated bait protein (headers), SAINT probability score representing the specificity of the interaction, and average number of unique peptides in control experiments (AvgCtrl, n=5). See also Table S9.

| *P. berghei* accession | *P. falciparum* accession | Description                                                                 | PB093 9100 # Peps | PB093 9100 SP | PB071 6000 # Peps | PB071 6000 SP | PB094 2700 # Peps | PB094 2700 SP | AvgCtrl 1 # Peps |
|------------------------|---------------------------|----------------------------------------------------------------------------|------------------|---------------|------------------|---------------|------------------|---------------|------------------|
| PBANKA_0939100         | PF3D7_11 07800            | transcription factor with AP2 domain(s), putative (ApiAP2)                 | 59 | 48 | 1 | 0 | 2 | 0.01 | 0 | 5 | 0.09 | 1 |
| PBANKA_1331400         | PF3D7_14 68100            | conserved Plasmodium protein, unknown function                           | 26 | 22 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 2 |
| PBANKA_0112100         | PF3D7_06 13800            | transcription factor with AP2 domain(s), putative (ApiAP2)                 | 35 | 27 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 2 |
| PBANKA_1234600         | PF3D7_05 19800            | EELM2 domain-containing protein, putative                               | 4 | 2 | 0.84 | 1 | 1 | 0 | 0 | 0 | 0 |
| PBANKA_0716000         | PF3D7_04 14000            | structural maintenance of chromosomes protein 3, putative               | 0 | 2 | 0 | 47 | 62 | 1 | 0 | 4 | 0.01 | 1 |
| PBANKA_0917500         | PF3D7_11 30700            | structural maintenance of chromosomes protein 1, putative               | 0 | 0 | 0 | 41 | 57 | 1 | 0 | 1 | 0 | 0 |
| PBANKA_1304000         | PF3D7_14 40100            | conserved Plasmodium protein, unknown function                           | 0 | 0 | 0 | 11 | 16 | 1 | 0 | 0 | 0 | 0 |
| PBANKA_0942700         | PF3D7_11 04200            | chromatin remodeling protein, putative (SNF2L)                            | 0 | 0 | 0 | 0 | 0 | 41 | 58 | 1 | 0 | 0 |
| PBANKA_1461600         | PF3D7_12 48700            | conserved Plasmodium protein, unknown function                           | 0 | 0 | 0 | 0 | 0 | 46 | 38 | 1 | 0 | 0 |


METHODS

Animal work

The parasite line used was the c15cy1 P. berghei ANKA reference clone (Hall et al., 2005). All animal work was performed under licenses from the UK Home Office, with protocols approved by the Animal Welfare and Ethical Review Body of the Wellcome Sanger Institute. Rodents were reared in specific-pathogen-free conditions, and were monitored, housed and maintained as previously described (Bushell et al., 2017). Parasitaemia of infected animals were determined by light microscopy of Giemsa stained of thin blood smears.

Eight-twelve week old female Theiler's original (TO) outbred mice (Envigo, UK) were used as a donors for the P. berghei schizont cultures. This mouse strain was chosen to attain robust P. berghei infections with a low frequency of cerebral malaria. For transfections, an eight week old female RCC Han Wistar outbred rat (Envigo, UK) was utilised to generate parasites for the schizont culture. Rats are used because they give rise to more schizonts with a higher transfection efficiency compared to mice. The very high transfection efficiency also means that no dilution cloning was required prior to commencing work with the 3xHA epitope tagged lines. Animals were infected via the intraperitoneal injection route, and on day two (mice) or five (rats) of infection at a parasitemia of ~1-5 %, the animals were terminally anaesthetised followed by cardiac puncture to collect the P. berghei infected blood. Following transfection parasites were injected intravenously into the tail vein of 8-12 week old female TO mice.

Isolation of Plasmodium schizonts from blood culture

Mouse-derived P. berghei infected erythrocytes were put into culture for 22-24 hours to generate schizonts and these were purified at room temperature as described previously (Janse et al., 2006) with some modifications. Parasite blood cultures were checked for mature schizonts by giemsa staining. Leucocytes were removed from the culture by passing through a Plasmodipur filter (EuroProxima). Blood cultures were pelleted at 300g for 14 min. Supernatants were removed leaving 3 ml behind and the pellets resuspended. The erythrocyte suspension was gently layered onto a Histodenz (Sigma) gradient and centrifuged at 300g for 20 min. Schizont-infected erythrocytes were collected from the brown layer at the interface of the two suspensions and diluted in the supernatant from the parasite blood culture. Isolated schizont-infected erythrocytes were pelleted by centrifugation at 450g for 3 min and washed twice with PBS.

P. falciparum schizonts were freed from erythrocytes by incubation with 0.1% saponin in PBS for 10 minutes and then washed with PBS to remove saponin.
P. knowlesi schizonts were purified by centrifugation at 1500g for 10 minutes onto a 55% (w/v) Nycodenz cushion. They were then washed in RPMI-1640, and then freed from erythrocytes by incubation with 0.1% saponin in PBS, followed by a PBS wash.

**Blue native PAGE of schizont lysates**

Purified *Plasmodium* schizonts were lysed in 50 mM Tris pH 8, 150 mM NaCl, 1 mM EDTA, containing Halt protease and phosphatase inhibitor cocktail (Thermo Scientific), supplemented with 0, 0.1 or 1% NP-40, as previously described (Pardo et al., 2010). The cleared lysate was dialysed into 20 mM Bis-Tris pH 7, 500 mM 6-aminocaproic acid, 12 mM NaCl, 2 mM EDTA, containing Halt protease and phosphatase inhibitor cocktail, supplemented with 0 (for lysates with no NP-40) or 0.1% (for lysates with 0.1 or 1% NP-40), by centrifugal filtration through a 10 kDa cut-off membrane (PES, Vivaspin 500, Sartorius). Protein concentration was determined using the Bradford protein quantification assay.

The dialysed lysate containing 50 μg of protein was separated by Blue Native PAGE in NativePAGE 3-12% Bis-Tris gels as previously described (Bode et al., 2016). Samples were prepared by addition of 4x native PAGE sample loading buffer (Life Technologies) and G-250 sample according to manufacturer’s instructions. Gels were fixed in 40% methanol and 2% acetic acid for 30 minutes and then left in water until further processing.

**In-gel protein digestion**

Gel lanes were each excised into 48 1.5 mm-slices with a grid cutter (THISTLE Scientific) and these placed into a 96-well plate for further processing (Pardo et al., 2017). Proteins were reduced with 5 mM TCEP, followed by alkylation with 10 mM iodoacetamide. After complete gel destaining, proteins were digested with 1 ng/μL trypsin (sequencing grade, Roche). Peptide extraction was performed as described previously (Pardo et al., 2017). Peptide solutions were supplemented with acetonitrile to 60% final concentration, filtered through a 0.65 μm pore membrane plate (Multiscreen HTS DV, Millipore) to remove particulate material and dried. Peptides were resuspended in 0.4% formic acid and 80 mM ammonium bicarbonate and frozen until further analysis.

**P. berghei transfections**

All *P. berghei* 3xHA tagging vectors were obtained from the PlasmoGEM project and details of the constructs are available at http://plasmogem.sanger.ac.uk. The constructs transfected in this study were PbGEM-094925 (PBANKA_093910), PbGEM-281970 (PBANKA_071600) and PbGEM-292320 (PBANKA_094270). For each construct 1-2 μg of NotI-HF (New England Biolabs) digested DNA (MIDI prep, Qiagen) was purified by standard ethanol precipitation prior to transfection. *P. berghei* schizonts were prepared by culturing 22 hours ex vivo, purified on a Histodenz gradient and
transfected using the Fl115 program on the Lonza 4D-Nucleofector core system (with the X Unit) together with the P3 Primary Cell 4D-Nucleofector solution (Lonza), as previously described (Bushell et al., 2017; Janse et al., 2006). Transfectant parasites were selected for and maintained under 0.07mg/mL pyrimethamine (Sigma) administered in drinking water, and were prepared for HA immunoprecipitation without prior dilution cloning.

**HA immunoprecipitation for MS analysis**

Purified *Plasmodium* schizonts (wild type or expressing an HA-tagged protein) were lysed in 50 mM Tris pH 8, 150 mM NaCl, 1 mM EDTA, containing Halt protease and phosphatase inhibitor cocktail (Thermo Scientific), supplemented with 0.1 or 1% NP-40, as previously described (Pardo et al., 2010). Anti-HA antibody (Roche) was coupled to Protein G Dynabeads (Life Technologies). Cleared lysates were incubated with anti-HA coupled Dynabeads for 2 hours at 4 °C. After removing the supernatant beads were washed four times with IPP150 buffer (Pardo et al., 2010) and then 3 times with 50 mM ammonium bicarbonate. Beads were then resuspended in 50 mM ammonium bicarbonate and 1 μg of trypsin (sequencing grade, Roche) was added. Digestion was carried out at 37 °C overnight with constant shaking. Peptide solutions were recovered from the beads, supplemented with acetonitrile to 60% final concentration and filtered through a 0.65 μm pore membrane plate (Multiscreen HTS DV, Millipore). Peptides were then dried, reduced with 40 mM TCEP for 15 minutes at room temperature, acidified with formic acid at 0.5% final concentration and frozen until further analysis.

**LC-MS/MS**

Peptides from blue native fractionation experiments were analysed by online nanoLC-MS/MS on an Orbitrap Velos mass spectrometer coupled with an Ultimate 3000 RSLCnano System. Samples were first loaded and desalted on a nanotrap (100 μm id x 2 cm) (PepMap C18, 5 μ) at 10 μL/min with 0.1% formic acid for 10 min and then separated on an analytical column (75 μm id x 25 cm) (PepMap C18, 2μ) over a 60 min linear gradient of 5 – 42% B (B = 80% CH3CN/0.1% formic acid) at 300 nL/min, and the total cycle time was 90 min. The Orbitrap Velos was operated in standard data-dependent acquisition. The survey scans (m/z 380-1500) were acquired in the Orbitrap at a resolution of 30,000 at m/z 400, and one microscan was acquired per spectrum. The 10 most abundant multiply charged ions with a minimal intensity of 2000 counts were subject to MS/MS in the linear ion trap at an isolation width of 2 Th. Dynamic exclusion width was set at ± 10 ppm for 45 s. The automatic gain control target value was regulated at 1x106 for the Orbitrap and 5000 for the ion trap, with maximum injection time at 200 ms for Orbitrap and 100 ms for the ion trap, respectively.

Peptides from immunoprecipitation experiments were analysed by online nanoLC-MS/MS on an Orbitrap Fusion Tribrid mass spectrometer coupled with an Ultimate 3000 RSLCnano System.
Samples were first loaded and desalted on a nanotrap (100 µm id x 2 cm) (PepMap C18, 5 µ) at 10 µL/min with 0.1% formic acid for 10 min and then separated on an analytical column (75 µm id x 25 cm) (PepMap C18, 2µ) over a 120 min linear gradient of 5 – 40% B (B = 80% CH3CN/0.1% formic acid) at 300 nL/min, and the total cycle time was 150 min. The Orbitrap Fusion was operated in the Top Speed mode at 3 s per cycle. The survey scans (m/z 375-1500) were acquired in the Orbitrap at a resolution of 120,000 at m/z 200 (AGC 4x105 and maximum injection time 50 ms). The multiply charged ions (2-7) with a minimal intensity of 1x104 counts were subject to MS/MS in HCD with a collision energy at 30% and an isolation width of 1.6 Th then detected in the linear ion trap (AGC 1x104 and maximum injection time 35 ms). Dynamic exclusion width was set at ± 10 ppm for 30 s.

Analysis of mass spectrometry data

Raw data files for each blue native PAGE fractionation experiment were analysed together using MaxQuant (version 1.5.5.1) (Cox and Mann, 2008) to identify and quantify proteins across all gel slices. Trypsin was set as digestion mode with a maximum of two missed cleavages allowed. Main search peptide tolerance was set to 20 ppm, and MS/MS match tolerance set to 0.5 Da. Carbamidomethylation of cysteine was set as a fixed modification, and acetylation at the N-terminus, oxidation of methionine, and deamidation of asparagine or glutamine were set as variable modifications. Peptide and protein identifications were set at 1% FDR. Protein identification required at least one peptide for maximum coverage (Table S1). Unique and razor peptides were used for quantification. Database searches were conducted against protein sequence databases from GeneDB (Logan-Klumpler et al., 2012) (P. falciparum – 5431 sequences, P. knowlesi – 5477 sequences, P. berghei – 5019 sequences) for each of P. berghei, P. falciparum and P. knowlesi, and additionally with mouse protein sequences from UniProt (2017)(Mus musculus - 16950 sequences) for samples without saponin treatment. Proteins identified as potential contaminants, reverse hits or mouse proteins were removed for further analysis. The fraction corresponding to the top of the gel was removed before further analysis because of the potential to contain proteins that failed to enter the gel. Intensity scores for a protein in each fraction were normalised by dividing by the sum of protein intensities across all fractions.

Raw data files from HA immunoprecipitation experiments were analysed using Mascot (version 2.4). Database search parameters were as above save the following: oxidation of methionine and acetylation at N-terminus set as variable modifications, peptide tolerance set to 10 ppm, peptide identification set at 1% using Mascot Percolator. SAINTexpress was used to discriminate specific interactions from background binding (Teo et al., 2014). Proteins with SAINT probability score > 0.8 (FDR< 5%) were deemed specific interactors.
The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno et al., 2016) partner repository with the dataset identifier PXD009039 and PXD010117.

**Bioinformatics Analysis**

Bioinformatics analysis was performed mostly using R or Perseus (Tyanova et al., 2016).

To determine whether a protein was running above its expected molecular weight in BN-PAGE a regression line was fit to the native protein markers and their corresponding gel slices. For each quantified protein, its expected Mw was compared against the regression-predicted Mw for the fraction of greatest intensity. A protein was determined to be running above its monomeric Mw if its regression-fitted MW was more than 30% higher than its expected Mw.

Pairwise PPI scores were calculated by taking Pearson correlation scores for each pair of protein profiles in each of the 18 fractionation datasets. Reproducibility between fractionation experiments was assessed by observing the frequency of binned absolute differences between pairwise scores of two datasets.

Hierarchical clustering of protein intensity profiles was performed with Perseus (Tyanova et al., 2016). Profiles of normalised intensities for each protein across all gel slices were assembled into a dendrogram using Pearson correlation as the distance measure with complete linkage and pre-processing with k-means. Data was visualised in a heatmap.

**Scoring of protein interactions by machine learning**

A gold standard set of PPIs was generated using interactions taken from STRING for each of *P. berghei*, *knowlesi*, and *falciparum*. We used only protein interactions derived from experimental evidence (active interaction source “Experiments” only) with a minimum required interaction score of 0.400, and the gene IDs were mapped into the *P. falciparum* orthologue. Interactions from the three species were pooled and duplicate interactions removed, producing a set of 21,671 positive gold standard pairwise interactions. Gold standard negative interactions were produced by taking proteins from each species with the GO component term ‘membrane’ or ‘nucleus’ into separate sets. From the ‘membrane’ set, proteins with descriptions or GO terms containing the patterns ‘RNA’, ‘DNA’, ‘nuclear’, ’nucleolar’, ’chromatin’, ’nucleotide’, ’nuclease’, ’replication’, ’transcription’, ’nucleic’, ’helicase’, ’spliceosom*’, ’ch TEMOSOM*’, ’spindle pole’, ’nucleocytoplasmic’ and ’centrosom*’ were removed. From the ‘nucleus’ set, proteins with putative transmembrane regions were removed, along with proteins containing the terms ‘membrane’, ’extracellular’ and ’cytoplasm’ in their annotation. Negative Interactions (112,032) were generated as a combination of the proteins between these sets.
205 interactions overlapped between the positive and negative sets, and were thus removed from the negative set.

The Random Forest classifier was used to elucidate high confidence PPIs from multiple data sources using the randomForest R package (Liaw and Wiener, 2001). We calculated Pearson correlation scores for each protein pair in each of the 18 fractionation datasets, and averaged them between biological replicates for each species and detergent concentration to provide 9 biochemical fractionation features, one for each experimental condition. Protein pairs with a correlation score of 0.4 or above were used in the classifier. The number of times a protein pair was detected with this criteria across all datasets was included as a feature for the classifier. The number of times a pair of proteins showed maximum intensity in the same gel fraction was also included as a feature. Additional features included domain interactions from DOMINE (Raghavachari et al., 2008; Yellaboina et al., 2011), gene co-expression (Bozdech et al., 2003; Hu et al., 2010; Modrzynska et al., 2017), co-evolution scores (Juan et al., 2008; Ochoa et al., 2015) and growth phenotype data from PlasmoGEM (Bushell et al., 2017). Only feature information pertaining to protein pairs already present in the fractionation datasets was included. Features were formatted to a binary or continuous score. For DOMINE, known domain interactions were assigned a score of 1, and high, medium and low confidence interactions from computational sources were assigned scores of 0.75, 0.5 and 0.25 respectively. To map these domain scores onto proteins, the highest score associated with domain interactions between two proteins was used. Gene co-expression scores were calculated by Pearson correlation of transcriptomic profiles from drug-induced growth perturbations (Hu et al., 2010), AP2 knockouts (Modrzynska et al., 2017) and gene expression across the asexual blood stages (Bozdech et al., 2003) in Plasmodium species. Co-evolution scores were taken as the output of pMirror-Tree (pMT) (Ochoa et al., 2015) and Context Mirror methodologies (Juan et al., 2008), using PPIs with p-values less than 10e-5 for pMT and PPIs from Context Mirror level 10 with p-values less than 10e-6.

Growth phenotypes (Bushell et al., 2017) were converted to a binary score, where pairs of proteins with the same phenotype ontology were placed in one category and those with differing ontology into another. Protein pairs found common to both the resultant data set and the gold standard were divided evenly into training and test datasets. The random forest classifier was run against the training dataset with 500 trees and allowed to choose from 4 random features at each branch split, and its performance assessed by ROC curve analysis on the test dataset. The contribution of each feature to the classifier was assessed by the mean decrease in Gini score across all trees. Protein pairs with a probabilistic output score of 0.9 or above were considered to represent putative interactions. Additional filtering criteria or raising the probabilistic score cut-off resulted in loss of experimentally-validated interactions from the network.

Receiver Operating Characteristic (ROC) curves using the gold standard were used to assess the quality of PPI datasets. To assess the validity of derived clusters, a cluster was considered to have
captured an annotated protein complex from STRING or CORUM if at least half of its proteins had interactions with proteins from the same cluster that were present in these datasets. REACTOME, BioGrid and STRING pathway/interaction databases were used to assess the quality of derived interactions. Interactions were taken from these databases were treated as undirected, and mapped into \textit{P. falciparum} gene IDs where possible. The extent to which derived interactions overlapped with these datasets was then observed.

Protein interactions were visualised as networks using Cytoscape version 3.4.0 (Shannon et al., 2003). ClusterONE was used to derive clusters from networks based on cohesiveness (Nepusz et al., 2012). Settings were chosen such as to capture the greatest number of known complexes.

A conserved \textit{Plasmodium} PPI network was derived from the high confidence probabilistic PPI network by including edges with co-migration scores present in at least one fractionation dataset from each of the three species.

Computed Gene Ontologies for process, function and component were retrieved from PlasmoDB (Aurrecoechea et al., 2009) for each species. GO term enrichment analysis for whole BN-PAGE datasets was performed with R using a Fisher’s exact test with Bonferroni multiple testing correction against each \textit{Plasmodium} species annotated genome as reference set. GO term enrichment analysis for clusters used the set of proteins seen in all clusters for a given network as reference. Clusters were deemed to recapitulate STRING or CORUM complexes if more than 50% of cluster members were subunits of a given complex. Enrichments were also calculated for asexual blood stage parasite growth phenotype using data from PlasmoGEM (Bushell et al., 2017).

\textit{Calculation of relative evolutionary rates and gene co-evolution}

Co-evolution between pairs of proteins was determined using the pMirror Tree (pMT) (Ochoa et al., 2015) and ContextMirror (Juan et al., 2008) methodologies. 3786 OrthoMCL protein families containing at least the species \textit{P. falciparum}, \textit{berghei}, \textit{yoelii}, \textit{knowlesi}, \textit{chabaudi} and \textit{vivax} were taken and protein FASTA sequences for all homologues retrieved. For each protein family, multiple sequence alignments were made using MUSCLE (Edgar, 2004a, b), and distance matrices produced with Protdist from PHYLIP, using the Probability Matrix from Blocks model for amino acid substitution (Veerassamy et al., 2003). For protein-protein co-evolution scores, distances matrices for each protein family were fed to pMT with the following settings: 40 groups, 0.05\% chance of branch switching and 1000 branch switching iterations. As an alternate measure, distance matrices were fed to ContextMirror and level 10 protein pairs with \textit{p}-values less than 10e-6 were determined to be co-evolved. To assign relative evolutionary rates, an average distance was calculated for each species to \textit{P. falciparum} across all distance matrices. For each protein family, the average distance for all species to \textit{P. falciparum} was calculated, and divided by the average of the average distances for the same
subset of species pairs. Relative evolutionary rates were used to compare protein sub-populations within interaction datasets by comparing their distributions using a Wilcoxon-Mann-Whitney test.

SUPPLEMENTAL INFORMATION TITLES AND LEGENDS

Figure S1. Related to Figure 1. Description of the BN-PAGE fractionation data. (A) Number of proteins detected for each species and experimental condition. Bars are coloured by percentage of detergent in the lysis buffer. Protein counts for pairs of replicates are shown as a union. (B) Comparison of positive and negative GO component term enrichments between two *P. knowlesi* fractionation experiments in 0% and 1% NP-40 lysis conditions. Significance was assessed using Fisher’s exact test with Bonferroni correction. The top 10 GO terms for each condition and enrichment with p values less than 0.05 are shown. The dashed line represents a p value of 0.05. (C) Migration profiles of membrane protein complexes only identified from lysates containing NP-40. (D) Number of detected proteins running above their expected BN-PAGE migration distance. Proteins with a predicted Mw at 30% higher than their literature Mw based on a regression model of known markers are considered to be running above their expected Mw. Error bars represent one standard deviation.

Figure S2. Related to Figure 1. Reproducibility of BN-PAGE fractionation. Histograms of absolute pairwise distances between co-migration scores of common pairs from (A) replicate lysis conditions (*P. falciparum* 0.1% NP-40) and (B) interspecies datasets (*P. berghei* 1% NP-40 and *P. falciparum* 0% NP-40).

Figure S3. Related to Figure 2. Evaluation of the high-confidence protein interaction network. Overlap between protein interactions from the random forest classifier output (grey) and PPIs annotated in public databases.

Table S1. Related to Figure 1. List of all proteins identified by LC-MS/MS with total spectral counts across all fractions per sample. Accession numbers are from PlasmoDB database (*P. berghei*). Column headers indicate the experimental sample.

Table S2. Related to Figure 1. GO term enrichment (positive and negative) analysis for each *Plasmodium* species and lysis condition. The enrichment was calculated using a Fisher’s exact test with Bonferroni correction.

Table S3. Related to Figure 2. Gini scores for features used in random forest analysis. Gini scores measure the relevance of the feature in the classification of the data. References indicate the source of the feature data used in the classifier.

Table S4. Related to Figure 2. List of proteins (nodes) in the *Plasmodium* protein interaction network. This table can be used to import attributes to the proteins in the interaction network when using network visualisation programs.
Table S5. Related to Figure 2. Pairwise interactions in the Plasmodium protein interaction network. The table includes scores for all the features used in the random forest classifier, in addition to other useful information, such as annotation in interaction databases for human, mouse, worm and fly orthologs, and whether the interaction pair has biochemical evidence (blue native fractionation) from only one Plasmodium species. This table can be used to import the protein interaction pairs together with attributes of the interaction (edge attributes) into network visualisation programs.

Table S6. Related to Figure 2. Clusters from the Plasmodium interaction network representing putative protein complexes. Clusters were derived using ClusterOne (Nepusz et al., 2012). Conserved Plasmodium proteins of unknown function are highlighted in red.

Table S7. Related to Figure 2. GO term enrichment and CORUM complex recapitulation of clusters. The GO_BPterms_enriched_clusters sheet lists clusters enriched in specific GO_Biological Process terms (where at least 50% of proteins in the cluster share a particular GO term). The Proteins_GO_BPenriched_clusters lists the proteins of the clusters enriched in specific GO_Biological Process terms. The CORUM clusters sheet lists clusters that recapitulate CORUM complex(es) (where at least 50% of proteins in the cluster belong to a particular CORUM complex).

Table S8. Related to Figures 2 and 3. Function prediction for selected conserved Plasmodium proteins of unknown function. The prediction was based on the protein being part of a cluster enriched in a particular GO term and/or CORUM complex, supported in some cases from other evidence indicative of a physical association.

Table S9. Related to Table 1 and Figure 3. Interaction validation through affinity purification coupled to mass spectrometry. The first sheet contains protein identifications from AP-MS experiments on selected candidates and controls. # Peptides, number of unique peptide sequences; # PSM, number of peptide spectrum matches. The second sheet contains the results of the SAINTexpress analysis. SP, SAINT probability; FC, fold change.

Table S10. Related to Figure 4. Conserved Plasmodium predicted complexes. The conserved network was derived from the overall network by including only edges (protein pairs) present in at least one fractionation dataset from each of the three species.
Fig. S1. Related to Figure 1. Description of the BN-PAGE fractionation data. (A) Number of proteins detected for each species and experimental condition. Bars are coloured by percentage of detergent in the lysis buffer. Protein counts for pairs of replicates are shown as a union. (B) Comparison of positive and negative GO component term enrichments between two P. knowlesi fractionation experiments in 0% and 1% NP-40 lysis conditions. Significance was assessed using Fisher’s exact test with Bonferroni correction. The top 10 GO terms for each condition and enrichment with p values less than 0.05 are shown. The dashed line represents a p value of 0.05. (C) Migration profiles of membrane protein complexes only identified from lysates containing NP-40. (D) Number of detected proteins running above their expected BN-PAGE migration distance. Proteins with a predicted Mw at 30% higher than their literature Mw based on a regression model of known markers are considered to be running above their expected Mw. Error bars represent one standard deviation.
Fig. S2. Related to Figure 1. Reproducibility of BN-PAGE fractionation. Histograms of absolute pairwise distances between co-migration scores of common pairs from (A) replicate lysis conditions (P. falciparum 0.1% NP-40) and (B) interspecies datasets (P. berghei 1% NP-40 and P. falciparum 0% NP-40).
Fig. S3. Related to Figure 2. Evaluation of the high-confidence protein interaction network. Overlap between protein interactions from the random forest classifier output (grey) and PPIs annotated in public databases.