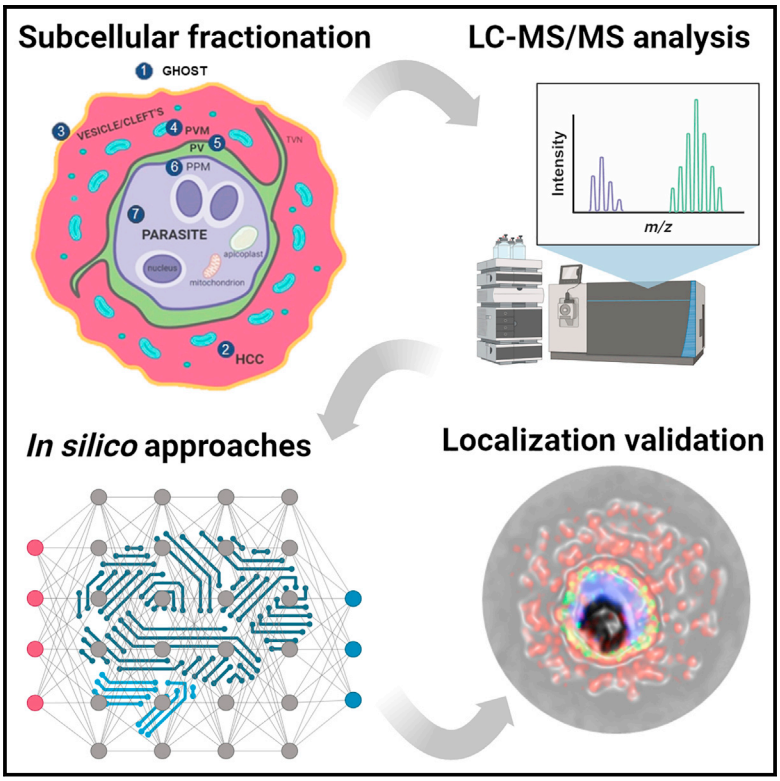
Cell Reports

Resource

Comparative spatial proteomics of *Plasmodium*-infected erythrocytes

Graphical abstract



Authors

Anthony Siau, Jing Wen Ang, Omar Sheriff, ..., Sze Siu Kwan, Marek Mutwil, Peter R. Preiser

Correspondence

prpreiser@ntu.edu.sg

In brief

Siau et al. examine the spatial proteome from different human, simian, and rodent *Plasmodium* species. Two approaches are employed to provide predictive localization for previously unassigned proteins. This represents a powerful tool to uncover conserved and diverged strategies deployed by the parasites to adapt to its hosts.

Highlights

- Predictive localization for ~8,500 previously unassigned parasite proteins
- Validation of 93 orthology groups supporting the consensus localization prediction strategy
- Increased species-specific protein novelty toward infected RBC periphery
- Increased m^{6}A modification of mRNA transcripts associated with peripheral proteome

Siau et al., 2023, Cell Reports *42*, 113419
November 28, 2023
https://doi.org/10.1016/j.celrep.2023.113419

CellPress

Cell Reports

CellPress
OPEN ACCESS

Resource

Comparative spatial proteomics of *Plasmodium*-infected erythrocytes

Anthony Siau,[1,2,3] Jing Wen Ang,[1,2] Omar Sheriff,[1] Regina Hoo,[1] Han Ping Loh,[1] Donald Tay,[1] Ximei Huang,[1] Xue Yan Yam,[1] Soak Kuan Lai,[1] Wei Meng,[1] Irene Julca,[1] Sze Siu Kwan,[1] Marek Mutwil,[1] and Peter R. Preiser[1,4,*]

[1]Nanyang Technological University, School of Biological Sciences, Singapore 637551, Singapore
[2]These authors contributed equally
[3]Deceased August, 2023
[4]Lead contact
*Correspondence: prpreiser@ntu.edu.sg
https://doi.org/10.1016/j.celrep.2023.113419

SUMMARY

Plasmodium parasites contribute to one of the highest global infectious disease burdens. To achieve this success, the parasite has evolved a range of specialized subcellular compartments to extensively remodel the host cell for its survival. The information to fully understand these compartments is likely hidden in the so far poorly characterized *Plasmodium* species spatial proteome. To address this question, we determined the steady-state subcellular location of more than 12,000 parasite proteins across five different species by extensive subcellular fractionation of erythrocytes infected by *Plasmodium falciparum, Plasmodium knowlesi, Plasmodium yoelii, Plasmodium berghei*, and *Plasmodium chabaudi*. This comparison of the pan-species spatial proteomes and their expression patterns indicates increasing species-specific proteins associated with the more external compartments, supporting host adaptations and post-transcriptional regulation. The spatial proteome offers comprehensive insight into the different human, simian, and rodent *Plasmodium* species, establishing a powerful resource for understanding species-specific host adaptation processes in the parasite.

INTRODUCTION

After invasion, the *Plasmodium* parasite extensively modifies its host erythrocyte. The parasite places highly specialized compartments or structures in the host cell by secreting, beyond the parasite plasma membrane (PPM), a range of proteins that constitute the remodelome. Some proteins are translocated to the parasite periphery, either in the parasitophorous vacuole (PV) and/or the PV membrane (PVM), two subcellular compartments conserved across the genus. The remainder are exported beyond the PVM and can be found (1) uniformly distributed as soluble proteins in the host cell cytosol (HCC), (2) localized in specialized vesicles/complexes in the HCC, or (3) associated/inserted into the erythrocyte membrane.[1] The role of the extra-parasitic compartments as well as the nature of the exported parasite proteins remain largely elusive, especially for non-*Plasmodium falciparum* species. This lack of information about definitive cellular localization poses a significant challenge to *in silico* gene function predictions and experimental study of protein interactions.[2–4] Additionally, efforts to identify new therapeutic targets against malaria is hindered by the lack of functional annotations for almost one-third of the parasite-specific proteome.[2]

In silico screens based on discriminative feature(s) have led to the identification of the *Plasmodium* translocon of exported proteins (PTEX) complex, the *Plasmodium* export element (PEXEL), and the vacuolar targeting sequence (VTS), shedding light on the mechanism used to translocate malaria proteins beyond the PVM,[5] as well as a repertoire of 300–400 *P. falciparum*-exported proteins.[6–9] These datasets do not reflect the full extent of the exported protein repertoire, and other predictive approaches were used to identify additional PV proteins in *Plasmodium berghei*[10] and PEXEL-negative exported proteins in *P. falciparum*.[11] Nevertheless, these approaches were limited in their ability to reliably identify remodeling proteins and required extensive individual validation, making the current predicted *Plasmodium* exportome incomplete, as indicated by the increasing number (>60) of PEXEL-negative exported proteins reported.[12] For other non-*P. falciparum* species, host cell remodeling events are even more elusive, with only a few PEXEL proteins being predicted.[7] These species are characterized by several variant protein families found at the parasite periphery and/or exported into the host cell.[13–16] The identification of a new export motif, PLASMED, in *Plasmodium yoelii*[14] indicates that *Plasmodium* uses several export motifs, and our current knowledge of the parasite exportome remains inadequate. Moreover, multigene protein families, such as *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) and *Plasmodium* interspersed repeat (PIR), are not constitutively expressed because of tight regulatory mechanisms[17–26] that could not be captured by any predictive model.

To overcome the limitations of *in silico* approaches, alternative proteomic analysis of parasite samples enriched for specific subcellular compartments expanded the known remodelome in the PV,[27] the red blood cell (RBC) membrane,[13,28,29] and the parasite-induced HCC structure.[30,31] Immunoprecipitation and proximity labeling identified selected protein complexes in specific subcellular compartments and provided additional information

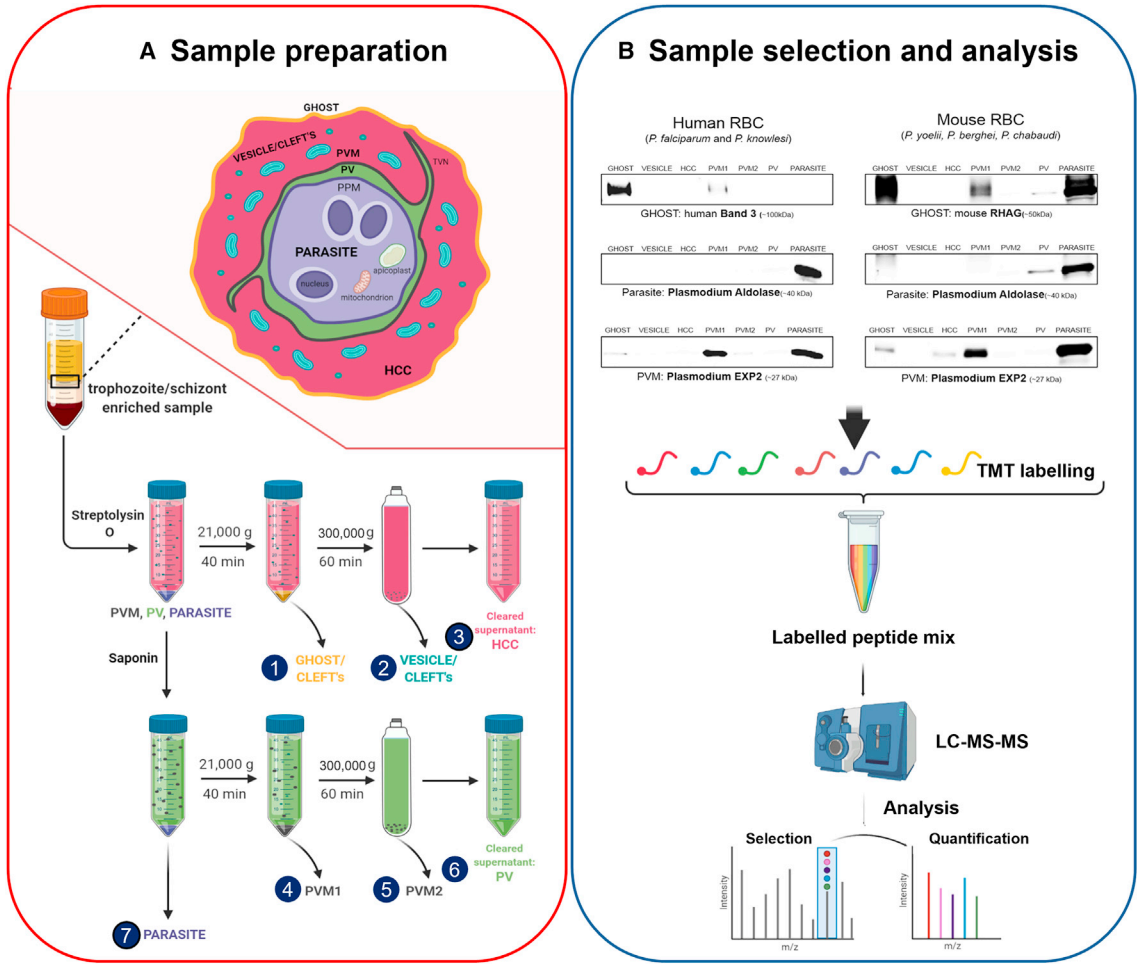


Figure 1. Subcellular fractionation of infected RBCs

(A) Workflow of the subcellular fractionation method used to map the whole *Plasmodium* exportome. Enriched late trophozoite/early schizont parasite samples were treated with SLO and, subsequently, saponin. Centrifugation was performed to obtain a supernatant and a pellet fraction after each treatment. The supernatants were then centrifugated to recover six subcellular fractions (1–6). The pellet obtained after saponin lysis contains the internal parasite fraction (7). In total, seven fractions were produced: GHOST, VESICLE, HCC, PVM1, PVM2, PV, and internal PARASITE.

(B) Representative western blot of a human infected RBC pellet (right) or mouse infected RBC (left) probed with antibodies against the PVM marker EXP2 (~27 kDa), the internal parasite protein aldolase (~40 kDa), and the human RBC membrane proteins band 3 (~100 kDa) or mouse rhesus blood group-associated glycoprotein (RHAG; ~50 kDa). The antibodies used are indicated below each western blot. The fractionated infected RBC samples displaying minimal cross-contamination between subcellular fractions were selected for LC-MS/MS analysis following TMT labeling.

about the interactomes of PTEX components,[32,33] PVM protein,[34] chimeric proteins targeted to the PV[35] or engineered to be arrested in their translocation process across the PVM,[36,37] in the J-dot,[38] as well as the RBC membrane.[39,40]

Together, *in silico* and experimental approaches expanded our knowledge but were unable to provide a comprehensive understanding of the highly dynamic host remodeling process. Moreover, these efforts are impeded by the large amount of contaminating intracellular proteins, reducing the extent of the findings and requiring extensive validation. In this study, we addressed those deficiencies by performing extensive subcellular fractionation and quantitative proteomics on infected RBCs to capture the steady-state localization of proteins found in five *Plasmodium* species independent of their protein characteristics. The proteomics datasets were mapped using two different prediction approaches, and the findings were compiled into an orthology map to derive consensus predictions. The analysis of the spatial proteome not only helps prioritize subcellular compartments with therapeutic potential but also reveals discrete remodeling features between primate and rodent parasites. Furthermore, it highlights the contribution of merozoite proteins to the interface of host-parasite interactions.

RESULTS

Subcellular fractionation of five *Plasmodium*-infected erythrocytes

Infected RBC samples of five *Plasmodium* species were sequentially lysed and differentially centrifuged to produce a comprehensive and resolved protein atlas (Figure 1A).[15,41] The human

parasite *P. falciparum* is responsible for most of the death associated with malaria. The simian parasite *Plasmodium knowlesi* is a zoonotic pathogen responsible for thousands of human infection cases yearly,[42] while the rodent surrogates *P. yoelii, P. berghei*, and *Plasmodium chabaudi* are easily amenable to genetic manipulation[43] and provide *in vivo* insight. Overall, seven fractions enriched for six subcellular compartments were obtained: (1) the RBC membrane (GHOST), (2) the HCC-specialized structure (VESICLE), (3) the HCC-soluble protein (HCC), (4) the PVM (PVM1 and PVM2), (5) the PV (PV), and (6) the internal parasite (PARASITE) (Figure 1A).

The quality and integrity of the subcellular fractions were assessed by western blot probed with antibodies against the human RBC membrane protein band 3 (*P. falciparum* and *P. knowlesi*), the mouse rhesus blood group-associated glycoprotein (RHAG) (*P. yoelii*, *P. berghei*, and *P. chabaudi*), the intra-parasitic protein aldolase, and the PVM protein EXP2 (Figure 1B). The large amount of band 3 (~100 kDa; Figure 1B, left) and RHAG (~50 kDa; Figure 1B, right) protein detected in the GHOST fractions indicated that streptolysin O (SLO) treatment lysed most of the RBC membranes and that these membranes are enriched in the GHOST fractions following differential centrifugation. Conversely, the non-exported aldolase (~40 kDa) was mainly detected in the PARASITE samples, indicating that the integrity of the PPM was largely preserved during the sequential lysis. Finally, EXP2 (~27 kDa) was detected mostly in the PVM (PVM1) and PARASITE fractions, in line with previously reported localization data.[27] This indicates that the PVM was mostly preserved during SLO treatment and then released into the supernatant after saponin treatment. Altogether, it validated that the subcellular fractionation achieved selective lysis of the RBC membrane and the PVM, while the integrity of the PPM was preserved.

Mapping of five *Plasmodium* proteomes

To accurately quantify the proteome of the subcellular fractions obtained from differential lysis, samples from three (*P. falciparum* and *P. knowlesi*) and two (*P. yoelii*, *P. berghei*, and *P. chabaudi*) independent experiments were analyzed by quantitative liquid chromatography-tandem mass spectrometry (LC-MS/MS).[44] Compiling peptide hits from 12 (3 + 3 + 2 + 2 + 2) parasite samples, 12,392 parasite proteins were identified across the 12 × 7 = 84 fractions. The 2,796 *P falciparum*, 2,677 *P knowlesi*, 2,780 *P yoelii*, 2,246 *P berghei*, and 1,893 *P chabaudi* proteins were quantified by comparing, for each protein, the intensities of the reporter ions in the fraction with the sum of the intensities across the fractions (see Table S1 for the composition of the infected RBC samples).

While the quantitative distribution of the different proteins in the subcellular fractions is a valuable indicator of their locations, relying solely on this information may be suboptimal because of the complexity of the study, which spans 84 fractionated samples, and the need to distinguish relative abundance patterns that are vastly similar. Further complicating the task, many proteins are present simultaneously with various abundance in the intra-parasitic and extra-parasitic compartments, with some of these proteins being exported into the host cell in a stage-specific manner.[45]

To overcome this, two alternative strategies, iterative Permutation (P) and machine learning (ML), were trained using datasets generated by searching the literature for localization information of proteins detected in the *Plasmodium* proteomes (Table S1, training sets). The approaches were designed to maximize the number of remodeling proteins identified without compromising the accuracy in the five *Plasmodium* species (Table S1, P and ML; Figure 2A). To evaluate the performance of the P and ML approaches, the proportions of training proteins that were correctly predicted were determined for each subcellular localization and then combined to determine overall accuracy (Figure 2B). Of note, no protein was predicted in the PVM2 fraction. The P and ML approaches displayed comparable performance, with accuracy ranging from 20%–100% for the extra-parasitic compartments and 88%–97% for the internal parasite fraction. The overall accuracy was greater than 85% in all species (Figure 2B). Interestingly, the low 29% accuracy predicted by the P approach in the *P. yoelii* clefts and the 20% correlation predicted by the ML approach in the *P. berghei* PVM were complemented by the 57% and 80% accuracies obtained with their cognate counterparts, indicating that a combination of both predictions can greatly help improve localization accuracy (Figure 2B).

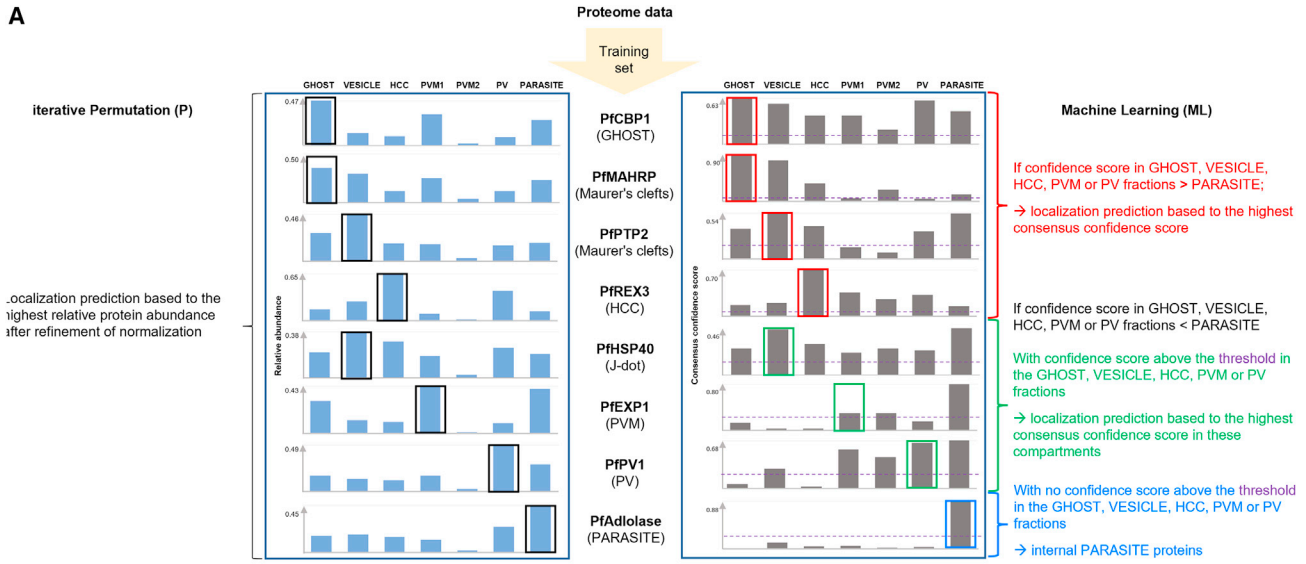
Determination of a consensus localization

The *in silico* approaches were designed to provide accurate localization predictions. However, variations in the predictions are expected because of biological differences, experimental variations, as well as the quality of the training sets used. Because orthologs are expected to share the same localization, interspecies variations can be reduced by inferring a consensus localization using the P, ML, or a combination of P and ML (P+ML) predictions. For this, an orthology map of 4,350 orthology groups annotated with their respective localization predictions was generated from the 12,392 parasite proteins identified by the proteomics approach (Table S2, proteome mapping). To derive a consensus localization for each orthology group, an iterative three-step process was used to provide consensus localizations and derive prediction confidence (Figure 3A; Table S2, proteome mapping, columns AM, AO, and AQ).

Overall, P, ML, and P+ML consensus predictions displayed a high level of correlation, with ~78% of the orthology groups having an identical prediction (P = ML = S). For 5% of the groups, P+ML predictions complemented the absence of prediction by one of the two approaches (P = ND and/or ML = no consensus prediction [ND]). In another 5% of the groups, the consensus process favored one of the two predictions based on the most frequent location count (ML = S ≠ P and P = S ≠ ML). The remaining differences resulting from discordant predictions in the P and ML approaches caused an absence of prediction in P+ML (S = ND, 8%) or resulted in a different prediction in P+ML (P ≠ S ≠ ML, <1%) (Figure 3B).

Validation of the pan-parasite protein atlas

One hundred orthology groups selected by random draw were characterized in subcellular location studies using protein fusion tagging, irrespectively of their known localizations or predicted features. Ninety orthology groups encoding genes of 3.5 kb or less were assessed using episomal GFP tagging, while the remaining



B

	Localization	*P. falciparum*			*P. knowlesi*			*P. yoelii*			*P. berghei*			*P. chabaudi*		
		P	ML	n	P	ML	n	P	ML	n	P	ML	n	P	ML	n
LITERATURE LOCALIZATION DATA	**GHOST**	74%	82%	65	65%	70%	43	63%	69%	48	60%	77%	43	70%	74%	46
	VESICLE	54%	46%	13	44%	22%	9	92%	83%	12	75%	75%	8	67%	78%	9
	HCC	63%	81%	16	97%	94%	31	86%	93%	29	71%	57%	14	64%	86%	14
	GHOST+VESICLE	80%	70%	30	60%	50%	10	29%	57%	7	50%	50%	6	60%	67%	6
	VESICLE+HCC	100%	67%	3				67%	67%	6	67%	75%	12	50%	33%	6
	PVM/PV/PPM (periphery)	94%	86%	35	83%	83%	12	94%	100%	35	97%	83%	30	86%	86%	28
	PVM	75%	33%	12	67%	44%	9	70%	80%	10	80%	20%	10	63%	75%	8
	PV	87%	87%	15	83%	83%	12	71%	86%	14	83%	58%	12	67%	42%	12
	PARASITE	92%	91%	492	93%	93%	468	94%	94%	483	90%	97%	437	88%	94%	366
PROTEIN FAMILIES	**GHOST/VESICLE/HCC** (fam/pyst, tryptophan-rich protein, Pfrifin, Pfstevor, Pfmc-2tm, Pf PEXEL protein, PkSICAvar)	67%	75%	85	91%	87%	53	95%	95%	98	88%	91%	32	96%	91%	95
	PVM/PV (etramp, sera)	86%	100%	7	100%	100%	2	80%	100%	5	100%	50%	4	83%	92%	12
	VESICLE/HCC/PVM/PV (pirs)				93%	86%	14	96%	99%	166	69%	50%	16	56%	38%	16
	VESICLE/HCC (non-*falciparum* PEXEL protein)				79%	64%	39	61%	78%	18	100%	100%	43	75%	75%	4
	GHOST (*P. chabaudi* erythrocyte membrane antigen 1)													100%	100%	6
	GHOST/VESICLE (PfEMP1)	50%	100%	8												
	Total	85%	85%	781	88%	87%	702	90%	92%	931	87%	90%	667	85%	87%	628

Figure 2. *In silico* mapping of *Plasmodium* proteomes

(A) Principle of the iterative permutation (P) and machine learning (ML) approaches used to map the *Plasmodium* proteomes. The localization of the proteins was deduced either based on their relative abundance or mean confidence score patterns across the seven subcellular fractions. The patterns obtained for eight representative proteins expected to be enriched in each of the subcellular fractions are displayed, along with their known localization. Maurer's cleft proteins are enriched either in the GHOST or VESICLE fractions, as shown for PfMAHRP and PfPTP2. For each protein, the threshold used for the ML method is defined by dividing the confidence score in the PARASITE fraction by a value defined for the whole dataset. The latter is calculated in order that more than 90% of the PARASITE proteins have confidence scores in the remodeling fractions that fall below the threshold as for PfAldolase.

(B) Evaluation of the P and ML approaches to predict the localization of proteins characterized in the literature (top) or known multigene protein families (bottom). Expected localization outputs are indicated in the left column. For each species, the composition of the training set given by the number of proteins is indicated in the column "n," while the "P" and "ML" columns depict the percentage of correlation between the prediction approach and the localization data for each subset of proteins associated with a specific subcellular localization. For each localization, the highest percentage of correlation between the P and ML approaches is highlighted in red.

groups that are mostly encoded by larger genes were evaluated using endogenous hemagglutinin (HA) tagging. For each ortholog group, one or two protein tagging attempts were made, using *P. falciparum*, *P. knowlesi*, and/or *P. yoelii* parasites (Table S3). Of more than130 transfections, 18 *P. falciparum*, 21 *P knowlesi*, and 44 *P yoelii* transfectant lines expressing C-terminally tagged protein were obtained (Figure S1). These 83 proteins represent 77 ortholog groups, with six groups having localization data obtained from two species (Figure 4A). Among these, 56 groups had no known localization, while 21 groups had information in the literature, mostly from *P. falciparum* and *P. berghei*. Localization of the tagged chimeras was ascertained using live-cell

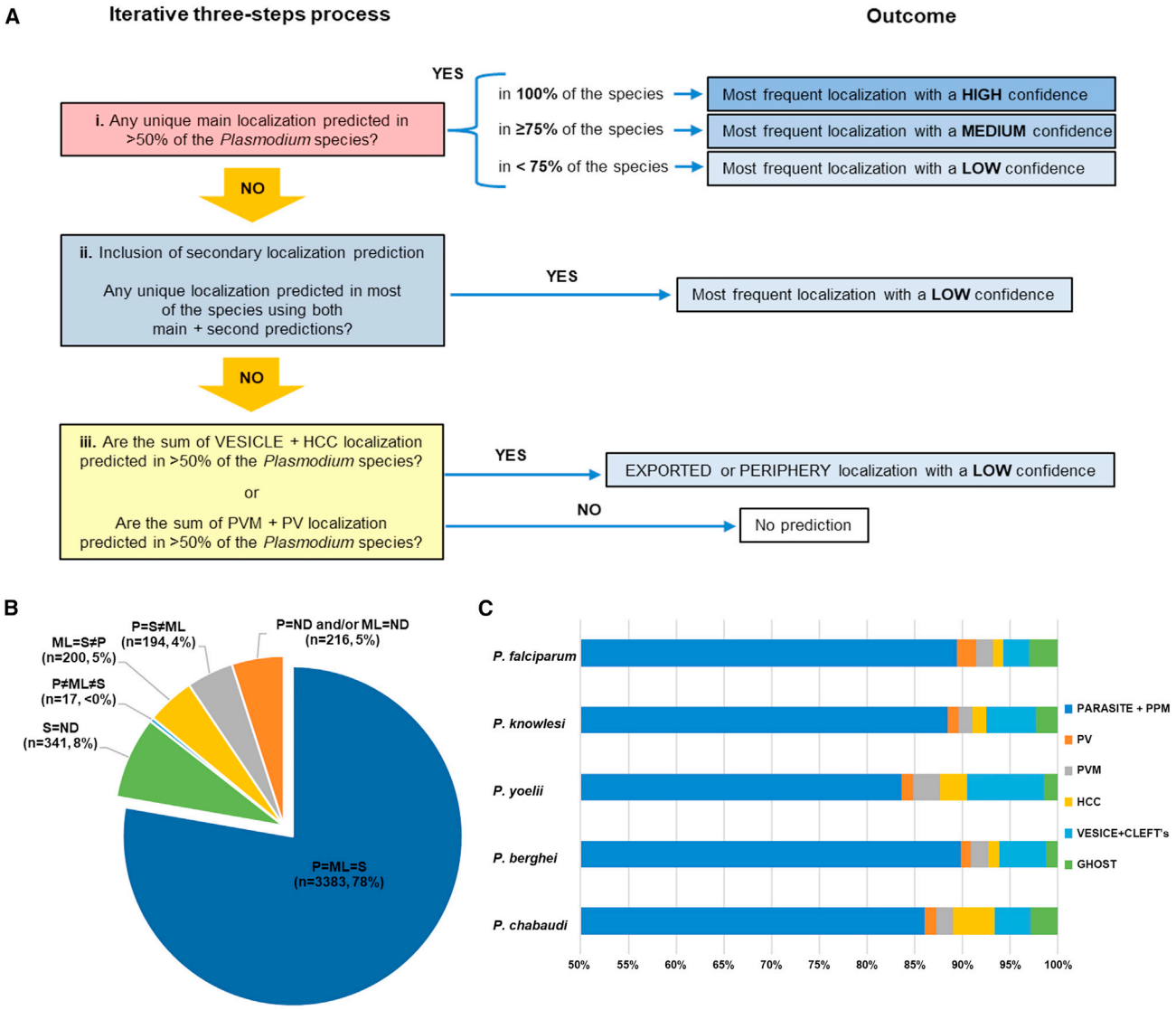


Figure 3. Determination of a consensus localization using gene orthology information

(A) The predictions obtained using the P and ML approaches were combined in the orthology map depicted in Table S2 and used to derive a consensus localization shared by each orthology group using an iterative three-steps process pipeline. (i) Orthology groups with a subcellular localization shared by more than 50% of the predictions were annotated with this specific localization, along with confidence levels of low, medium, and high when the localizations were shared by more than 50% and less than 75%, more than 75% and less than 100%, and 100% of the predictions, respectively (Table S2, proteome mapping, columns AM, AO, and AQ). (ii) Because experimental variations might affect the outcome when predictions are within the same range of relative abundance (P) or confidence scores (ML), secondary localizations were used. For the P approach, it consists of the subcellular compartment with the second-highest abundance, while for the ML approach, we considered the second compartment with the highest confidence scores greater than the background threshold. Only secondary localizations found within 20% of the main localization's relative abundance or confidence scores were used to refine predictions considered to have a low confidence level (Table S1, P and ML; Table S2, proteome mapping). (iii) The identification of Clefts' proteins may be suboptimal because they are expected to be enriched in the GHOST and VESICLE fractions. Unannotated orthology groups were annotated as "Cleft's" with a low confidence level when GHOST and VESICLE localization represented more than 50% of the predictions included in the orthology group.

(B) Comparison of the consensus prediction obtained using P, ML, or P+ML (S) predictors. ND, no consensus prediction.

(C) Composition of the *Plasmodium* protein atlas following localizations derived from the compilation of experimental validations and P+ML predictions.

fluorescence microscopy on mixed-stage transfectants. Four localization patterns could be distinguished by combining the microscopy pictures with the localization data available in the literature and retention of the most external localization (Figure 4B).

The performance of the P, ML, and P + ML approaches was determined by comparing the localization data of the 77 orthology groups with their localization predictions. For proteins found in multiple localizations by microscopy, predictions were considered

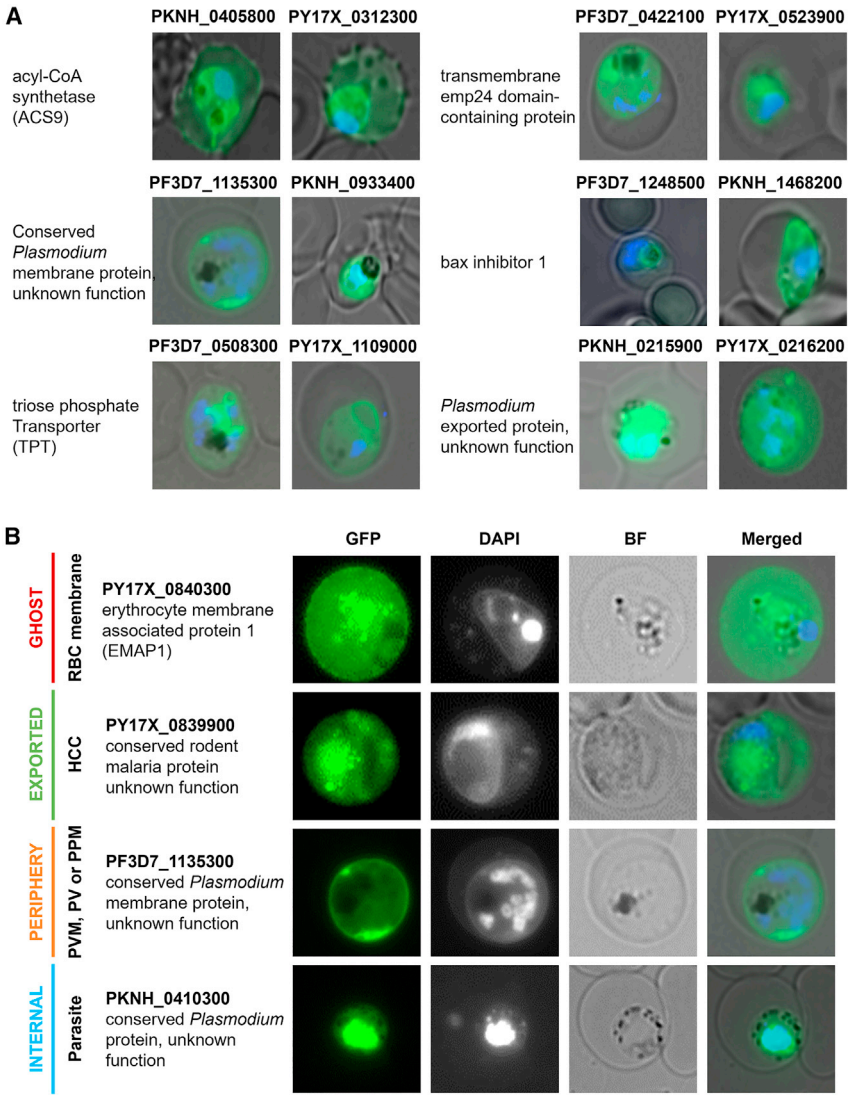


Figure 4. Live-cell fluorescence microscopy of GFP-tagged proteins

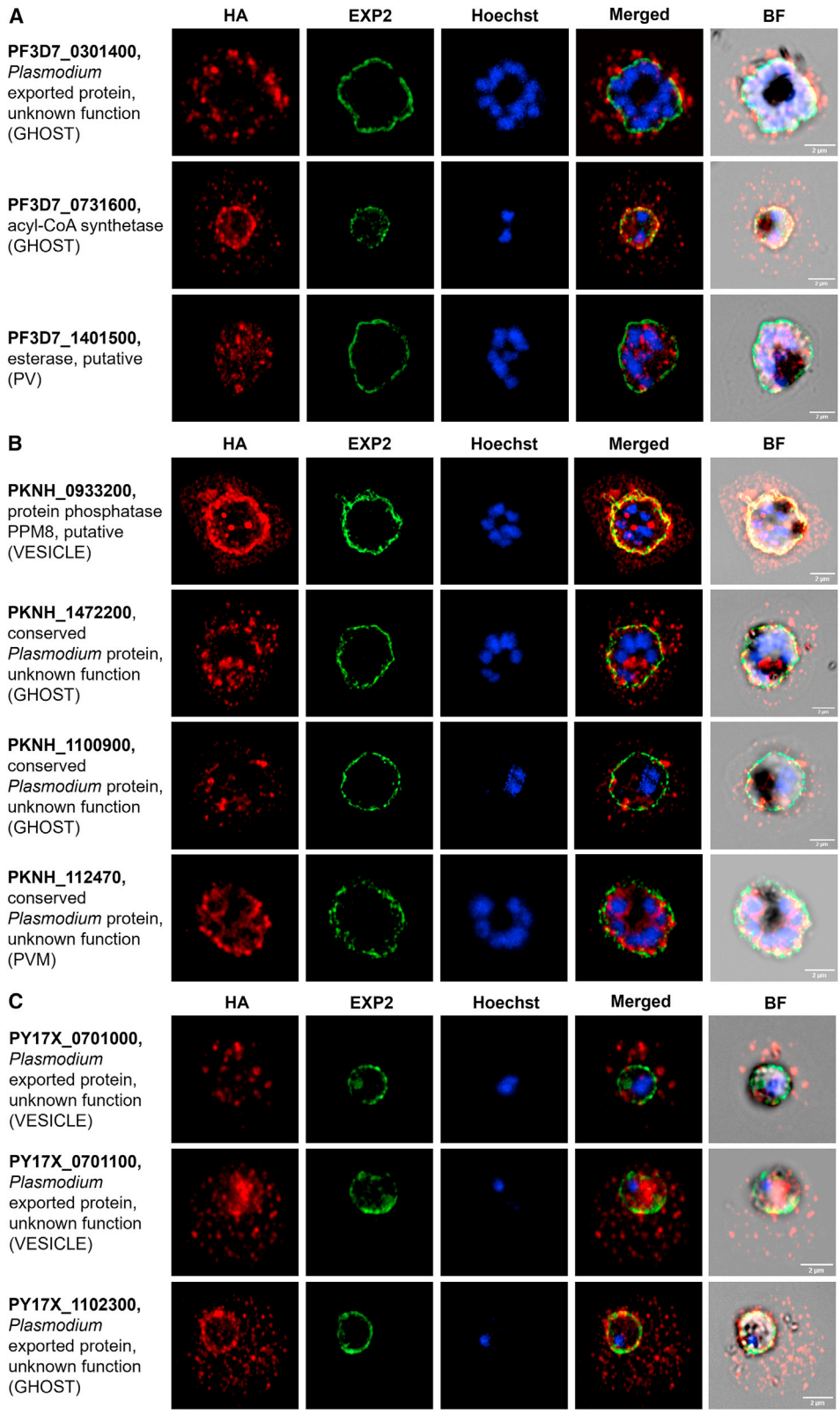
Representative images of *P. falciparum*-, *P. knowlesi*-, and *P. yoelii*-infected RBCs depicting (A) proteins from six orthology groups having localization data obtained from two parasite species and (B) the four localization patterns (ghost, exported, periphery, or internal) of the GFP-tagged protein (green), distinguished by microscopy imaging. The parasite nucleus was stained with DAPI (blue).

correct when they corresponded to one of the predicted localizations for the orthology group (Table S3). The locations of the proteins with an equivocal peripheral localization were refined using the literature and a malaria database to assign them to the PV or the PVM (Table S3). Based on this, the predictions obtained using P, ML, and P+ML were accurate for 53 (69%), 62 (81%), and 66 (86%) orthology groups, respectively. Moreover, incorrectly predicted orthology groups had mostly no or a low confidence level, supporting the confidence rating developed here (Figure 3A). While the reason(s) behind those discrepancies is not clear, it is possible that the expression of proteins fused to a large fluorescent reporter using a strong constitutive promoter, EF1-α, could mistarget PEXEL-positive exported proteins predicted in the VESICLE to the periphery, as shown for *P. yoelii* PY17X_1402300 (Table S3, row 47). Orthology groups such as those including the 60S ribosomal protein L7-3 were predicted with a VESICLE pattern (Table S3, row 73), likely because of their high degree of conservation with host ribosomal proteins, which increases the number of unique peptides available for identification and abundance calculation in the HCC. Hence, the spatial proteome presented in Table S2 was updated to reflect the localization evidence by favoring (1) localization data obtained by individual validation (this study and literature data) and (2) P+ML predictions (Table S2, proteome mapping, column BD). Furthermore, orthology groups with experimental data indicating an ambiguous peripheral pattern were annotated according to the proteomics prediction when available (either the PV, PVM, or PPM for those predicted to be intra-parasitic), while ribosomal proteins were considered intra-parasitic. The proportion of *Plasmodium* proteins found in each subcellular compartment across the five species studied is summarized in Figure 3C.

Validation of the species-specific protein atlas

The P+ML prediction was able to determine the spatial location of orthologous proteins with ~86% accuracy. To evaluate whether this approach was also accurate for exported proteins that were unique to human, simian, or rodent malaria species, 17 proteins predicted to be exported into the spatial proteome were examined using episomal (*P. knowlesi*) and endogenous (*P. falciparum* and *P. yoelii*) HA tagging. This included three *P. falciparum*-specific (Figure 5A), four *P. knowlesi*-specific (Figure 5B), and 10 *P. yoelii*-specific (Figure 5C; Figure S2) proteins.

Of the *P. falciparum* proteins, PF3D7_0301400, containing a PEXEL motif, can be seen to be localized as a punctate pattern in the host cell. Similarly, the PEXEL-negative PF3D7_0731600 is exported into the host cell, while PF3D7_1401500, predicted to be located in the periphery, appears to remain internal (Figure 5A). In line with the P+ML prediction, three of the *P. knowlesi* proteins (PKNH_0933200, PKNH_1472200, and PKNH_1100900) are exported, while PKNH_112470 is associated with its predicted location at the parasite periphery (Figure 5B). Of the 10 rodent-specific proteins, PEXEL-positive PY17X_0701000, PY17X_0701100, and PY17X_1102300 are exported as predicted (Figure 5C). The remaining seven (PY17X_1001800, PY17X_0526100, PY17X_1210300, PY17X_0802800,



(legend on next page)

PY17X_0839600, PY17X_1321800, and PY17X_0701300) are effectively exported out of the parasite (Figure S2).

Taken together, the P+ML approach was able to accurately predict 16 of the 17 proteins analyzed here with an overall accuracy greater than 90%, suggesting that this strategy is highly reliable in predicting the spatial location of the parasite proteome.

Comparative analysis of *Plasmodium* proteomes

The resolved spatial proteome across multiple *Plasmodium* species provides new insights into conserved as well as divergent strategies deployed by these parasites to adapt to diverse hosts. This includes predictive localization of 1,063 orthology groups with no functional annotation. We focused on the distribution, features, essentiality, and conservation of the spatial proteome across the species and incorporated information in relation to evolutionary distance, peak of transcription, and protein function. In line with previous indications, this approach showed an increasing level of novelty and species differences toward the host and parasite periphery.

Intra-parasitic compartments (PARASITE and PPM)

Parasite internal compartments include more than 80% of the spatial proteome (Figure 6A, distribution), displaying the highest level of conservation, with ~95% being shared across the species analyzed (Table S2, proteome mapping, column AS), and contain more than 70% of proteins that are essential or required for parasite development (Figure 6A, essentiality; Table S2, proteome mapping, columns AV–AY). These proteins also display a high phylogenetic percentage with an ancestral origin (Figure 6B), indicating a prevalence of essential functions shared with their metazoan hosts. PARASITE proteins are more than 70% soluble (Figure 6A, features) and are involved in core functions that are transcriptionally co-regulated to exhibit specific timing of induction (Figures 6A, expression peak, and 6C) as reported previously.[46] PPM proteins have a similar proportion of signal peptide (SP) and/or transmembrane (TM) domain(s), with most of the latter containing 4 or more TM domains indicative of multimembrane-spanning proteins (Figure 6A, features). The PPM proteome is preferentially expressed during the ring-to-trophozoite transition (Figure 6A, expression peak) and is mostly involved in transport pathways important for metabolic needs of the subsequent stages (Figure 6C). This includes transporters such as the (V)-H+-ATPases, which form highly conserved, large, membrane-bound, multisubunit complex-containing interacting partner proteins without hydrophobic sequences involved in the regulation of intracellular pH.[47,48] It also includes a relatively small number of conserved proteins with unknown function, of which ~50% are essential.

Despite the high conservation of the PARASITE proteins across species, ~30 of these proteins are *P. falciparum* and *P. knowlesi* specific, indicating a unique function in primate malaria species (Table S2, proteome mapping). In addition, the spatial proteome identified ~640 of the close to 2,900 PARASITE orthology group members as conserved proteins with no functional annotation, of which ~60% are essential.

Peripheral compartments (PV and PVM)

The peripheral compartments contain ~4% of the parasite proteome (Figure 6A, distribution) and display an increased level of diversity, with ~70% of the peripheral proteins being conserved across the species analyzed, most of which are either essential or required for parasite development (Figure 6A, essentiality). Phylostratographic gene enrichment analysis revealed a prevalence of genes specific to *Plasmodium* species, and for species-specific genes, we observed enrichment only in *P. falciparum* (Figure 6B; Table S4, enrichment analysis). This suggests a *Plasmodium*-centered accelerated evolution essential for establishment in the host cell. The PV is enriched in proteins with SP (Figure 6A, features), which are mainly expressed during the transition from schizont to ring (Figure 6A, expression peak). They are mostly involved in proteolytic cascades regulating parasite egress and invasion (Figure 6C, GO:0002377), including the serine-repeat antigen (SERA) gene family, involved in egress following subtilisin-like protease 1 (SUB1) cleavage.[55] Unlike SERA, SUB1 is identifiable in the eukaryotic common ancestor,[56–58] suggesting an evolutionary adaptation of these proteolysis pathways.[59,60] The PVM, which forms the main host-parasite interface, is enriched in proteins with TM domain(s) (Figure 6A, features), that are predominantly expressed during early ring stages (Figure 6A, expression peak). The PVM proteome contains proteins involved in protein translocation (Figure 6C), such as the PTEX complex[5] and its accessory proteins,[32] as well as in nutrient uptake, like EXP2[61] and EXP1.[62] The PV and PVM contain a significant number of known merozoite invasion proteins, including MSP7, RAP1, RAP2, and 6-CYS proteins, indicating that merozoite proteins may play a role in PV/PVM function during the establishment of the parasite immediately after invasion. In addition, about 25% of the proteins have no annotated function, with ~35% of them being essential for parasite survival. The remaining peripheral subset includes mainly non-syntenic proteins unique to a parasite lineage that can be grouped based on literature annotation or sequence similarities that overlap with the exported subset (Figure 6D), suggesting that proteins are transitioning into the HCC, while others may be genuine PVM proteins, such as tryptophan-rich antigens (TRAGs) interacting with early transcribed membrane proteins (ETRAMPs) in the periphery.[63] ETRAMPs form a prominent subtelomeric variant protein family, with 6, 4, 6, 3, and 9 members found in *P. falciparum*, *P. knowlesi*, *P. yoelii*, *P. berghei*, and *P. chabaudi*, respectively, that are shared across the genus, specific to rodent malaria, or a unique species,[64] indicating that the ETRAMP family expanded after speciation to facilitate parasite-specific features. The ETRAMPs identified here constitute a subset of the family repertoire, suggesting their unique role in PVM function.[65]

Exported proteins (exportome)

About 10% of the parasite proteome is exported beyond the PVM and is dominated by lineage- or species-specific protein families with limited functional annotation. The HCC is enriched in proteins with SP, VESICLE/CLEFT are enriched in proteins with TM domain(s), while the GHOST has a similar proportion

Figure 5. Immunofluorescence microscopy of HA-tagged protein

Representative images of (A) *P. falciparum*, (B) *P. knowlesi*, and (C) *P. yoelii* trophozoite/schizont-stage parasites labeled with rat anti-HA (red) and rabbit anti-EXP2 antibodies (green). The parasite nucleus was stained with Hoechst 33342 (blue). Consensus predicted localization is indicated for each protein. Scale bars, 2 μM.

CellPress
OPEN ACCESS

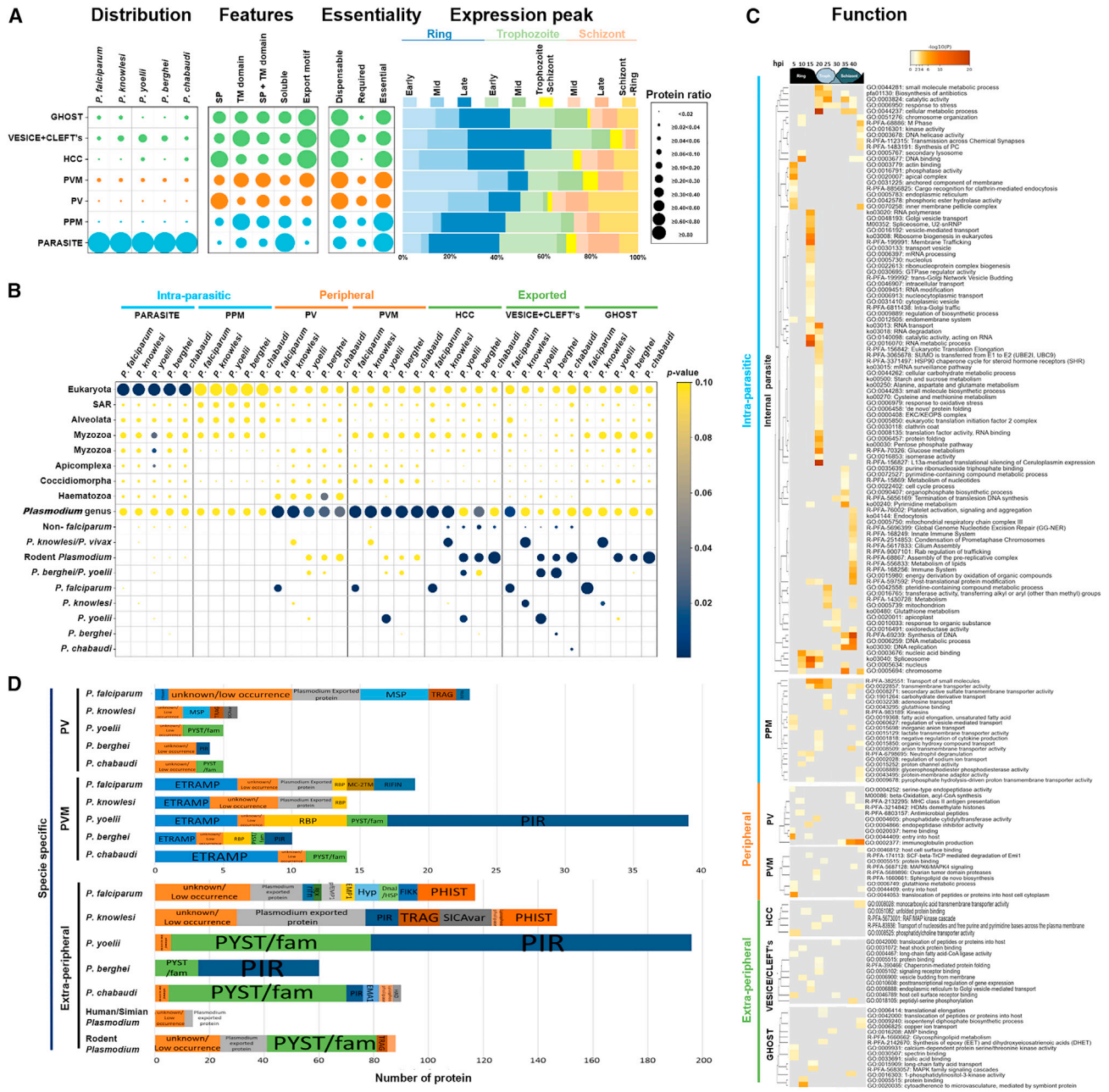


Figure 6. Fact sheet for the malaria spatial proteome

(A) The distribution, features, essentiality, and RNA expression peak of the protein atlas are represented using dot plots. Distribution: the dots are sized to represent the proportion of malaria proteins across the five species studied. Features: the dots reflect the proportion of proteins without (soluble) or with hydrophobic feature(s) (SP and/or TM[s]). Essentiality: the dots indicate the proportion of cognate gene found to be dispensable, required, or essential using genetic screen data obtained for *P. falciparum*[49] and/or rodent *Plasmodium*[50] (RMgmDB). Expression peak: stage of development during which the genes encoding the proteins found in the *P. falciparum* atlas display their highest RNA abundance based on the IDC pattern compiled from 4 RNA sequencing databases[51–54] (Table S2, RNA vs. protein, columns C–AM). The stages of intraerythrocytic parasites correspond to nine time point of the IDC found in these datasets with a 5-h interval between them (5–45 h post invasion).

(B) Phylogenetic analysis. The distribution of the protein atlas over phylogenetic distance (phylostratum) was used to access the broader evolution of each cellular compartment using an evolutionary tree of 19 organisms, including *P. vivax* and the five *Plasmodium* species studied here (Figure S3). In each *Plasmodium* species, the size of the dots indicates the fraction of proteins in each phylostratum relative to the total of proteins found in the compartment. The color of the dots indicates the p value ($p < 0.05$, blue) of the phylostratographic gene enrichment analysis (Table S4, enrichment analysis).

(legend continued on next page)

of proteins with either a SP, a TM domain, or both, indicating that some of these proteins may be membrane associated (Figure 6A, features). The exportome is mainly expressed during the transition from ring to trophozoite (Figure 6A, expression peak) and is largely dispensable for development (Figure 6A, essentiality), in line with previous studies in *P. falciparum*.[66] It also displays the lowest level of conservation, with ~40% of the exported proteins being shared across the species, of which ~50% are essential or required for parasite development. Of the remaining exportome, ~50% is either species specific or rodent parasite specific. Most of these conserved parasite proteins are without functional annotations, in line with unknown conserved remodeling feature(s). The *P. chabaudi* exportome appears to be evenly distributed across the three extracellular compartments, *P. yoelii* and *P. berghei* are characterized by an expanded VESICLE/CLEFT subset, while the *P. falciparum* and *P. knowlesi* exportomes are mainly divided between the GHOST and the VESICLE/CLEFT fractions (Figure 6A, distribution). Phylostratographic gene enrichment analysis highlighted an increasing evolutionary protein divergence toward external compartments that are likely required for host-specific adaptation (Figure 6B; Table S4, enrichment analysis). Primate malaria parasite HCC proteins are enriched in *Plasmodium* proteins, while those found in the VESICLE/CLEFT and GHOST are increasingly specific to *P. falciparum* or *P. knowlesi*/*Plasmodium vivax* lineages (Figure 6B). Rodent parasite exportomes are mainly shared across rodent *Plasmodium.* The increasing presence of proteins common to *P. yoelii*/*P. berghei or* specific to a single species in the VESICLE/CLEFT of rodent parasites supports species-specific radiations of multigene families in non-*P. falciparum* parasites, especially in *P. yoelii* (Figure 6B).

Despite this species-specific divergence, the analysis here has identified a subset of conserved exported proteins, including 20 parasite GHOST proteins of which ~25% are essential, 28 HCC proteins of which ~40% are essential, and 38 VESICLE/CLEFT proteins of which ~20% are essential. The majority of these conserved proteins have no functional annotations, suggesting an unknown conserved *Plasmodium* species role in host cell modifications.

P. falciparum *exportome*

In this parasite, exported proteins are mainly transcribed during the transition from ring to trophozoite stage and involved in multiple functions, such as molecule translocation (GO:0042000) and cytoadherence (GO:0020035), protein families of limited size, and a large subset of proteins encoded by single-copy genes found in the J-dot vesicles, Maurer's clefts, and RBC membrane (Figure 6D). These exported proteins are involved in modulating the physical properties of the RBCs, exporting molecules to the RBC surface and mediating parasite cytoadherence and antigenic variation, with the majority being largely uncharacterized[49] (Table S2, low conservation exportome).

P. knowlesi *exportome*

In *P. knowlesi,* the exportome is enriched in proteins linked to host cell remodeling features found in human parasites (Figure 6D; Table S2, low conservation exportome), as supported by the prevalence of protein families such as PHIST shared with *P. falciparum* and TRAG with *P. vivax* , suggesting broadly conserved functions. TRAG is partially localized in the caveola-vesicle complexes (CVCs) and has been implicated in immune evasion in *P. vivax*,[67] supporting a similar role in *P. knowlesi*. While CVCs have been described on the periphery of *P. vivax*- and *P. cynomolgi*-infected RBCs, evidence of their presence in *P. knowlesi* is limited,[68] and it is interesting to note that *P. knowlesi* knob-associated histidine-rich protein (KAHRP) may be located in the CVC unlike the knob structures in *P. falciparum*.[69] The *P. knowlesi* exportome also includes *P. falciparum* Maurer's clefts orthologs likely localized in the Sinton Mulligan's structures[70,71] proposed to be a translocation machinery functional analogous for immune evasion proteins such as the Schizont Infected Cell Agglutination variant antigens (SICAvar) and PIRs to the RBC surface.[72–74] However, *P. knowlesi*-exported protein shared with other human parasites may also have species-specific functions, as for the PHIST protein PKNH_0117900, whose syntenic orthologs were identified in *P. falciparum* mobile J-dots[38] and *P. vivax* CVCs.[75] Finally, the existence of species-specific remodeling features is further supported by the existence of a large subset of non-syntenic proteins of unknown function unique to *P. knowlesi.*

Rodent parasite exportomes

The rodent parasite exportomes are characterized by the broad expansion of distinct variant protein families. The spatial proteome provides a more detailed characterization of these protein families. The *P. yoelii* and *P. berghei* exportomes are mainly associated with an expanded repertoire of PIRs and confirms previous studies indicating that they are located in the vesicles and, to some extent, the HCC but not the RBC membrane,[13,14,16] unlike in *P. vivax*, where some PIRs are expressed on the RBC surface.[76,77] The *P. chabaudi* exportome is characterized by a prevalence of PYST/fam proteins, and our data show them to be located in multiple compartments, including the RBC membrane.[13–15] Some PYST/fam proteins might be involved in lipid metabolism, similarly to the *P. falciparum* START-domain-containing protein,[78] along with exported lysophospholipases specific to non-*P. falciparum*[13] that are expanded in *P. chabaudi*. The spatial proteome provides information about the location of a significant number of distinct members of each gene family that will aid in fully understanding their function. Finally, the exportomes include proteins of unknown function conserved in all or a subset of rodent parasites that, like SMAC and MAHRP1b, could contribute to functions like the CD36-mediated sequestration of schizont-infected RBCs described in *P. berghei*.[29,79]

(C) Function of 2,707 proteins included in the *P. falciparum* protein atlas, predicted using the Metascape analysis resource. The clustered heatmaps generated using Metascape depict the top enriched functions/pathways shared between or selectively ascribed to the proteins induced at a development time of the IDC. The color scale denotes the statistical significance.

(D) Composition of the non-syntenic proteins included in the peripheral and exported proteome following their annotations in PlasmoDB. Only protein groups with more than three occurrences in the exportome are depicted. The unknown/low occurrence group includes proteins encoded by single-copy genes, with most of them having an unknown function.

CellPress
OPEN ACCESS

Figure 7. Protein and RNA correlation across subcellular localization

(A) Distribution of RNA expression in abundance percentiles for atlas and undetected proteins.

(B) (i) Pathway enrichment analysis of atlas and undetected protein subsets. (ii) RNA and protein abundance peak patterns of proteins undetected in the atlas. The RNA abundance patterns of 269 undetected proteins highlighted by pathway enrichment analysis were compiled from RNA sequencing datasets available in

(legend continued on next page)

Differences in transcriptional regulation of the spatial proteome

Plasmodium species utilize a range of different mechanisms to regulate overall gene expression, including epigenetic regulation of multigene family member expression,[80] tRNA modifications driving codon-biased gene expression,[81] and N6-methyladenosine (m^6A) modifications impacting transcript stability and translation rates.[82] To evaluate whether there are differences in the way the expression of the spatial proteome is regulated, we evaluated the correlation between mRNA levels and peak transcription with overall protein abundance and localization.

A clear relationship between the overall RNA abundance and protein detection can be observed, with undetected proteins mainly associated with transcripts with an expression value below the 20th percentile of the overall transcriptomics data (Figure 7A). However, a significant number of proteins (>20th percentile) were still not detected, even at higher RNA abundance (Figure 7A), because of degradation, transcriptional or post-transcriptional regulation, and/or limitations of MS. Some of those proteins (n = 95) were found in the iTRAQ intraerythrocytic developmental cycle (IDC) proteome[81] and have their highest abundance during the early stage (Figure 7Bii, protein), indicating that proteins undetected at the trophozoite-schizont stage may have an earlier role. The undetected subset is mainly encoded by transcripts with a peak abundance at the schizont stage (Figure 7Bii, mRNA) and enriched in apicoplast (GO:0009536) and mitochondrion (GO:0005739) proteins (Figure 7Bi) involved in metabolic pathways not essential for blood stage development, such as fatty acid (GO:0006633), heme biosynthesis (ko00860), or lipoic acid metabolism (ko00785) (Figure 7Biv).[83–85] The spatial proteome includes only 111 of 315 apicoplast[86] and 105 of 280 mitochondrial[87] proteins previously predicted in other studies compared with 718 of ~760 for core nuclear proteins[88] (Figure 7Biii). Collectively, it indicates that the delay between RNA and protein peaks because of regulatory effects on transcription, mRNA stability, and/or translation as well as the protein function and essentiality contributed to the missing spatial proteome.

Most proteins detected in the spatial proteome are encoded by transcripts above the 20th percentile of the overall transcriptomics data (n = 2,571), with RNA peaks during ring-trophozoite transition (Figure 7CI, mRNA). These proteins are increasingly abundant during the schizont stage, according to the cyclic abundance profiles of 1,164 internal proteins obtained from the iTRAQ IDC data[51–54] (Figure 7CI, protein). To investigate whether the dynamic between RNA and protein accumulation differs depending on final protein location, we calculated the Spearman rank correlation between the IDC transcriptomes and the proteomes found in each subcellular localization following their abundance percentiles (Figure 7Cii). Internal proteins showed the best correlation with their corresponding transcript profile; however, the moderately high Spearman correlation suggests a role of other regulatory mechanisms, like rates of translation and protein degradation.[89] In contrast, peripheral proteins displayed the highest correlation with the late schizont transcriptome, while exported proteins correlated best with early trophozoite transcriptomes. The observation that peripheral proteins are enriched in merozoite proteins suggests that, during merozoite invasion, many of the peripheral proteins are directly delivered to their eventual location, with ring-stage transcription and translation playing a minor role. Conversely, exported proteins are only transcribed and translated after invasion and establishment of the peripheral export machinery.

To explore the role of post-transcriptional regulation on the spatial proteome, we compared the abundance percentile rankings of each protein with its corresponding RNA to identify incongruence by at least the 20th percentile and analyzed the variation of RNA levels across IDC of those discrepancies (Figures 7Ciii and 7Civ). Intra-parasitic compartments displayed the highest proportion, with an increased RNA-to-protein abundance level, characterized by a high level of transcript and a reduced level of protein ($p < 0.05$) (Figures 7Ciii and 7Civ). This incongruence displayed a reduced transcript level variability, which contributed to the decreased mean coefficient of variation computed for RNA encoding intra-parasitic proteins (Figure 7Cv), supporting the existence of translational repression and/or rapid protein degradation. Conversely, peripheral and exported subsets were enriched in proteins with a decreased RNA-to-protein abundance level that are associated with a low RNA level, an increased protein level ($p < 0.05$) (Figures 7Ciii and 7Civ), and a broader transcript level variability (Figure 7Cv), suggesting the presence of increased translational activity and/or rapid RNA degradation. Pathway enrichment analysis showed that proteins with decreased RNA-to-protein abundance levels are involved in nucleic acid metabolism, such as spliceosome (ko03040) and helicase (GO:0004386) activities, while those with an increased RNA-to-protein abundance level broadly fall into three groups related to virulence and survival in the host: parasite egress mediated by SERA and SUB3 proteases (GO:0048583), host cell invasion and nutrient acquisition (GO:0044409) involving Rhoph/CLAG proteins, and fatty acid scavenging (M0086) mediated by acyl-coenzyme A (CoA) synthetases targeted to the RBC membrane, such as the ACS8 characterized in the study (Figure 7vii). Finally, post-transcriptional m^6A modification of

PlasmodB. Of these 269 proteins, 95 were identified by IDC iTRAQ-based quantitative proteomics, including six time points across the IDC.[81] (iii) distribution of atlas and undetected proteins in published apicoplast,[86] mitochondrial,[87] and nuclear[88] proteomes. (iv) pathway enrichment analysis of the apicoplast and mitochondrial proteins included in the atlas and undetected subsets.

(C) (i) RNA and protein abundance peak patterns of atlas proteins compiled from published RNA sequencing or IDC iTRAQ-based quantitative proteomics.[81,51–54] (ii) Spearman rank correlation between mRNA level and protein localization at various parasite developmental stages. (iii) proportion of proteins with concordant, increased, or decreased RNA levels across subcellular fractions. (iv) RNA (top) and protein (bottom) abundance percentiles of the proteins with concordant, increased, or decreased RNA levels across subcellular fractions. The horizontal bar indicates the median percentile, and an asterisk indicates a significant difference (*$p < 0.05$). (v) RNA-level dynamic of the proteins with concordant, increased, or decreased RNA levels across subcellular fractions. (vi) proportion of proteins with concordant, increased, or decreased RNA levels with m^6A-methylated RNA. (vii) pathway enrichment analysis of proteins with concordant, increased, or decreased RNA levels.

mRNA plays a role in decreased mRNA stability and translational efficiency.[82] Here, the investigation of m^6A levels in the genes encoding for different components of the spatial proteome revealed an increased level of m^6A in transcripts with an increased RNA-to-protein abundance ratio (Figure 7vi). Strikingly, the increase in m^6A in transcripts encoding for peripheral proteins suggests that these genes are particularly subjected to epi-transcriptomic regulation.

Altogether, the discordance in RNA and protein enrichment supports the existence of various post-transcriptional regulation patterns controlling the level of proteins involved in functions essential for parasite virulence and survival. This could provide a rapid translational response to environmental changes despite a higher energy cost.[90]

DISCUSSION

Many of the proteins identified in the genomes of the *Plasmodium* species have no or limited functional annotations. To address this significant gap in our understanding, we determined the spatial cellular location of the asexual blood-stage proteins across multiple species in an unbiased manner. This study reports a comparative spatial atlas of five *Plasmodium* proteomes, identifying and quantifying 12,392 malaria proteins representing 4,350 orthology groups. The combination of subcellular fractionation together with quantitative multiplex LC-MS/MS provided an extensive dataset that served as a starting point for two unbiased computational approaches. The *in silico* approaches were trained with a pan-species set of 3,709 molecules (~30% of the experimental proteome) with localization data inferred from the literature and protein features to predict the localization of the remaining 8,683 proteins (~70% of the experimental proteome). The experimental validation, consisting of 93 orthology groups across three parasite species, supported the overall consensus localization prediction strategy deployed here. Based on this, the predictive subcellular localization of 1,124 single-copy orthology groups with no functional annotation is provided with an overall accuracy of 86%. Additionally, 80 previously unassigned single-copy gene orthology groups with limited functional annotation were identified in this study to be potentially exported beyond the parasite PPM, using consensus localization prediction. This new information provides a powerful discovery tool for future studies of their possible roles in parasite biology.

The spatial data also provide new information about *Plasmodium* species-specific differences that indicate host-specific adaptation, particularly of exported proteins. This confirms previous studies indicating a relatively high conservation of the intra-parasitic proteomes, which make up more than 80% of the proteome. These fractions also contain the highest proportion of essential or required proteins for parasite development and have the highest level of conservation with eukaryotic ancestor proteins, in line with the many core molecular activities they represent. The abundance of multi-TM-containing proteins found in the PPM highlights its role in cellular homeostasis. However, the presence of numerous intra-parasitic proteins unique to malaria parasites (~1,300 of 2,800 in *P. falciparum* according to orthoMCL DB) highlights the importance of parasite-specific core activities, like pyrimidine metabolism. The dataset generated here provides a starting point for characterizing these proteins for further functional studies.

Proteins located in the peripheral compartment represent about 4% of the total proteome and show a species-specific divergence, with ~70% of the proteins identified being conserved across all five species. This suggests host-specific selection pressure encountered by the parasite. Moreover, many of the proteins in this location are unique to *Plasmodium*, in line with the unique role this compartment plays in the biology of the parasite. A surprising finding is the relative abundance of merozoite proteins previously linked to invasion in this compartment, suggesting a role in establishing a viable PV/PVM. It also indicates that many of the early post-invasion functions required by the parasite are not provided by the ring-stage parasite but by the invading merozoite and are supplemented subsequently by additional expression in the developing parasite. Moreover, it is possible that some of these proteins have a dual role in terms of merozoite invasion as well as establishment of the parasite in the host cell, as seen in the case of RhopH3.[91]

The parasite proteins exported beyond the PVM represent the evolutionarily most divergent part of the parasite genome, including a wider variety of species-specific genes. Many of the exported proteins are members of multigene families not essential for blood-stage development. Comparison of the exported proteome across species shows a striking divergence with an increased repertoire of proteins associated with lipid metabolism in rodent parasite that is further expanded in *P. chabaudi*, supporting the theory that each parasite species encounters unique host-dependent selection pressure that needs to be overcome. This also suggest that the functions mediated by these exported proteins are of particular relevance for zoonotic jumps, as seen in the case of *P. knowlesi*.

The data here also identified many essential parasite ortholog groups with no functional annotation that are predicted to be localized in the internal parasite and that are likely involved in critical parasite-specific metabolic or developmental activities. These proteins not only represent attractive parasite-specific drug targets but also open a doorway to gaining a better understanding of the complex parasite biology. While the total number of parasite orthologs that are exported is significantly lower, the fact that many of them are essential for parasite biology identifies these as potential critical regulators of host cell remodeling that is conserved across *Plasmodium* species. Their relatively small number now unlocks an opportunity to identify a pan-parasite Achilles' heel that can be exploited for new therapeutic interventions.

Post-transcriptional regulation of expression is known to play an important role in *Plasmodium* biology, and our data indicate that some compartments of the spatial proteome are under different levels of post-transcriptional regulation. Specifically, the peripheral proteome appears to be regulated by an increase in m^6A modification of its mRNA, suggesting an important role of this epi-transcriptomic regulator in parasite biology.

Taken together, the comprehensive and validated spatial proteome provides a powerful resource for improving our understanding of the unique parasite-specific biology. Specifically, it

helps facilitate the establishment of conserved as well as divergent features across the parasite species, which will be an important starting point for informing future studies. At the same time, the data also provide insights into unique parasite targets that could be explored for future therapy development.

Limitations of the study

Although significant efforts have been made to ensure that the subcellular fractionation protocol avoided spillover effects from other fractions, it cannot be totally excluded, which could impact the final localization assignment for some proteins. In addition, the training set containing known locations of parasite proteins is potentially biased toward proteins with eukaryotic orthologs of known function, which could lead to a bias in the overall analysis. It is also important to note that, for some proteins, the final localization was not possible to assign or is different across parasite species. To resolve these discrepancies, it will be necessary to experimentally validate the location of some of these in the future.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Human samples
  - Cultivation of *P. falciparum* 3D7 and *P. knowlesi* A1-H.1
  - Animal samples
  - Cultivation of P. yoelii yoelii 17X 1.1, P. berghei ANKA and P. Chabaudi chabaudi AS
- METHOD DETAILS
  - Fractionation of the parasite samples
  - Western blot analysis of parasite preparations after subcellular fractionation
  - Proteome extraction, purification, and preparation
  - High-pressure liquid chromatography fractionation
  - LC-MS/MS analysis of TMT labeled proteins
  - Mass Spectrometric data analysis
  - Generation of training dataset
  - Prediction of the malaria protein localization
  - Derivation of a consensus localization
  - *P. falciparum* plasmid construction and transfection
  - *P. knowlesi* plasmid construction and transfection
  - *P. yoelii* plasmid construction and transfection
  - Determination of the protein localization
  - Transcriptional profile analysis of spatial proteome
  - Post-transcriptional regulation on the spatial proteome
  - Pathway analysis
  - Phylogenomic and phylostratigraphic enrichment analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.113419.

ACKNOWLEDGMENTS

This work is in memory of Dr. Anthony Siau, who died following a tragic accident in August, 2023. We thank Dr. Xiaohong Gao for critical reading and Ms. Shiqin Howe for administrative support. This research is supported by the Singapore Ministry of Health's National Medical Research Council under an Open Fund individual research grant (NMRC/OFIRG/0058/2017), NMRC Cooperative Basic Research grant (NMRC/CBRG/0040/2013), Singapore Immunology Network grant 07-009, and Ministry of Education grant MOE2019-T3-1-007. J.W.A. and H.P.L. were funded by the A*STAR Graduate Academy (AGA). This paper is dedicated to Ms. Nathalie Heng, who died following a long illness after contributing to initiate the iterative permutation approach.

AUTHOR CONTRIBUTIONS

Conceptualization, A.S. and P.R.P.; methodology, A.S., S.S.K., and W.M.; investigation, A.S., J.W.A., O.S., R.H., H.P.L., D.T., X.H., X.Y.Y., S.K.L., W.M., and I.J.; writing – original draft, A.S. and P.R.P.; writing – review & editing, A.S., J.W.A., M.M., and P.R.P.; funding acquisition, P.R.P.; resources, J.W.A.; supervision, P.R.P.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: February 15, 2022
Revised: July 14, 2023
Accepted: October 25, 2023
Published: November 11, 2023

REFERENCES

1. de Koning-Ward, T.F., Dixon, M.W.A., Tilley, L., and Gilson, P.R. (2016). Plasmodium species: master renovators of their host cells. Nat. Rev. Microbiol. *14*, 494–507. https://doi.org/10.1038/nrmicro.2016.79.
2. Hillier, C., Pardo, M., Yu, L., Bushell, E., Sanderson, T., Metcalf, T., Herd, C., Anar, B., Rayner, J.C., Billker, O., and Choudhary, J.S. (2019). Landscape of the Plasmodium Interactome Reveals Both Conserved and Species-Specific Functionality. Cell Rep. *28*, 1635–1647.e5. https://doi.org/10.1016/j.celrep.2019.07.019.
3. 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. https://doi.org/10.1038/nature04104.
4. Tan, Q.W., and Mutwil, M. (2020). Malaria.tools-comparative genomic and transcriptomic database for Plasmodium species. Nucleic Acids Res. *48*, D768–D775. https://doi.org/10.1093/nar/gkz662.
5. 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., and Crabb, B.S. (2009). A newly discovered protein export machine in malaria parasites. Nature *459*, 945–949. https://doi.org/10.1038/nature08104.
6. Marti, M., Good, R.T., Rug, M., Knuepfer, E., and Cowman, A.F. (2004). Targeting malaria virulence and remodeling proteins to the host erythrocyte. Science *306*, 1930–1933. https://doi.org/10.1126/science.1102452.
7. Sargeant, T.J., Marti, M., Caler, E., Carlton, J.M., Simpson, K., Speed, T.P., and Cowman, A.F. (2006). Lineage-specific expansion of proteins

exported to erythrocytes in malaria parasites. Genome Biol. *7*, R12. https://doi.org/10.1186/gb-2006-7-2-r12.

8. van Ooij, C., Tamez, P., Bhattacharjee, S., Hiller, N.L., Harrison, T., Liolios, K., Kooij, T., Ramesar, J., Balu, B., Adams, J., et al. (2008). The malaria secretome: from algorithms to essential function in blood stage infection. PLoS Pathog. *4*, e1000084. https://doi.org/10.1371/journal.ppat.1000084.
9. Hiller, N.L., Bhattacharjee, S., van Ooij, C., Liolios, K., Harrison, T., Lopez-Estraño, C., and Haldar, K. (2004). A host-targeting signal in virulence proteins reveals a secretome in malarial infection. Science *306*, 1934–1937. https://doi.org/10.1126/science.1102737.
10. Matz, J.M., and Matuschewski, K. (2018). An in silico down-scaling approach uncovers novel constituents of the Plasmodium-containing vacuole. Sci. Rep. *8*, 14055. https://doi.org/10.1038/s41598-018-32471-6.
11. Heiber, A., Kruse, F., Pick, C., Grüring, C., Flemming, S., Oberli, A., Schoeler, H., Retzlaff, S., Mesén-Ramírez, P., Hiss, J.A., et al. (2013). Identification of new PNEPs indicates a substantial non-PEXEL exportome and underpins common features in Plasmodium falciparum protein export. PLoS Pathog. *9*, e1003546. https://doi.org/10.1371/journal.ppat.1003546.
12. Spielmann, T., and Gilberger, T.W. (2010). Protein export in malaria parasites: do multiple export motifs add up to multiple export pathways? Trends Parasitol. *26*, 6–10. https://doi.org/10.1016/j.pt.2009.10.001.
13. Pasini, E.M., Braks, J.A., Fonager, J., Klop, O., Aime, E., Spaccapelo, R., Otto, T.D., Berriman, M., Hiss, J.A., Thomas, A.W., et al. (2013). Proteomic and genetic analyses demonstrate that Plasmodium berghei blood stages export a large and diverse repertoire of proteins. Mol. Cell. Proteomics *12*, 426–448. https://doi.org/10.1074/mcp.M112.021238.
14. Siau, A., Huang, X., Yam, X.Y., Bob, N.S., Sun, H., Rajapakse, J.C., Renia, L., and Preiser, P.R. (2014). Identification of a new export signal in Plasmodium yoelii: identification of a new exportome. Cell Microbiol. *16*, 673–686. https://doi.org/10.1111/cmi.12293.
15. Siau, A., Huang, X., Weng, M., Sze, S.K., and Preiser, P.R. (2016). Proteome mapping of Plasmodium: identification of the P. yoelii remodellome. Sci. Rep. *6*, 31055. https://doi.org/10.1038/srep31055.
16. Yam, X.Y., Brugat, T., Siau, A., Lawton, J., Wong, D.S., Farah, A., Twang, J.S., Gao, X., Langhorne, J., and Preiser, P.R. (2016). Characterization of the Plasmodium Interspersed Repeats (PIR) proteins of Plasmodium chabaudi indicates functional diversity. Sci. Rep. *6*, 23449. https://doi.org/10.1038/srep23449.
17. Chen, Q., Fernandez, V., Sundström, A., Schlichtherle, M., Datta, S., Hagblom, P., and Wahlgren, M. (1998). Developmental selection of var gene expression in Plasmodium falciparum. Nature *394*, 392–395. https://doi.org/10.1038/28660.
18. Scherf, A., Rivière, L., and Lopez-Rubio, J.J. (2008). SnapShot: var gene expression in the malaria parasite. Cell *134*, 190. https://doi.org/10.1016/j.cell.2008.06.042.
19. Voss, T.S., Healer, J., Marty, A.J., Duffy, M.F., Thompson, J.K., Beeson, J.G., Reeder, J.C., Crabb, B.S., and Cowman, A.F. (2006). A var gene promoter controls allelic exclusion of virulence genes in Plasmodium falciparum malaria. Nature *439*, 1004–1008. https://doi.org/10.1038/nature04407.
20. Horrocks, P., Pinches, R., Christodoulou, Z., Kyes, S.A., and Newbold, C.I. (2004). Variable var transition rates underlie antigenic variation in malaria. Proc. Natl. Acad. Sci. USA *101*, 11129–11134. https://doi.org/10.1073/pnas.0402347101.
21. Dzikowski, R., Frank, M., and Deitsch, K. (2006). Mutually exclusive expression of virulence genes by malaria parasites is regulated independently of antigen production. PLoS Pathog. *2*, e22. https://doi.org/10.1371/journal.ppat.0020022.
22. Frank, M., and Deitsch, K. (2006). Activation, silencing and mutually exclusive expression within the var gene family of Plasmodium falciparum. Int. J. Parasitol. *36*, 975–985. https://doi.org/10.1016/j.ijpara.2006.05.007.
23. Lawton, J., Brugat, T., Yan, Y.X., Reid, A.J., Böhme, U., Otto, T.D., Pain, A., Jackson, A., Berriman, M., Cunningham, D., et al. (2012). Characterization and gene expression analysis of the cir multi-gene family of Plasmodium chabaudi chabaudi (AS). BMC Genom. *13*, 125. https://doi.org/10.1186/1471-2164-13-125.
24. Brugat, T., Reid, A.J., Lin, J., Cunningham, D., Tumwine, I., Kushinga, G., McLaughlin, S., Spence, P., Böhme, U., Sanders, M., et al. (2017). Antibody-independent mechanisms regulate the establishment of chronic Plasmodium infection. Nat. Microbiol. *2*, 16276. https://doi.org/10.1038/nmicrobiol.2016.276.
25. Cunningham, D.A., Jarra, W., Koernig, S., Fonager, J., Fernandez-Reyes, D., Blythe, J.E., Waller, C., Preiser, P.R., and Langhorne, J. (2005). Host immunity modulates transcriptional changes in a multigene family (yir) of rodent malaria. Mol. Microbiol. *58*, 636–647. https://doi.org/10.1111/j.1365-2958.2005.04840.x.
26. Fonager, J., Cunningham, D., Jarra, W., Koernig, S., Henneman, A.A., Langhorne, J., and Preiser, P. (2007). Transcription and alternative splicing in the yir multigene family of the malaria parasite Plasmodium y. yoelii: identification of motifs suggesting epigenetic and post-transcriptional control of RNA expression. Mol. Biochem. Parasitol. *156*, 1–11. https://doi.org/10.1016/j.molbiopara.2007.06.006.
27. Nyalwidhe, J., and Lingelbach, K. (2006). Proteases and chaperones are the most abundant proteins in the parasitophorous vacuole of Plasmodium falciparum-infected erythrocytes. Proteomics *6*, 1563–1573. https://doi.org/10.1002/pmic.200500379.
28. Florens, L., Liu, X., Wang, Y., Yang, S., Schwartz, O., Peglar, M., Carucci, D.J., Yates, J.R., 3rd, and Wu, Y. (2004). Proteomics approach reveals novel proteins on the surface of malaria-infected erythrocytes. Mol. Biochem. Parasitol. *135*, 1–11. https://doi.org/10.1016/j.molbiopara.2003.12.007.
29. Fonager, J., Pasini, E.M., Braks, J.A.M., Klop, O., Ramesar, J., Remarque, E.J., Vroegrijk, I.O.C.M., van Duinen, S.G., Thomas, A.W., Khan, S.M., et al. (2012). Reduced CD36-dependent tissue sequestration of Plasmodium-infected erythrocytes is detrimental to malaria parasite growth in vivo. J. Exp. Med. *209*, 93–107. https://doi.org/10.1084/jem.20110762.
30. Vincensini, L., Richert, S., Blisnick, T., Van Dorsselaer, A., Leize-Wagner, E., Rabilloud, T., and Braun Breton, C. (2005). Proteomic analysis identifies novel proteins of the Maurer's clefts, a secretory compartment delivering Plasmodium falciparum proteins to the surface of its host cell. Mol. Cell. Proteomics *4*, 582–593. https://doi.org/10.1074/mcp.M400176-MCP200.
31. Di Girolamo, F., Raggi, C., Birago, C., Pizzi, E., Lalle, M., Picci, L., Pace, T., Bachi, A., de Jong, J., Janse, C.J., et al. (2008). Plasmodium lipid rafts contain proteins implicated in vesicular trafficking and signalling as well as members of the PIR superfamily, potentially implicated in host immune system interactions. Proteomics *8*, 2500–2513. https://doi.org/10.1002/pmic.200700763.
32. Elsworth, B., Sanders, P.R., Nebl, T., Batinovic, S., Kalanon, M., Nie, C.Q., Charnaud, S.C., Bullen, H.E., de Koning Ward, T.F., Tilley, L., et al. (2016). Proteomic analysis reveals novel proteins associated with the Plasmodium protein exporter PTEX and a loss of complex stability upon truncation of the core PTEX component. Cell Microbiol. *18*, 1551–1569. https://doi.org/10.1111/cmi.12596.
33. Chisholm, S.A., Kalanon, M., Nebl, T., Sanders, P.R., Matthews, K.M., Dickerman, B.K., Gilson, P.R., and de Koning-Ward, T.F. (2018). The malaria PTEX component PTEX88 interacts most closely with HSP101 at the host-parasite interface. FEBS J. *285*, 2037–2055. https://doi.org/10.1111/febs.14463.
34. Schnider, C.B., Bausch-Fluck, D., Brühlmann, F., Heussler, V.T., and Burda, P.C. (2018). BioID Reveals Novel Proteins of the Plasmodium Parasitophorous Vacuole Membrane. mSphere *3*, e00522-17. https://doi.org/10.1128/mSphere.00522-17.

35. Khosh-Naucke, M., Becker, J., Mesén-Ramírez, P., Kiani, P., Birnbaum, J., Fröhlke, U., Jonscher, E., Schlüter, H., and Spielmann, T. (2018). Identification of novel parasitophorous vacuole proteins in P. falciparum parasites using BioID. Int. J. Med. Microbiol. *308*, 13–24. https://doi.org/10.1016/j.ijmm.2017.07.007.

36. Mesén-Ramírez, P., Reinsch, F., Blancke Soares, A., Bergmann, B., Ullrich, A.K., Tenzer, S., and Spielmann, T. (2016). Stable Translocation Intermediates Jam Global Protein Export in Plasmodium falciparum Parasites and Link the PTEX Component EXP2 with Translocation Activity. PLoS Pathog. *12*, e1005618. https://doi.org/10.1371/journal.ppat.1005618.

37. Batinovic, S., McHugh, E., Chisholm, S.A., Matthews, K., Liu, B., Dumont, L., Charnaud, S.C., Schneider, M.P., Gilson, P.R., de Koning-Ward, T.F., et al. (2017). An exported protein-interacting complex involved in the trafficking of virulence determinants in Plasmodium-infected erythrocytes. Nat. Commun. *8*, 16044. https://doi.org/10.1038/ncomms16044.

38. Zhang, Q., Ma, C., Oberli, A., Zinz, A., Engels, S., and Przyborski, J.M. (2017). Proteomic analysis of exported chaperone/co-chaperone complexes of P. falciparum reveals an array of complex protein-protein interactions. Sci. Rep. *7*, 42188. https://doi.org/10.1038/srep42188.

39. Nilsson Bark, S.K., Ahmad, R., Dantzler, K., Lukens, A.K., De Niz, M., Szucs, M.J., Jin, X., Cotton, J., Hoffmann, D., Bric-Furlong, E., et al. (2018). Quantitative Proteomic Profiling Reveals Novel Plasmodium falciparum Surface Antigens and Possible Vaccine Candidates. Mol. Cell. Proteomics *17*, 43–60. https://doi.org/10.1074/mcp.RA117.000076.

40. Alampalli, S.V., Grover, M., Chandran, S., Tatu, U., and Acharya, P. (2018). Proteome and Structural Organization of the Knob Complex on the Surface of the Plasmodium Infected Red Blood Cell. Proteomics. Clin. Appl. *12*, e1600177. https://doi.org/10.1002/prca.201600177.

41. Shastri, S., Zeeman, A.M., Berry, L., Verburgh, R.J., Braun-Breton, C., Thomas, A.W., Gannoun-Zaki, L., Kocken, C.H.M., and Vial, H.J. (2010). Plasmodium CDP-DAG synthase: an atypical gene with an essential N-terminal extension. Int. J. Parasitol. *40*, 1257–1268. https://doi.org/10.1016/j.ijpara.2010.03.006.

42. Millar, S.B., and Cox-Singh, J. (2015). Human infections with Plasmodium knowlesi–zoonotic malaria. Clin. Microbiol. Infect. *21*, 640–648. https://doi.org/10.1016/j.cmi.2015.03.017.

43. Janse, C.J., Franke-Fayard, B., Mair, G.R., Ramesar, J., Thiel, C., Engelmann, S., Matuschewski, K., van Gemert, G.J., Sauerwein, R.W., and Waters, A.P. (2006). High efficiency transfection of Plasmodium berghei facilitates novel selection procedures. Mol. Biochem. Parasitol. *145*, 60–70. https://doi.org/10.1016/j.molbiopara.2005.09.007.

44. Rauniyar, N., and Yates, J.R., 3rd. (2014). Isobaric labeling-based relative quantification in shotgun proteomics. J. Proteome Res. *13*, 5293–5309. https://doi.org/10.1021/pr500880b.

45. Bachmann, A., Petter, M., Tilly, A.K., Biller, L., Uliczka, K.A., Duffy, M.F., Tannich, E., and Bruchhaus, I. (2012). Temporal expression and localization patterns of variant surface antigens in clinical Plasmodium falciparum isolates during erythrocyte schizogony. PLoS One *7*, e49540. https://doi.org/10.1371/journal.pone.0049540.

46. Bozdech, Z., Llinás, 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. https://doi.org/10.1371/journal.pbio.0000005.

47. Nelson, N., and Harvey, W.R. (1999). Vacuolar and plasma membrane proton-adenosinetriphosphatases. Physiol. Rev. *79*, 361–385. https://doi.org/10.1152/physrev.1999.79.2.361.

48. Stevens, T.H., and Forgac, M. (1997). Structure, function and regulation of the vacuolar (H+)-ATPase. Annu. Rev. Cell Dev. Biol. *13*, 779–808. https://doi.org/10.1146/annurev.cellbio.13.1.779.

49. Zhang, M., Wang, C., Otto, T.D., Oberstaller, J., Liao, X., Adapa, S.R., Udenze, K., Bronner, I.F., Casandra, D., Mayho, M., et al. (2018). Uncovering the essential genes of the human malaria parasite Plasmodium falciparum by saturation mutagenesis. Science *360*, eaap7847. https://doi.org/10.1126/science.aap7847.

50. Bushell, E., Gomes, A.R., Sanderson, T., Anar, B., Girling, G., Herd, C., Metcalf, T., Modrzynska, 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.e8. https://doi.org/10.1016/j.cell.2017.06.030.

51. Bártfai, R., Hoeijmakers, W.A.M., Salcedo-Amaya, A.M., Smits, A.H., Janssen-Megens, E., Kaan, A., Treeck, M., Gilberger, T.W., Françoijs, K.J., and Stunnenberg, H.G. (2010). H2A.Z demarcates intergenic regions of the plasmodium falciparum epigenome that are dynamically marked by H3K9ac and H3K4me3. PLoS Pathog. *6*, e1001223. https://doi.org/10.1371/journal.ppat.1001223.

52. Hoeijmakers, W.A., Bártfai, R., and Kensche, P.R. (2015). Intraerythrocytic cycle transcriptome (3D7). PlasmoDB. *13* (*0*). https://plasmodb.org/plasmo/app/record/dataset/DS_715bf2deda#description.

53. Siegel, T.N., Hon, C.C., Zhang, Q., Lopez-Rubio, J.J., Scheidig-Benatar, C., Martins, R.M., Sismeiro, O., Coppée, J.Y., and Scherf, A. (2014). Strand-specific RNA-Seq reveals widespread and developmentally regulated transcription of natural antisense transcripts in Plasmodium falciparum. BMC Genom. *15*, 150. https://doi.org/10.1186/1471-2164-15-150.

54. Toenhake, C.G., Fraschka, S.A.K., Vijayabaskar, M.S., Westhead, D.R., van Heeringen, S.J., and Bártfai, R. (2018). Chromatin Accessibility-Based Characterization of the Gene Regulatory Network Underlying Plasmodium falciparum Blood-Stage Development. Cell Host Microbe *23*, 557–569.e9. https://doi.org/10.1016/j.chom.2018.03.007.

55. Yeoh, S., O'Donnell, R.A., Koussis, K., Dluzewski, A.R., Ansell, K.H., Osborne, S.A., Hackett, F., Withers-Martinez, C., Mitchell, G.H., Bannister, L.H., et al. (2007). Subcellular discharge of a serine protease mediates release of invasive malaria parasites from host erythrocytes. Cell *131*, 1072–1083. https://doi.org/10.1016/j.cell.2007.10.049.

56. Collins, C.R., Hackett, F., Atid, J., Tan, M.S.Y., and Blackman, M.J. (2017). The Plasmodium falciparum pseudoprotease SERA5 regulates the kinetics and efficiency of malaria parasite egress from host erythrocytes. PLoS Pathog. *13*, e1006453. https://doi.org/10.1371/journal.ppat.1006453.

57. Ruecker, A., Shea, M., Hackett, F., Suarez, C., Hirst, E.M.A., Milutinovic, K., Withers-Martinez, C., and Blackman, M.J. (2012). Proteolytic activation of the essential parasitophorous vacuole cysteine protease SERA6 accompanies malaria parasite egress from its host erythrocyte. J. Biol. Chem. *287*, 37949–37963. https://doi.org/10.1074/jbc.M112.400820.

58. Thomas, J.A., Tan, M.S.Y., Bisson, C., Borg, A., Umrekar, T.R., Hackett, F., Hale, V.L., Vizcay-Barrena, G., Fleck, R.A., Snijders, A.P., et al. (2018). A protease cascade regulates release of the human malaria parasite Plasmodium falciparum from host red blood cells. Nat. Microbiol. *3*, 447–455. https://doi.org/10.1038/s41564-018-0111-0.

59. Goldberg, D.E., and Zimmerberg, J. (2020). Hardly Vacuous: The Parasitophorous Vacuolar Membrane of Malaria Parasites. Trends Parasitol. *36*, 138–146. https://doi.org/10.1016/j.pt.2019.11.006.

60. Koussis, K., Withers-Martinez, C., Yeoh, S., Child, M., Hackett, F., Knuepfer, E., Juliano, L., Woehlbier, U., Bujard, H., and Blackman, M.J. (2009). A multifunctional serine protease primes the malaria parasite for red blood cell invasion. EMBO J. *28*, 725–735. https://doi.org/10.1038/emboj.2009.22.

61. Garten, M., Nasamu, A.S., Niles, J.C., Zimmerberg, J., Goldberg, D.E., and Beck, J.R. (2018). EXP2 is a nutrient-permeable channel in the vacuolar membrane of Plasmodium and is essential for protein export via PTEX. Nat. Microbiol. *3*, 1090–1098. https://doi.org/10.1038/s41564-018-0222-7.

62. Mesén-Ramírez, P., Bergmann, B., Tran, T.T., Garten, M., Stäcker, J., Naranjo-Prado, I., Höhn, K., Zimmerberg, J., and Spielmann, T. (2019). EXP1 is critical for nutrient uptake across the parasitophorous vacuole

membrane of malaria parasites. PLoS Biol. *17*, e3000473. https://doi.org/10.1371/journal.pbio.3000473.

63. Tyagi, K., Hossain, M.E., Thakur, V., Aggarwal, P., Malhotra, P., Mohmmed, A., and Sharma, Y.D. (2016). Plasmodium vivax Tryptophan Rich Antigen PvTRAg36.6 Interacts with PvETRAMP and PvTRAg56.6 Interacts with PvMSP7 during Erythrocytic Stages of the Parasite. PLoS One *11*, e0151065. https://doi.org/10.1371/journal.pone.0151065.
64. MacKellar, D.C., Vaughan, A.M., Aly, A.S.I., DeLeon, S., and Kappe, S.H.I. (2011). A systematic analysis of the early transcribed membrane protein family throughout the life cycle of Plasmodium yoelii. Cell Microbiol. *13*, 1755–1767. https://doi.org/10.1111/j.1462-5822.2011.01656.x.
65. Spielmann, T., Montagna, G.N., Hecht, L., and Matuschewski, K. (2012). Molecular make-up of the Plasmodium parasitophorous vacuolar membrane. Int. J. Med. Microbiol. *302*, 179–186. https://doi.org/10.1016/j.ijmm.2012.07.011.
66. Maier, A.G., Rug, M., O'Neill, M.T., Brown, M., Chakravorty, S., Szestak, T., Chesson, J., Wu, Y., Hughes, K., Coppel, R.L., et al. (2008). Exported proteins required for virulence and rigidity of Plasmodium falciparum-infected human erythrocytes. Cell *134*, 48–61. https://doi.org/10.1016/j.cell.2008.04.051.
67. Wang, B., Lu, F., Cheng, Y., Chen, J.H., Jeon, H.Y., Ha, K.S., Cao, J., Nyunt, M.H., Han, J.H., Lee, S.K., et al. (2015). Immunoprofiling of the tryptophan-rich antigen family in Plasmodium vivax. Infect. Immun. *83*, 3083–3095. https://doi.org/10.1128/IAI.03067-14.
68. Aikawa, M., Miller, L.H., and Rabbege, J. (1975). Caveola–vesicle complexes in the plasmalemma of erythrocytes infected by Plasmodium vivax and P cynomolgi. Unique structures related to Schuffner's dots. Am. J. Pathol. *79*, 285–300.
69. Liu, B., Blanch, A.J., Namvar, A., Carmo, O., Tiash, S., Andrew, D., Hanssen, E., Rajagopal, V., Dixon, M.W.A., and Tilley, L. (2019). Multimodal analysis of Plasmodium knowlesi-infected erythrocytes reveals large invaginations, swelling of the host cell, and rheological defects. Cell Microbiol. *21*, e13005. https://doi.org/10.1111/cmi.13005.
70. Asare, K.K., Sakaguchi, M., Lucky, A.B., Asada, M., Miyazaki, S., Katakai, Y., Kawai, S., Song, C., Murata, K., Yahata, K., and Kaneko, O. (2018). The Plasmodium knowlesi MAHRP2 ortholog localizes to structures connecting Sinton Mulligan's clefts in the infected erythrocyte. Parasitol. Int. *67*, 481–492. https://doi.org/10.1016/j.parint.2018.04.005.
71. Lucky, A.B., Sakaguchi, M., Katakai, Y., Kawai, S., Yahata, K., Templeton, T.J., and Kaneko, O. (2016). Plasmodium knowlesi Skeleton-Binding Protein 1 Localizes to the 'Sinton and Mulligan' Stipplings in the Cytoplasm of Monkey and Human Erythrocytes. PLoS One *11*, e0164272. https://doi.org/10.1371/journal.pone.0164272.
72. al-Khedery, B., Barnwell, J.W., and Galinski, M.R. (1999). Antigenic variation in malaria: a 3' genomic alteration associated with the expression of a P. knowlesi variant antigen. Mol Cell *3*, 131–141. https://doi.org/10.1016/s1097-2765(00)80304-4.
73. Corredor, V., Meyer, E.V.S., Lapp, S., Corredor-Medina, C., Huber, C.S., Evans, A.G., Barnwell, J.W., and Galinski, M.R. (2004). A SICAvar switching event in Plasmodium knowlesi is associated with the DNA rearrangement of conserved 3' non-coding sequences. Mol. Biochem. Parasitol. *138*, 37–49. https://doi.org/10.1016/j.molbiopara.2004.05.017.
74. Korir, C.C., and Galinski, M.R. (2006). Proteomic studies of Plasmodium knowlesi SICA variant antigens demonstrate their relationship with P. falciparum EMP1. Infect. Genet. Evol. *6*, 75–79. https://doi.org/10.1016/j.meegid.2005.01.003.
75. Wang, B., Lu, F., Han, J.H., Lee, S.K., Cheng, Y., Nyunt, M.H., Ha, K.S., Hong, S.H., Park, W.S., and Han, E.T. (2016). Characterization of Caveola-Vesicle Complexes (CVCs) Protein, PHIST/CVC-8195 in Plasmodium vivax. Korean J. Parasitol. *54*, 725–732. https://doi.org/10.3347/kjp.2016.54.6.725.
76. Bernabeu, M., Lopez, F.J., Ferrer, M., Martin-Jaular, 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. https://doi.org/10.1111/j.1462-5822.2011.01726.x.
77. Fernandez-Becerra, C., Bernabeu, M., Castellanos, A., Correa, B.R., Obadia, T., Ramirez, M., Rui, E., Hentzschel, F., López-Montañés, M., Aylon-Hermida, A., et al. (2020). Plasmodium vivax spleen-dependent genes encode antigens associated with cytoadhesion and clinical protection. Proc. Natl. Acad. Sci. USA *117*, 13056–13065. https://doi.org/10.1073/pnas.1920596117.
78. Fougère, A., Jackson, A.P., Bechtsi, D.P., Braks, J.A.M., Annoura, T., Fonager, J., Spaccapelo, R., Ramesar, J., Chevalley-Maurel, S., Klop, O., et al. (2016). Variant Exported Blood-Stage Proteins Encoded by Plasmodium Multigene Families Are Expressed in Liver Stages Where They Are Exported into the Parasitophorous Vacuole. PLoS Pathog. *12*, e1005917. https://doi.org/10.1371/journal.ppat.1005917.
79. De Niz, M., Ullrich, A.K., Heiber, A., Blancke Soares, A., Pick, C., Lyck, R., Keller, D., Kaiser, G., Prado, M., Flemming, S., et al. (2016). The machinery underlying malaria parasite virulence is conserved between rodent and human malaria parasites. Nat. Commun. *7*, 11659. https://doi.org/10.1038/ncomms11659.
80. Duffy, M.F., Selvarajah, S.A., Josling, G.A., and Petter, M. (2014). Epigenetic regulation of the Plasmodium falciparum genome. Brief. Funct. Genomics *13*, 203–216. https://doi.org/10.1093/bfgp/elt047.
81. Ng, C.S., Sinha, A., Aniweh, Y., Nah, Q., Babu, I.R., Gu, C., Chionh, Y.H., Dedon, P.C., and Preiser, P.R. (2018). tRNA epitranscriptomics and biased codon are linked to proteome expression in Plasmodium falciparum. Mol. Syst. Biol. *14*, e8009. https://doi.org/10.15252/msb.20178009.
82. Baumgarten, S., Bryant, J.M., Sinha, A., Reyser, T., Preiser, P.R., Dedon, P.C., and Scherf, A. (2019). Transcriptome-wide dynamics of extensive m(6)A mRNA methylation during Plasmodium falciparum blood-stage development. Nat. Microbiol. *4*, 2246–2259. https://doi.org/10.1038/s41564-019-0521-7.
83. Ke, H., Sigala, P.A., Miura, K., Morrisey, J.M., Mather, M.W., Crowley, J.R., Henderson, J.P., Goldberg, D.E., Long, C.A., and Vaidya, A.B. (2014). The heme biosynthesis pathway is essential for Plasmodium falciparum development in mosquito stage but not in blood stages. J. Biol. Chem. *289*, 34827–34837. https://doi.org/10.1074/jbc.M114.615831.
84. Storm, J., and Müller, S. (2012). Lipoic acid metabolism of Plasmodium–a suitable drug target. Curr. Pharm. Des. *18*, 3480–3489. https://doi.org/10.2174/138161212801327266.
85. Tarun, A.S., Vaughan, A.M., and Kappe, S.H.I. (2009). Redefining the role of de novo fatty acid synthesis in Plasmodium parasites. Trends Parasitol. *25*, 545–550. https://doi.org/10.1016/j.pt.2009.09.002.
86. Boucher, M.J., Ghosh, S., Zhang, L., Lal, A., Jang, S.W., Ju, A., Zhang, S., Wang, X., Ralph, S.A., Zou, J., et al. (2018). Integrative proteomics and bioinformatic prediction enable a high-confidence apicoplast proteome in malaria parasites. PLoS Biol. *16*, e2005895. https://doi.org/10.1371/journal.pbio.2005895.
87. van Esveld, S.L., Meerstein-Kessel, L., Boshoven, C., Baaij, J.F., Barylyuk, K., Coolen, J.P.M., van Strien, J., Duim, R.A.J., Dutilh, B.E., Garza, D.R., et al. (2021). A Prioritized and Validated Resource of Mitochondrial Proteins in Plasmodium Identifies Unique Biology. mSphere *6*, e0061421. https://doi.org/10.1128/mSphere.00614-21.
88. Oehring, S.C., Woodcroft, B.J., Moes, S., Wetzel, J., Dietz, O., Pulfer, A., Dekiwadia, C., Maeser, P., Flueck, C., Witmer, K., et al. (2012). Organellar proteomics reveals hundreds of novel nuclear proteins in the malaria parasite Plasmodium falciparum. Genome Biol. *13*, R108. https://doi.org/10.1186/gb-2012-13-11-r108.
89. Foth, B.J., Zhang, N., Chaal, B.K., Sze, S.K., Preiser, P.R., and Bozdech, Z. (2011). Quantitative time-course profiling of parasite and host cell proteins in the human malaria parasite Plasmodium falciparum. Mol. Cell. Proteomics *10*, M110.006411. https://doi.org/10.1074/mcp.M110.006411.

90. Hausser, J., Mayo, A., Keren, L., and Alon, U. (2019). Central dogma rates and the trade-off between precision and economy in gene expression. Nat. Commun. *10*, 68. https://doi.org/10.1038/s41467-018-07391-8.

91. Ito, D., Schureck, M.A., and Desai, S.A. (2017). An essential dual-function complex mediates erythrocyte invasion and channel-mediated nutrient uptake in malaria parasites. Elife *6*, e23485. https://doi.org/10.7554/eLife.23485.

92. Moon, R.W., Hall, J., Rangkuti, F., Ho, Y.S., Almond, N., Mitchell, G.H., Pain, A., Holder, A.A., and Blackman, M.J. (2013). Adaptation of the genetically tractable malaria pathogen Plasmodium knowlesi to continuous culture in human erythrocytes. Proc. Natl. Acad. Sci. USA *110*, 531–536. https://doi.org/10.1073/pnas.1216457110.

93. Przyborski, J.M., and Lanzer, M. (2005). Protein transport and trafficking in Plasmodium falciparum-infected erythrocytes. Parasitology *130*, 373–388. https://doi.org/10.1017/s0031182004006729.

94. Omelianczyk, R.I., Loh, H.P., Chew, M., Hoo, R., Baumgarten, S., Renia, L., Chen, J., and Preiser, P.R. (2020). Rapid activation of distinct members of multigene families in Plasmodium spp. Commun. Biol. *3*, 351. https://doi.org/10.1038/s42003-020-1081-3.

95. Pedregosa, F.V., Alexandre, G., Michel, G., Thirion, V., Grisel, B., Blondel, O., Prettenhofer, M., Weiss, P., Dubourg, R., Vanderplas, V., et al. (2011). Scikit-learn: Machine Learning in Python. https://www.jmlr.org/papers/v12/pedregosa11a.html.

96. Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat. Commun. *10*, 1523. https://doi.org/10.1038/s41467-019-09234-6.

97. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238. https://doi.org/10.1186/s13059-019-1832-y.

98. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods *12*, 59–60. https://doi.org/10.1038/nmeth.3176.

99. Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Mol. Biol. Evol. *33*, 1635–1638. https://doi.org/10.1093/molbev/msw046.

100. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675. https://doi.org/10.1038/nmeth.2089.

101. Walliker, D., Quakyi, I.A., Wellems, T.E., McCutchan, T.F., Szarfman, A., London, W.T., Corcoran, L.M., Burkot, T.R., and Carter, R. (1987). Genetic analysis of the human malaria parasite Plasmodium falciparum. Science *236*, 1661–1666. https://doi.org/10.1126/science.3299700.

102. Trager, W., and Jensen, J.B. (1976). Human malaria parasites in continuous culture. Science *193*, 673–675. https://doi.org/10.1126/science.781840.

103. Lambros, C., and Vanderberg, J.P. (1979). Synchronization of Plasmodium falciparum erythrocytic stages in culture. J. Parasitol. *65*, 418–420.

104. Kutner, S., Breuer, W.V., Ginsburg, H., Aley, S.B., and Cabantchik, Z.I. (1985). Characterization of permeation pathways in the plasma membrane of human erythrocytes infected with early stages of Plasmodium falciparum: association with parasite development. J. Cell. Physiol. *125*, 521–527. https://doi.org/10.1002/jcp.1041250323.

105. Ngernna, S., Chim-Ong, A., Roobsoong, W., Sattabongkot, J., Cui, L., and Nguitragool, W. (2019). Efficient synchronization of Plasmodium knowlesi in vitro cultures using guanidine hydrochloride. Malar. J. *18*, 148. https://doi.org/10.1186/s12936-019-2783-1.

106. Janse, C.J., Ramesar, J., and Waters, A.P. (2006). High-efficiency transfection and drug selection of genetically transformed blood stages of the rodent malaria parasite Plasmodium berghei. Nat. Protoc. *1*, 346–356. https://doi.org/10.1038/nprot.2006.53.

107. Serra, A., Gallart-Palau, X., Park, J.E., Lim, G.G.Y., Lim, K.L., Ho, H.H., Tam, J.P., and Sze, S.K. (2018). Vascular Bed Molecular Profiling by Differential Systemic Decellularization In Vivo. Arterioscler. Thromb. Vasc. Biol. *38*, 2396–2409. https://doi.org/10.1161/atvbaha.118.311552.

108. Sinton, J.A., and Mulligan, H.W. (1933). A critical review of the literature relating to the identification of the malarial parasites recorded from monkeys of the families Cercopithecidae and Colobidae. Records of the Malaria Survey of India *3*, 381–443.

109. Ingmundson, A., Nahar, C., Brinkmann, V., Lehmann, M.J., and Matuschewski, K. (2012). The exported Plasmodium berghei protein IBIS1 delineates membranous structures in infected red blood cells. Mol. Microbiol. *83*, 1229–1243. https://doi.org/10.1111/j.1365-2958.2012.08004.x.

110. Grüring, C., Heiber, A., Kruse, F., Ungefehr, J., Gilberger, T.W., and Spielmann, T. (2011). Development and host cell modifications of Plasmodium falciparum blood stages in four dimensions. Nat. Commun. *2*, 165. https://doi.org/10.1038/ncomms1169.

111. McMillan, P.J., Millet, C., Batinovic, S., Maiorca, M., Hanssen, E., Kenny, S., Muhle, R.A., Melcher, M., Fidock, D.A., Smith, J.D., et al. (2013). Spatial and temporal mapping of the PfEMP1 export pathway in Plasmodium falciparum. Cell Microbiol. *15*, 1401–1418. https://doi.org/10.1111/cmi.12125.

112. Birnbaum, J., Flemming, S., Reichard, N., Soares, A.B., Mesén-Ramírez, P., Jonscher, E., Bergmann, B., and Spielmann, T. (2017). A genetic system to study Plasmodium falciparum protein function. Nat. Methods *14*, 450–456. https://doi.org/10.1038/nmeth.4223.

113. Fidock, D.A., and Wellems, T.E. (1997). Transformation with human dihydrofolate reductase renders malaria parasites insensitive to WR99210 but does not affect the intrinsic activity of proguanil. Proc. Natl. Acad. Sci. USA *94*, 10931–10936. https://doi.org/10.1073/pnas.94.20.10931.

114. Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005). Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol. Cell. Proteomics *4*, 1265–1272. https://doi.org/10.1074/mcp.M500061-MCP200.

115. Burki, F., Roger, A.J., Brown, M.W., and Simpson, A.G.B. (2020). The New Tree of Eukaryotes. Trends Ecol. Evol. *35*, 43–55. https://doi.org/10.1016/j.tree.2019.08.008.

116. Burki, F., Kudryavtsev, A., Matz, M.V., Aglyamova, G.V., Bulman, S., Fiers, M., Keeling, P.J., and Pawlowski, J. (2010). Evolution of Rhizaria: new insights from phylogenomic analysis of uncultivated protists. BMC Evol. Biol. *10*, 377. https://doi.org/10.1186/1471-2148-10-377.

117. Escalante, A.A., Freeland, D.E., Collins, W.E., and Lal, A.A. (1998). The evolution of primate malaria parasites based on the gene encoding cytochrome b from the linear mitochondrial genome. Proc. Natl. Acad. Sci. USA *95*, 8124–8129. https://doi.org/10.1073/pnas.95.14.8124.

118. Galen, S.C., Borner, J., Martinsen, E.S., Schaer, J., Austin, C.C., West, C.J., and Perkins, S.L. (2018). The polyphyly of Plasmodium: comprehensive phylogenetic analyses of the malaria parasites (order Haemosporida) reveal widespread taxonomic conflict. R. Soc. Open Sci. *5*, 171780. https://doi.org/10.1098/rsos.171780.

119. Miroliubova, T.S., Simdyanov, T.G., Mikhailov, K.V., Aleoshin, V.V., Janouškovec, J., Belova, P.A., and Paskerova, G.G. (2020). Polyphyletic origin, intracellular invasion, and meiotic genes in the putatively asexual agamococcidians (Apicomplexa incertae sedis). Sci. Rep. *10*, 15847. https://doi.org/10.1038/s41598-020-72287-x.

120. Domazet-Loso, T., Brajković, J., and Tautz, D. (2007). A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. Trends Genet. *23*, 533–539. https://doi.org/10.1016/j.tig.2007.08.014.

121. Benjamini, Y., and Hochberg, Y. (1995). Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing. J. Royal Statist. Soc., Series B *57*, 289–300. https://doi.org/10.2307/2346101.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rat anti-HA	Roche	Cat#11867423001; RRID:AB_390918
Rabbit anti-band3	Abcam	Cat#ab108414; RRID:AB_10862297
Rabbit anti- RHAG	Abcam	Cat#ab155094
Rabbit anti-EXP2	Siau et al.[15]	N/A
Rabbit anti-aldolase	This paper	N/A
Goat anti-rat AF594	Jackson ImmunoResearch	Cat#112-585-062; RRID:AB_2338377
Goat anti-rabbit DyLight 488	Invitrogen	Cat#35553; RRID:AB_1965947
Goat anti-rabbit AF594	Invitrogen	Cat#A11012; RRID:AB_2534079
Goat anti-rabbit AF647	Invitrogen	Cat#A21245; RRID:AB_141775
Bacterial and virus strains		
XL10-Gold Ultracompetent Cells	Agilent	Cat#200315
Chemicals, peptides, and recombinant proteins		
Percoll®	MP biochemicals	Cat#0219536990
D-Sorbitol	Sigma	Cat#240850
Histodenz™	Sigma	Cat#D2158
Guanidine hydrochloride	Sigma	Cat#50950
WR99210	Jacobus Pharmaceuticals	N/A
G418 Sulfate	Gold Biotechnology	Cat#G-418-25
Pyrimethamine	MP biochemicals	Cat#0219418025
Hoechst33342	Thermo Scientific	Cat#62249
DAPI, FluoroPure™ grade	Invitrogen	Cat#D21490
Saponin	Sigma	Cat#47036
Streptolysin O	Abcam	Cat#ab126650
Imperial™ Protein Stain	Thermo Scientific	Cat#24615
Sequencing-grade modified trypsin	Promega	Cat#V5111
Halt™ Protease Inhibitor Cocktail	Thermo Scientific	Cat#78430
Critical commercial assays		
ProteoExtract® Complete Proteome Extraction Kit	Millipore	Cat#539779
Micro BCA™ Protein Assay Kit	Thermo Scientific	Cat#23235
TMT10plex™ Isobaric Label Reagent Set	Thermo Scientific	Cat#90406
KOD -Plus- *Neo*	Toyobo	Cat#KOD-401
NEBuilder HiFi DNA Assembly Master Mix	NEB	Cat#E2621
FavorPrepTM Plasmid Extraction Mini Kit	Favorgen	Cat#FAPDE 100
NucleoBond Xtra Midi Plus EF, Midi kit	Macherey-Nagel	Cat#740422.50
Basic Parasite Nucleofector® Kit 2	Lonza	Cat#VMI-1021
P3 Primary Cell 4D-Nucleofector® X Kit L	Lonza	Cat#V4XP-3012
NucleoSpin Blood, Mini kit for DNA from blood	Macherey-Nagel	Cat#740951.50
Experimental models: Organisms/strains		
P. falciparum 3D7	MR4	N/A
P. knowlesi A1-H.1	Moon et al.[92]	N/A
P. yoelii yoelii 17X 1.1	MR4	N/A
P. berghei ANKA	MR4	N/A
P. chabaudi chabaudi AS	MR4	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
BALB/cAnNTac, Male, 5–6 weeks	Invivos	N/A
Recombinant DNA		
pARL	Przyborski et al.[93]	N/A
pSLI-2×FKBP-HA	Omelianczyk et al.[94]	N/A
PKePL-HA	This paper	N/A
PKePL-eGFP	This paper	N/A
PYePL-eGFP	Siau et al.[14]	N/A
PY-T2A-HA	This paper	N/A
Software and algorithms		
PlasmoDB	VEuPathDB	plasmodb.org
Xcalibur™ Software	Thermo Scientific	Cat#OPTON-30965
Proteome Discoverer	Thermo Scientific	Cat#OPTON-31099
Scikit-Learn	Pedregosa et al.[95]	scikit-learn.org
Metascape	Zhou et al.[96]	metascape.org
OrthoFinder	Emms and Kelly.[97]	davidemms.github.io
DIAMOND	Buchfink et al.[98]	github.com/bbuchfink/diamond
ETE tookit	Huerta-Cepas et al.[99]	etetoolkit.org
ImageJ	Schneider et al.[100]	ImageJ.net
Deposited data		
Mass Spectrometric raw data	This Paper; PRIDE	PXD031796
Homo sapiens reference proteome	EMBL-EBI	UP000005640
Mus musculus reference proteome	EMBL-EBI	UP000000589
Common contaminant proteins	GPMDB	thegpm.org/crap/index.html
Others		
Sep-Pak C18 3 cc Vac Cartridge	Waters	Cat#WAT054945
XBridge Peptide BEH C18 Column 30Å, 3.5 μm	Waters	Cat#186003570
EASY-Spray™ Source	Thermo Scientific	Cat#ES081
Acclaim™ PepMap™ 100 C18 HPLC Columns	Thermo Scientific	Cat#164570

RESOURCE AVAILABILITY

Lead contact

Further information and requests should be directed to the Lead Contact, Peter R. Preiser (prpreiser@ntu.edu.sg).

Materials availability

Request for plasmids, antibodies or resources used/generated in this study should be directed to the Lead Contact, Peter R. Preiser (prpreiser@ntu.edu.sg).

Data and code availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD031796. This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Human samples

Human whole blood was donated by healthy male adult volunteers with no previous malaria history either at the National University Hospital, Singapore or Fullerton Healthcare Clinic, Nanyang Technological University. Donors were required to be between the age of 18–65 years old. Informed consents were obtained from all donors in accordance with the protocol approved by the Institutional Review Board of Nanyang Technological University, Singapore (IRB-2018-02-031, IRB-2019-09-047). Ancestry, race, ethnicity, and socioeconomic status were not taken into consideration for this study.

Cultivation of *P. falciparum* 3D7 and *P. knowlesi* A1-H.1

P. falciparum[101] were cultured in human red blood cells donated within 30 days prior to usage with RPMI1640 media supplemented with 10 mg/L gentamicin, 2 g/L sodium bicarbonate, 0.25% Albumax II, and 0.1mM hypoxanthine.[102] Cultures were synchronized using 5% D-sorbitol[103] (Sigma) or by 68% Percoll[104] (MP biochemicals) purification. *P. knowlesi*[92] were cultured in blood donated within 14 days prior to usage with RPMI1640 media supplemented with 10 mg/L gentamicin, 2 g/L sodium bicarbonate, 4 g/L dextrose, 0.5% Albumax II, 25mM HEPES, 0.1mM hypoxanthine, 2μM L-glutamine and 10% horse serum. Cultures were synchronized using 140 mM guanidine hydrochloride[105] (Sigma) or by 55% Histodenz[92] (Sigma) purification. All cultures were maintained at 37°C under microaerophilic conditions and synchronization of the parasites were verified by stage morphology of the parasites using Giemsa-stained thin blood smears.

Animal samples

All animal work was carried out in strict accordance with the recommendations of the NACLAR (National Advisory Committee for Laboratory Animal Research) guidelines under the Animal & Birds (Care and Use of Animals for Scientific Purposes) Rules of Singapore. The protocol was approved by the Institutional Animal Care and Use Committee (IACUC) of Nanyang Technological University, Singapore (ARF SBS/NIE-A-0223, ARF SBS/NIE-A-0379). All efforts were made to minimize suffering.

Cultivation of P. yoelii yoelii 17X 1.1, P. berghei ANKA and P. Chabaudi chabaudi AS

BALB/cAnNTac male mice of 5–6 weeks old (Invivos) were infected with cryopreserved parasite stabilates by intraperitoneal injection and used as parasite donor at 15–60% parasitemia. Infected RBCs was collected through cardiac puncture with syringe containing 2 mg/ml heparin (Sigma) solution. Infected RBC pellets were passed through a 50%–60% Histodenz gradient to isolate trophozoite/schizonts stage parasite.[106] Synchronization of the parasites were verified by stage morphology of the parasites using Giemsa-stained thin blood smears.

METHOD DETAILS

Fractionation of the parasite samples

Highly purified samples of late trophozoite/early schizont parasites obtained by Percoll (*P. falciparum*) or Histodenz (*P. knowlesi*, *P. yoelii, P. berghei and P. chabaudi)* gradient were verified using Giemsa-stained thin blood smears to ensure that they were all stage matched. The parasites samples were fractionated by sequential lysis using streptolysin O (SLO) and saponin followed by differential centrifugation in the presence of protease inhibitor (Thermo Scientific).[15,41] Briefly, 100-200μL of infected RBC pellet was first treated with ∼10U/μl of SLO (Abcam) for 2–4 min at room temperature to permeabilize the RBC membrane without affecting the PVM and centrifugated at 800 g to separate the parasite within the PVM (pellet) from the RBC cytoplasm and RBC membrane (supernatant). Following three washes using incomplete RPMI, the pellet was then treated with 0.05% of saponin (Sigma) for ∼2 min at room temperature and centrifuged to separate the PV/PVM component (supernatant) from the intraerythrocytic parasite (pellet). Following three washes using incomplete RPMI, the pellet fraction that contains the internal parasite was dissolved in Laemmli sample buffer (Sigma). SLO and saponin supernatants were clarified by centrifugation at 20,000 g for 40 min to isolate the RBC membrane (GHOST) or the PVM (PVM1), respectively. Both RBC membrane and PVM1 pellets were washed 4 times by resuspension in incomplete RPMI and centrifuged at 20,000 g for 40 min. Following the first centrifugation at 20,000 g, the resulting SLO and Saponin supernatants were ultracentrifuged at 300,000 g for 60 min to separate the specialized structure (VESICLES/CLEFT's) from the soluble content of the HCC (HCC) and the PVM content that was not isolated by the first centrifugation (PVM2) from the PV content (PV). In total, seven fractions were produced and dissolved in equivalent amount of Laemmli sample buffer: GHOST, VESICLES/CLEFTS, HCC, PVM1, PVM2, PV and PARASITE. For each parasite sample fractionated, equivalents to 3μL of infected RBC pellet were analyzed by Western blot as described below. Out of >50 samples fractionated across the five *Plasmodium* species studied, 3 *P falciparum*, 3 *P knowlesi*, 2 *P yoelii*, 2 *P berghei*, and 2 *P chabaudi* (n = 12) displaying Western blot patterns similar to those shown in the Figure 1B, with no or minimal cross contamination between subcellular fractions, were retained for LC-MS/MS analysis.

Inherent to all sub-cellular fractionations, a small amount of cross-contaminant was observed across the fractions likely due to over/under lysis of the infected RBC membranes or insufficient enrichment during the differential centrifugation. However, such cross-contaminating effect was reduced by the subsequent analysis of the subcellular fraction using quantitative LC-MS/MS.

Western blot analysis of parasite preparations after subcellular fractionation

Fluorescent Western blot was performed on subcellular fractions of infected RBCs lysed with SLO and saponin using Licor Near-Infrared Western Blot Detection protocol (https://www.licor.com/bio/support/). The membrane proteins were probed using anti-human Band 3 (*P. falciparum* and *P. knowlesi*) (Abcam), anti-mouse RHAG (*P. yoelii, P. berghei and P. chabaudi*) (Abcam), anti-aldolase (Genscript) raised against the peptide ADESTQTIKKRFDN and anti-EXP2 (Genscript) recognizing the peptide KNIESGKYEFDVD. All the proteins were revealed using DyLight 488, Alexa Fluor 594 or Alexa Fluor 647 secondary antibodies (Invitrogen), coupled to a GE typhoon trio scanner.

Proteome extraction, purification, and preparation

To map the *Plasmodium* proteome, proteins included in the subcellular samples derived from three (*P. falciparum* and *P. knowlesi*) and two (*P. yoelii, P. berghei* and *P. chabaudi*) biological replicates were dissolved in TEAB lysis buffer (2% SDS in 100mM triethylammonium bicarbonate) supplemented with protease inhibitor except the PARASITE pellets which were dissolved using ProteoExtract Complete Mammalian Proteome Extraction Kit (Millipore) following manufacturer instruction at 4°C. The homogenates were centrifuged for 10 min at 10,000 g to collect the supernatant. As the sub-proteomes contain detergents, <200μg of proteins from each fraction as determined by Micro BCA Protein Assay (Thermo Scientific) were separated on 12% SDS-PAGE at 50V and visualized by staining with imperial protein stain (Thermo Scientific). Each lane of the SDS gel was cut into 1 mm^2 cubes. The gel cubes were distained with three cycles of alternation of 0.25M TEAB with or without 50% (v/v) acetonitrile (Thermo Scientific) prior to gel shrinking following addition of 100% acetonitrile and air drying. The protein contained in the gel pieces were reduced with 10mM dithiothreitol (Sigma) in 25mM TEAB for 30 min at 60°C and alkylated with 55mM iodoacetamide (Sigma) in 25mM TEAB for 30 min at room temperature, followed by digestion at 37°C with sequencing-grade modified trypsin (Promega) at 1:50 trypsin to protein w/w ratio overnight. The peptides were extracted using 50% acetonitrile with 5% acetic acid buffer (Merck) three times with vortexing. The extracted peptides were pooled, dried and concentrated by spin vacuum concentrator (Eppendorf). The tryptic peptides were re-dissolved in 100mM TEAB for labeling with TMT 10-plex isobaric tags (Thermo Scientific) according to manufacturer's protocol. As there are seven subcellular proteomes (GHOST, VESICLE, HCC, PVM1, PVM2, PV and PARASITE) for TMT-10plex, 3 subproteomes (GHOST, VESICLE, HCC) were labeled with two different TMT tags determined by drawing lots and served as internal technical replicates for a total of ten tags used. The ten labeled peptide samples were pooled, subsequently desalted using a C-18 Sep-pack 200 mg cartridge (Waters), and the eluates were dried using vacuum concentrator.

High-pressure liquid chromatography fractionation

Each set of TMT labeled sample was subjected to HPLC fractionation as previously described.[107] Briefly, dried peptides were reconstituted in 200μL of mobile phase A (10 mmol/L ammonium hydroxide in water) and fractionated using a Xbridge BEH130 C18 column (3.5μm 4.6 × 250mm; Waters) on a Shimadzu Prominence UFLC system (Shimadzu) with UV monitoring of peptide intensities at 280nm. Peptide separation was performed at 1 mL/min using a 60-min gradient as follows: 0–5% B (0 mmol/L ammonium hydroxide in ACN) for 3 min, 5–35% B for 40 min, 35–70% B for 12 min and 70–100% for 5 min. Fractions were collected at 1 min intervals and combined by concatenation. Concatenated fractions were then completely dried in the vacuum concentrator.

LC-MS/MS analysis of TMT labeled proteins

Dried sample fractions were carefully reconstituted in mobile phase A (3% ACN, 0.1% FA) before their analysis by LC-MS/MS using a Dionex Ultimate 3000 RSLCnano system (Thermo Scientific) coupled with a Q-Exactive tandem mass spectrometer (Thermo Scientific). Spray was generated using an EASY-Spray ion source (Thermo Scientific) working at 1.5 kV. Peptide separation was performed using a PepMap C18 column (Thermo Scientific) maintained at 35°C. Separation of peptides was performed over a 60 min gradient with mobile phase A (0.1% FA in water) and mobile phase B (0.1% FA in 100% ACN) as follows: 3−30% B for 45 min, 30−50% B for 9 min, 50−80% B for 1 min, 80% B for 2 min, and finally maintained isocratic at 3% B for 3 min. Q-Exactive data acquisition was performed in positive ion mode using Xcalibur 3.0.63 software (Thermo Scientific) alternating Full Fourier transform MS (FT-MS; 350-1,600 m/z range, resolution of 70,000 at m/z 200, 1 μscan per spectrum) and FT-MS/MS (resolution 35,000) for the 10 most intense ions with charge >+2 and isolated within a 2 Da window. Fragmentation of ions was performed by high energy collisional dissociation fragmentation mode using 28% normalized collision energy. A threshold of 500 counts was enabled. For full FT-MS and FT-MS/MS, automatic gain control was set to 5×10^6 and 2×10^5, respectively.

Mass Spectrometric data analysis

The acquired data was processed using Proteome Discoverer version 1.4 software (Thermo Scientific). The raw files were directly imported into PD1.4 and further processed using designed workflow. Briefly, this workflow includes eight processing nodes numbered from 0 to 8. Node 0 named "spectrum file" allows selecting raw files, node 1 labeled as "spectrum selector" extracts the spectra within a retention time window and precursor ion mass window. Node 2 is an MS2-Spectrum Processor for deisotoping and deconvoluting isotopic cluster in MS/MS spectra. Node 3 connect to Mascot search engine, while node 5 (Sequest HT) select search engine SEQUEST. In both search engines, the search parameters set were enzyme: trypsin; maximum miss cleavage: 2; minimum peptide length: 6, maximum peptide length: 144; precursor mass tolerance: 10 ppm, fragment mass tolerance: 0.02 Da; static/fixed modification: carbamydomethylation (C) and TMT 10plex tag (K, peptide N-terminal); dynamic/variables modification: oxidation (M), and deamidation (N, Q). For each malaria specie, a database including 1) the ~20000 human proteins included in the Homo sapiens reference proteome (UP000005640) or the ~22000 mouse proteins obtained from the Mus musculus reference proteome (UP000000589) (https://www.ebi.ac.uk/reference_proteomes), 2) the cognate *Plasmodium* proteomes from PlasmoDB release 34 (https://plasmodb.org/plasmo/app/downloads/release-34) and 3) 2 sets of 245 and 116 common contaminant proteins (http://www.coxdocs.org/doku.php?id= maxquant:start_downloads.html and https://thegpm.org/crap/index.html, respectively) was constituted and used for the searches. Node 4 called "Fixed Value PSM Validator" or "Targeted Decoy PSM Validator" where target FDR (strict) was set as 0.01, target FDR (relaxed) was set as 0.05. Node 6 named reporter ions quantifier, where the parameter set were Quantification method: TMT 10plex, peak integration tolerance: 10 ppm, integration method: most confident centroid, mass

analyzer: FTMS, MS order: MS2, activation type: HCD, minimum collision energy; 0 and maximum collision energy: 100. Node 7 is the Event Detector with Mass precision of 2ppm; Node 8 called Precursor Ions Area Detector for label-free quantitation of the identified proteins. The obtained peptide/protein list was exported to Microsoft Excel or processed using an in-house script for further analysis.

Generation of training dataset

Training sets were generated by searching among the proteins detected in the proteomic approach, those with localization information in literature (Table S1 training sets). Only proteins with localization data obtained by immunolabelling or protein tagging approaches using non-invasive asexual blood stage parasites were included. The proteins were sorted according to their most external localization: RBC membrane (GHOST), the host-cell-cytoplasm specialized structure (VESICLE), the host-cell-cytoplasm soluble protein (HCC), the parasitophorous vacuole membrane (PVM), the parasitophorous vacuole (PV) and the internal parasite (PARASITE) (Table S1 training sets). Of note, the additional CLEFT's subset includes proteins characterized in the *P. falciparum* Maurer's clefts and in its lesser studied counterparts *P. knowlesi* Sinton and Mulligan's stippling[108] and *P. berghei* Maurer's clefts-like structure.[109] These proteins are mostly enriched in the GHOST and to some extent in the VESICLE fractions as Maurer's clefts are known to be mobile in the host cytosol and then anchored to the RBC membrane via tethers throughout the blood stage development.[110,111] Similarly, proteins previously annotated as localized at the periphery based on fluorescence microscopy observations were considered as enriched in the PVM, the PV or the PARASITE fractions as such technique cannot differentiate the PVM and PV compartments from the PPM included in the PARASITE fraction. Reflecting the low number of remodeling proteins currently known, the training sets were imbalanced with a predominance of intra-parasitic protein which could lead to a suboptimal refinement (Figure 2B, columns "n"). To improve the refinement outcome, the training sets were further enriched by inclusion of a) multigene families previously characterized in the periphery and/or the host cell, b) malaria molecules annotated as "exported" or "tryptophan-rich protein" in the malaria databases as well as c) host proteins localized in the RBC membrane (Table S1 training sets).

Prediction of the malaria protein localization

In the first prediction strategy, an "iterative Permutation" (P) algorithm was used to derive for each proteomic dataset, the normalization factors providing the highest level of correlation with the cognate training set (Figure 2A, right). Following normalization, the localization of the remaining proteome was then assigned based on their relative abundance patterns since plasmodial molecules found preferentially in one fraction compared to the other fractions are likely associated to this fraction.[15] The iterative algorithm used to derive the optimal values was built using Microsoft Excel. For each proteomic dataset, the normalization values were first set to 1 in each fraction and a calculation is made to determine the percentage of correlation with cognate training sets, with internal parasite and remodeling proteins expected to have their peaks of relative abundance in the intra-parasitic or extra-parasitic compartment(s), respectively (Table S1 iterative Permutation). For each subcellular fraction, this calculation is repeated over 1200 permutation of the normalization factors used, by adding 0.001 at each permutation (i.e., +0.01, +0.02 + 0.03 etc. …). One thousand two hundred other permutations were also calculated by subtracting 0.01 at each permutation (i.e., −0.01, −0.02 -0.03 etc. …). The normalization values were further refined in a second round of iterative permutation by adding or subtracting 0.001 to the normalization factors identified following the first round. Except for the PARASITE faction which normalization factor was set to 1, the normalization factors applied to the remaining subcellular fraction were refined using the iterative algorithm. For each proteomic dataset, the combination of normalization factor providing the highest level of correlation with the cognate training sets was retained (Table S1 iterative permutation). The proteomic datasets refined with the optimal normalization factor were then mapped by differential analysis of the relative abundance of the proteins across the subcellular fractions and identification of the compartment that contains most of the proteins.

The second prediction strategy use a parallel ensemble "Machine Learning" (ML) approach that combine the outputs obtained from several supervised machine learning classifiers to achieve better predictive performances (Table S1 machine learning; Figure 2A, left). The proteomics data (Table S1) was analyzed in Python with scikit-learn.[95] Subcellular localizations from literature (Table S1 training sets) were used in a supervised learning procedure to predict whether a given protein is found in a compartment. The training sets use '1′ and '0′ to indicate presence and absence, respectively, of a protein in a compartment. Overall, five classifiers were used including 1) Support Vector Machine (SVM), 2) k-Nearest Neighbors (kNN), 3) Multi-layer Perceptron (MLP), 4) random forest (RF) and 5) Decision Trees (DT). These classifiers were trained with the cognate localization datasets to determine the optimal parameters required to connect in each subcellular fraction, the relative abundances of the proteins included in the trainings sets to the experimental evidence of their presences. Training sets were split into training (80% of data) and test (20%) datasets with scikit-learn train_test_split function, to ensure that the percentage of positive (1) and negative (0) localizations of proteins to a given compartment are similar between the train and test datasets. The optimal parameters determined using an exhaustive 5-fold cross-validated grid-search over a parameter grid were as follow: SVM ('svm__kernel': ['linear', 'poly', 'rbf', 'sigmoid'], 'svm__gamma': [1e-5,1e-4,1e-3,1e-2,1e-1,'auto'],'svm__C': [1, 10, 100, 1000]), kNN ({'knn__n_neighbors': list(range(2,21,2)), 'knn__weights': ['uniform', 'distance'], "knn__leaf_size": [3,10,30,60]}), DC ({'dcc__max_depth': list(range(10, 251,5))+[None], "dcc__min_samples_split": range(2, 10)}), MLP ({'mlp__solver': ['lbfgs', 'sgd', 'adam'], 'mlp__activation': ['identity', 'logistic', 'tanh', 'relu'], "mlp__alpha": [1e-5,1e-4,1e-3,1e-2,1e-1], "mlp__hidden_layer_sizes":(500, 50)}) and RF({'rfc__max_depth': list(range(5, 50,10))+[None],"rfc__n_estimators": [2,5,10,20,50], "rfc__min_samples_split": range(2, 10)}). The likelihood of presence of the protein in a specific compartment was predicted by each classifier as a confidence score ranging from 0 (no evidence) to 1 (high confidence) (Table S1 machine learning). The five classifiers were running in parallel, and their

outputs combined to calculate for each protein a mean confidence score in each subcellular fraction. The patterns formed by the mean confidence scores were then analyzed. The remodeling proteins with confidence score(s) in the remodeling fractions superior to this in the PARASITE fraction are predicted based to the compartment that contains the highest confidence score (Figure 2A, right red). The remaining proteins are either remodeling proteins that are concomitantly present in the internal parasite or true intra-parasitic protein. To distinguish them, we defined for each dataset, a relative background threshold imposed according to the confidence score in the intra-parasitic fraction for which the >90% of the known internal parasite proteins are correctly predicted (i.e., with a confidence score in the remodeling fractions inferior to the threshold). This allowed us to distinguish the remodeling proteins characterized by confidence score(s) superior to the threshold in the extra-parasitic compartments (Figure 2A, right green) from the internal proteins (Figure 2A, right blue). Finally, the proteomic datasets were mapped by differential analysis of the confidence score patterns versus the relative thresholds defined for each *Plasmodium* species.

Derivation of a consensus localization

To identify consensus localizations for each orthology group constituted using PlasmoDB (https://plasmodb.org/) and an iterative three-steps process was used. (Figure 3A)

i) Identification of consensus localizations using the main localization prediction (Figure 3Ai). The orthology groups with a subcellular localization shared by > 50% of the predictions were annotated with this specific localization. Furthermore, an estimation of the prediction confidence was also provided with confidence levels ranging from "low", "medium" and "high" for localizations shared by >50% and <75%, >75% and <100% and 100% of the prediction, respectively.

ii) Addition of secondary localization (Figure 3Aii). While P and ML approaches were designed to predict the most probable localization, malaria protein may have multiple localizations. In addition, the whole approach may be subjected to experimental/technical/predictive variations that might affect the outcome when the predictions are within the same range of relative abundance or confidence score. Thus, secondary localizations, if available, may contribute to refining the prediction. For P approach, it consists of the subcellular compartment with the second highest abundance while for ML approach, we considered the second compartment with the highest confidence scores > the background threshold. To maximize accuracy, only secondary localizations considered as within a similar range of the main localization were used to refine prediction. This range was arbitrarily defined at 20% of the main localization relative abundances or confidence scores (Table S1 iterative permutation and machine learning; Table S2 proteome mapping). Following inclusion of secondary localizations, the unannotated orthology groups were reanalyzed to search for a predominant localization. These predictions were considered as having a "low" confidence level.

iii) Identification of Cleft's proteins (Figure 3Aiii). Cleft's proteins are concurrently enriched in the GHOST and the VESICLE fractions as Maurer's clefts are known to be both mobile in the host-cytosol and anchored to the host RBC membrane. Thus, orthology groups with GHOST and VESICLE localization representing >50% of the predictions were annotated as Cleft's proteins and considered as having a "low" confidence level. The remaining orthology groups were left without consensus localization.

P. falciparum plasmid construction and transfection

To generate the selection-linked integration (SLI) knock-in constructs for HA-tagging, the GFP tag was replaced with a codon optimized 3xHA tag in the pSLI-2×FKBP-GFP plasmid[112] to obtain the pSLI-2×FKBP-HA construct.[94] Target homology arm fragments of the gene of interests (without stop codon) were amplified using KOD-Plus-Neo polymerase (Toyobo) from *P. falciparum* genomic DNA and inserted between the NotI/MluI sites of pSLI-2×FKBP-HA.

For episomal overexpression, appropriate coding sequences of the full-length target genes were amplified from *P. falciparum* genomic DNA and assembled via NEbuilder assembly (NEB) into the AvrII/XhoI sites of the pARL plasmid for GFP-tagging.[93]

Synchronized ring stage parasites were transfected with 100–200μg of midiprep kit purified plasmids (Macherey-Nagel) using Bio-Rad gene pulser Xcell electroporator[113] Transfectants were selected with 2.5nM WR99210 (Jacobus Pharmaceuticals) 24 h post transfection. Secondary selection of the pSLI constructs for endogenous tagging was done with 400 μg/ml of G418 Sulfate (Gold Biotechnology), until parasites were observed. Drug selection pressure was maintained throughout the maintenance of the parasites. Recombinant protein expressions were verified by Western blot and genomic integrations were confirmed with appropriate PCR primers.

P. knowlesi plasmid construction and transfection

Episomal overexpression plasmids were constructed using appropriate coding sequences of the full-length genes amplified from *P. knowlesi* genomic DNA or complementary DNA. The target fragments were then cloned into the AvrII/SmaI sites of the modified PKePL plasmid for HA and GFP tagging respectively.[14]

Synchronized schizonts stage parasites were transfected with 40μg of purified plasmids (Macherey-Nagel) using Lonza 4D-Nucleofector electroporator and P3 Primary Cell 4D-Nucleofector X Kit L as previously described.[92] Transfected cells were selected with 2.5nM WR99210 (Jacobus Pharmaceuticals) 24 h post transfection and maintained throughout the maintenance of the parasites. Recombinant protein expressions were verified by Western blot.

P. yoelii plasmid construction and transfection

SLI knock-in construct for HA-tagging was generated by replacing the GFP tag of the T2A-ePL-GFP plasmid[94] with a codon optimized 3xHA tag. Target homology arm fragments of the gene of interests (without stop codon) were amplified using KOD -Plus-*Neo* polymerase (Toyobo) from *P. yoelii* genomic DNA and inserted between the AvrII/ApaI sites of the modified PY-T2A-HA plasmid.

Episomal overexpression plasmids for GFP-tagging were constructed using appropriate full-length sequences of the target genes amplified from *P. yoelii* genomic DNA or constructed by chemical gene synthesis (Integrated DNA Technologies). All fragments were then cloned into the AvrII/SmaI sites of the PYePL-GFP plasmid.[14]

Synchronized schizonts stage parasites were transfected with 12μg of linearized plasmids using Lonza Amaxa Nucleofector II/2b electroporator and Basic Parasite Nucleofector Kit as previously described.[92] Transfected parasites were injected into naive mice by intravenous injection and selected with 10 mg/kg of pyrimethamine (MP Biomedicals) 24 h post transfection. Drug selection pressure was maintained by supplementing pyrimethamine (70 μg/ml) in drinking water throughout the maintenance of the parasites. Recombinant protein expressions were verified by Western blot and genomic integrations were confirmed with appropriate PCR primers.

Determination of the protein localization

Transfected parasites obtained after *in vitro* (*P. falciparum* and *P. knowlesi*) and *in vivo* (*P. yoelii*) selection were stained with 1 μg/ml DAPI (Thermo Scientific) for 5 min at 37°C for live-cell microscopy. The infected RBCs were observed with Nikon ECLIPSE Ti-E fluorescent microscope under 100× oil immersion objective. DAPI was detected using a Chromas 11000v3 filter set revealing the internal structures and membranes of live parasites whereas GFP was detected using a Chroma 49011 Filter Set.

Transfected parasites expressing HA-tagged proteins were fixed with either methanol or 4% formaldehyde (Thermo Scientific), permeabilized with 0.1% Triton X-100 (Bio-Rad) and blocked with 3% BSA (Sigma). The infected RBCs were then stained with rat anti-HA (Roche), rabbit anti-EXP2 (Abcam) and DAPI/Hoechst33342 (Thermo Scientific). Secondary antibody coupled to Alexa Fluor 594 (Jackson ImmunoResearch) and to DyLight 488 (Invitrogen) were used to reveal the localization of the HA-tagged protein and PVM protein respectively. The slides were examined by fluorescence microscopy.

Images were captured using an Andor Zyla sCMOS 4.2 or Zeisss AxioCam ICC1 camera and processed using ImageJ.[100] Throughout the study, pictures of 15–20 independent observations of trophozoite/early schizont stage parasites were collected. The localizations of the investigated proteins were determined by quantifying the intensities of pixels corresponding to the green fluorescence signal along the yellow line from A to B in the picture depicted using ImageJ and represented using a profile plot using a scale ranging from 0 to 80-pixel intensities (Figure S1).

Transcriptional profile analysis of spatial proteome

The *P. falciparum* proteomic data was re-analyzed using Proteome Discoverer version 2.1 software to derive the overall abundance of 2571 protein with emPAI[114] at trophozoite-schizont stage. The resulting data was then compared to the IDC abundance profiles of 4860 transcripts obtained from RNA-sequencing databases[51–54] and 1713 parasite internal proteins determined using iTRAQ quantitative proteomics[81] (Table S2 RNA vs. protein).

Post-transcriptional regulation on the spatial proteome

The abundance percentile rankings of each protein to its corresponding RNA were determined to identify incongruences by at least 20th percentile and variation of RNA levels across IDC was analyzed. The ''concordant'' proteins with similar RNA abundance levels (i.e., within 20th percentile ranking) and representing half of the proteome in each compartment were used as reference point.

Pathway analysis

Function and Pathway Enrichment Analysis were performed by analysing the various *P. falciparum* subproteomes using Metascape (www.metascape.org) against the whole proteome detected by proteomic approach.[96]

Phylogenomic and phylostratigraphic enrichment analysis

The proteomes of 19 species, representing key phylogenetic positions (Table S4 species list) were used to construct orthologous gene groups (orthogroups) with Orthofinder v2.4.0,[97] where Diamond v0.9.24.125 was used as a sequence aligner.[98] A species tree based on a consensus of the current status of eukaryotic phylogeny[115–119] was used for the phylostratigraphic analysis (Figure S3). Each orthogroups was assigned to a phylostratum (node) by identifying the oldest clade found in the orthogroup[120] using ETE toolkit v3.0.[99]

To test whether a specific phylostratum is enriched in proteins belonging to different subcellular fractions in *Plasmodium*, we used a subset of proteins described in the spatial proteome atlas (Table S4 enrichment analysis). The proteins were classified into seven subcellular fractions, two for intra-parasite (PARASITE, PPM), two for peripheral (PV, PVM), three for exported (VESICLE+Cleft's, GHOST, HCC). An average of 3238 proteins were analyzed for each *Plasmodium* species: *P. falciparum* (n = 3249), *P. knowlesi* (n = 3229), *P. yoelii* (n = 3361 proteins), *P. berghei* (n = 3156) and *P. chabaudi* (n = 3195), (Table S4 protein phylostratum). For each node in the species tree (Figure S3), we randomly selected (without replacement) the number of observed subcellular fraction-specific genes 1000 times (i.e., number of proteins present in NODE_3 (phylostratum SAR) that belong to the ghost subcellular fraction).

QUANTIFICATION AND STATISTICAL ANALYSIS

Samples from three (*P. falciparum* and *P. knowlesi*) and two (*P. yoelii, P. berghei and P. chabaudi*) independent biological replicates were used for this study. Individual peptides identified were quantified by comparing the intensity of reporter ions in the tandem mass (MS/MS) spectra. The relative abundance of a protein in a subcellular fraction is expressed as its subcellular abundance over its total abundance across all the samples. Empirical p values were obtained for phylostratum analysis by calculating whether the observed number of genes was larger than the number obtained from the sampling procedure. The p values were adjusted using a false discovery rate (FDR) correction[121] with a cut-off of 0.05.