The proteome of the malaria plastid organelle, a key anti-parasitic target

Michael J. Boucher[1,2], Sreejoyee Ghosh[2], Lichao Zhang[3], Avantika Lal[4,10], Se Won Jang[5,10], An Ju[6,11], Shuying Zhang[6,11], Xinzi Wang[6,11], Stuart A. Ralph[8], James Zou[7,9], Joshua E. Elias[3], and Ellen Yeh[1,2,4,9,12*]

Departments of [1]Microbiology and Immunology, [2]Biochemistry, [3]Chemical and Systems Biology, [4]Pathology, [5]Computer Science, [6]Bioengineering, and [7]Biomedical Data Science, Stanford University, Stanford, CA 94305, United States of America

[8]Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Parkville, Vic 3010, Australia

[9]Chan Zuckerberg Biohub, San Francisco, CA 94158, United States of America

[10]These authors contributed equally

[11]These authors contributed equally

[12]Lead Contact

*Correspondence: ellenyeh@stanford.edu

Keywords: malaria, apicoplast, BioID, proteome, neural network

Summary

The apicoplast is an essential plastid organelle in malaria parasites (*Plasmodium* spp.) and a validated anti-parasitic target. A major hurdle to uncovering cryptic apicoplast pathways required for malaria pathogenesis is the lack of an organellar proteome. Here we combine proximity biotinylation-based proteomics (BioID) and a new machine learning algorithm to generate the first high-confidence apicoplast proteome consisting of 346 proteins. Critically, the accuracy of this proteome significantly outperforms previous prediction-based methods. Half of identified proteins have unknown function, and 77% are predicted to be important for normal blood-stage growth. We validate the apicoplast localization of a subset of novel proteins and show that an ATP-binding cassette protein ABCF1 is essential for blood-stage survival and plays a previously unknown role in apicoplast biogenesis. These findings indicate critical organellar functions for newly-discovered apicoplast proteins. The apicoplast proteome will be an important resource for elucidating unique pathways and prioritizing antimalarial drug targets.

Introduction

Identification of new antimalarial drug targets is urgently needed to address emerging resistance to all currently available therapies. However, nearly half of the *Plasmodium falciparum* genome encodes conserved proteins of unknown function (Aurrecoechea et al., 2017), obscuring critical pathways required for malaria pathogenesis. The apicoplast is an essential, non-photosynthetic plastid found in *Plasmodium* spp. and related apicomplexan pathogens (McFadden et al., 1996; Kohler et al., 1997). This unusual organelle is an enriched source of both novel cellular pathways and parasite-specific drug targets (van Dooren and Striepen, 2013). It was acquired by secondary (i.e., eukaryote-eukaryote) endosymbiosis and has

evolutionarily diverged from the primary endosymbiotic organelles found in model organisms. While some aspects of apicoplast biology are shared with bacteria, mitochondria, and primary chloroplasts, many are unique to the secondary plastid in this parasite lineage. For example, novel translocons import apicoplast proteins through specialized membranes derived from secondary endosymbiosis (Agrawal et al., 2009; Kalanon et al., 2009; Spork et al., 2009; Agrawal et al., 2013), while the parasite's pared-down metabolism necessitates export of key metabolites from the apicoplast using as-yet unidentified small molecule transporters (Ralph et al., 2004; Sheiner et al., 2013).

These novel cellular pathways, which are also distinct from human host cells, can be exploited for antimalarial drug discovery. Indeed, antimalarials that target apicoplast pathways are currently in use as prophylactics or partner drugs (doxycycline, clindamycin) or have been tested in clinical trials (fosmidomycin) (Jomaa et al., 1999; Dahl et al., 2006; Dahl and Rosenthal, 2007; Goodman et al., 2007; Stanway et al., 2009). However, known apicoplast drug targets have been limited to the handful of pathways identified by homology to plastid-localized pathways in model organisms. Meanwhile the number of druggable apicoplast targets, including those in unique secondary plastid pathways, is likely more extensive (Amberg-Johnson et al., 2017).

A major hurdle to identifying novel, parasite-specific pathways and prioritizing new apicoplast targets is the lack of a well-defined organellar proteome. So far, the apicoplast has not been isolated in sufficient yield or purity for traditional organellar proteomics. Instead, large-scale, unbiased identification of apicoplast proteins has relied on bioinformatic prediction of apicoplast targeting sequences (Zuegge et al., 2001; Foth et al., 2003; Cilingir et al., 2012). These prediction algorithms identify hundreds of putative apicoplast proteins but contain

numerous false positives. Validation of these low-confidence candidate apicoplast proteins is slow due to the genetic intractability of *P. falciparum* parasites. Unbiased identification of apicoplast proteins in an accurate and high-throughput manner would significantly enhance our ability to study novel apicoplast pathways and would suggest new antimalarial drug targets.

BioID and other cellular proximity labeling methods are attractive techniques that enable accurate identification of organellar proteins (Roux et al., 2012). In BioID, a promiscuous biotin ligase, BirA*, is fused to a bait protein and catalyzes biotinylation of neighbor proteins in intact cells. Proximity labeling methods have been used to spatially map the proteomes of subcellular compartments in diverse organisms, including the *Plasmodium* parasitophorous vacuole and parasitophorous vacuole membrane (Khosh-Naucke et al., 2018; Schnider et al., 2018), *Toxoplasma gondii* micronemes and dense granules (Chen et al., 2015; Nadipuram et al., 2016; Chen et al., 2017), *Trypanosoma brucei* bilobe and basal bodies (Morriswood et al., 2013; Dang et al., 2017), and human mitochondria (Rhee et al., 2013; Hung et al., 2014), nuclear pore complexes (Kim et al., 2014), centrosome-cilium interface (Gupta et al., 2015), synaptic clefts (Loh et al., 2016), and mitochondrial-ER contact sites (Hung et al., 2017). Here, we used BioID to perform large-scale identification of the *P. falciparum* apicoplast proteome during asexual blood-stage growth. These proteomic identifications helped us develop an improved neural network prediction algorithm, PlastNN. We now report a high-confidence, compiled apicoplast proteome of 346 proteins rich in novel and essential functions.

Results

The promiscuous biotin ligase BirA* is functional in the *P. falciparum* apicoplast and endoplasmic reticulum

To target the promiscuous biotin ligase BirA* to the apicoplast, the *N*-terminus of a GFP-BirA* fusion protein was modified with the apicoplast-targeting leader sequence from acyl carrier protein (ACP) (Figure 1A). Since apicoplast proteins transit the parasite endoplasmic reticulum (ER) en route to the apicoplast (Waller et al., 2000), we also generated a negative control in which GFP-BirA* was targeted to the ER via an *N*-terminal signal peptide and a *C*-terminal ER-retention motif (Figure 1A). Each of these constructs was integrated into an ectopic locus in Dd2^{attB} parasites (Nkrumah et al., 2006) to generate BioID-Ap and BioID-ER parasites (Figure S1). Live imaging of these parasites confirmed GFP-BirA* localization to a branched structure characteristic of the apicoplast or a perinuclear region characteristic of the ER, respectively (Figure 1B).

To test the functionality of the GFP-BirA* fusions in the apicoplast and ER, we labeled either untransfected Dd2^{attB}, BioID-Ap, or BioID-ER parasites with DMSO or 50 µM biotin and assessed biotinylation by western blotting and fixed-cell fluorescent imaging. As has been reported (Khosh-Naucke et al., 2018), significant labeling of GFP-BirA*-expressing parasites above background was achieved even in the absence of biotin supplementation, suggesting that the 0.8 µM biotin in RPMI growth medium is sufficient for labeling (Figure 1C). Addition of 50 µM biotin further increased protein biotinylation. Fluorescence imaging of biotinylated proteins revealed staining consistent with apicoplast morphology in BioID-Ap parasites and the ER and other endomembrane structures in BioID-ER parasites (Figure 1D). These results confirm that GFP-BirA* fusions are active in the *P. falciparum* apicoplast and ER and can be used for compartment-specific biotinylation of proteins.

Proximity-dependent labeling (BioID) generates an improved apicoplast proteome dataset

For large-scale identification of apicoplast proteins, biotinylated proteins from late-stage BioID-Ap and BioID-ER parasites were purified using streptavidin-conjugated beads and identified by mass spectrometry. A total of 728 unique *P. falciparum* proteins were detected in the apicoplast and/or ER based on presence in at least 2 of 4 biological replicates and at least 2 unique spectral matches in any single mass spectrometry run (Figure 2A and Table S1). The abundance of each protein in apicoplast and ER samples was calculated by summing the total MS1 area of all matched peptides and normalizing to the total MS1 area of all detected *P. falciparum* peptides within each mass spectrometry run.

To assess the ability of our dataset to distinguish between true positives and negatives, we then generated control lists of 96 known apicoplast and 451 signal peptide-containing non-apicoplast proteins based on published localizations and pathways (Table S2). Consistent with an enrichment of apicoplast proteins in BioID-Ap samples, we observed a clear separation of known apicoplast and non-apicoplast proteins based on apicoplast:ER abundance ratio (Figure 2A). Based on the apicoplast:ER abundance ratio, we considered the 187 proteins that were ≥5-fold enriched in apicoplast samples (Figure 2A, dotted line) to be the BioID apicoplast proteome (Table S1). This dataset included 50 of the 96 positive control proteins for a sensitivity of 52%. Further, manual inspection of the proteins on the ≥5-fold enriched apicoplast list identified 54 true positives and 5 likely false positives (Table S1) for a positive predictive value (PPV; the estimated fraction of proteins on the list that are true positives) of 92%.

To benchmark our dataset against the current standard for large-scale identification of apicoplast proteins, we compared the sensitivity and PPV of our apicoplast BioID proteome to those from three published bioinformatic algorithms: PATS (Zuegge et al., 2001), PlasmoAP (Foth et al., 2003), and ApicoAP (Cilingir et al., 2012) (Table S3). At 52% sensitivity, apicoplast

BioID identified fewer known apicoplast proteins than PATS or PlasmoAP, which had sensitivities of 89% and 84%, respectively, but outperformed the 40% sensitivity of ApicoAP (Figure 2B). However, we expected that the advantages of apicoplast BioID would be the ability to detect proteins without classical targeting presequences and its improved discrimination between true and false positives (Figure 2A). Indeed, bioinformatic algorithms had poor PPVs ranging from 19%-36% compared to the 92% PPV of BioID (Figure 2C). Even a dataset consisting only of proteins predicted by all three algorithms achieved a PPV of just 25%. Consistent with these low PPVs, many proteins predicted by the bioinformatic algorithms are not enriched in BioID-Ap samples, suggesting that many of these proteins are likely to be false positives (Figure S2). Altogether, identification of apicoplast proteins using BioID provided a dramatic improvement in prediction performance over bioinformatic algorithms.

Apicoplast BioID identifies proteins of diverse functions in multiple subcompartments

To determine whether lumenally targeted GFP-BirA* exhibited any labeling preferences, we assessed proteins identified based on the presence of transmembrane domains, their sub-organellar localization, and their functions. First, we determined the proportion of the 187 proteins identified by apicoplast BioID that are membrane proteins. To ensure that proteins were not classified as membrane proteins solely due to misclassification of a signal peptide as a transmembrane domain, we considered a protein to be in a membrane only if it contained at least one predicted transmembrane domain more than 80 amino acids from the protein's *N*-terminus (as determined by annotation in PlasmoDB). These criteria suggested that 11% of identified proteins (20/187) were likely membrane proteins (Figure 3A), indicating that lumenal GFP-BirA* can label apicoplast membrane proteins.

Second, apicoplast proteins may localize to one or multiple sub-compartments defined by the four apicoplast membranes. It was unclear whether BirA* targeted to the lumen would label proteins in non-lumenal compartments. Based on literature descriptions, we classified the 96 known apicoplast proteins on our positive control list as either lumenal (present in lumenal space or on the innermost apicoplast membrane) or non-lumenal (all other sub-compartments) and determined the proportion that were identified in our dataset. Apicoplast BioID identified 56% (45/81) of the classified lumenal proteins and 33% (5/15) of the non-lumenal proteins (Figure 3B), suggesting that the GFP-BirA* bait used can label both lumenal and non-lumenal proteins but may have a preference for lumenal proteins (though this difference did not reach statistical significance).

Finally, we characterized the functions of proteins identified by apicoplast BioID. We grouped positive control apicoplast proteins into functional categories and assessed the proportion of proteins identified from each functional group (Figure 3C). BioID identified a substantial proportion (67-100%) of proteins in four apicoplast pathways that are essential in blood stage and localize to the apicoplast lumen, specifically DNA replication, protein translation, isoprenoid biosynthesis, and iron-sulfur cluster biosynthesis. Conversely, BioID identified few proteins involved in heme or fatty acid biosynthesis (0% and 17%, respectively), which are lumenal pathways that are non-essential in the blood-stage and which are likely to be more abundant in other life cycle stages (Yu et al., 2008; Vaughan et al., 2009; Pei et al., 2010; Nagaraj et al., 2013; Ke et al., 2014). We achieved moderate coverage of proteins involved in protein quality control (44%) and redox regulation (38%). Consistent with the reduced labeling of non-lumenal apicoplast proteins, only a small subset (29%) of proteins involved in import of nuclear-encoded apicoplast proteins were identified. Overall, apicoplast BioID identified soluble

and membrane proteins of diverse functions in multiple apicoplast compartments with higher coverage for lumenal proteins required during blood-stage infection.

The PlastNN algorithm expands the predicted apicoplast proteome with high accuracy

Apicoplast BioID provided the first experimental profile of the blood-stage apicoplast proteome but is potentially limited in sensitivity due to 1) difficulty in detecting low abundance peptides in complex mixtures; 2) inability of the promiscuous biotin ligase to access target proteins that are buried in membranes or protein complexes; or 3) stage-specific protein expression. Currently-available bioinformatic predictions of apicoplast proteins circumvent these limitations, albeit at the expense of a low PPV (Figure 2C). We reasoned that increasing the number of high-confidence apicoplast proteins used to train algorithms could improve the accuracy of a prediction algorithm while maintaining high sensitivity. In addition, inclusion of exported proteins that traffic through the ER, which are common false positives in previous prediction algorithms, would also improve our negative training set.

We used our list of previously known apicoplast proteins (Table S2) as well as newly-identified apicoplast proteins from BioID (Table S1) to construct a positive training set of 205 apicoplast proteins (Table S4). As a negative training set, we used our previous list of 451 signal peptide-containing non-apicoplast proteins (Table S2). For each of the 656 proteins in the training set, we calculated the frequencies of all 20 canonical amino acids in a 50 amino acid region immediately following the predicted signal peptide cleavage site. In addition, given that apicoplast proteins have a characteristic transcriptional profile in blood-stage parasites (Bozdech et al., 2003) and that analysis of transcriptional profile has previously enabled identification of apicoplast proteins in the related apicomplexan *Toxoplasma gondii* (Sheiner et al., 2011), we

obtained transcript levels at 8 time points during intraerythrocytic development from previous RNA-Seq data (Bartfai et al., 2010). Altogether, each protein was represented by a vector of dimension 28 (20 amino acid frequencies plus 8 transcript levels). These 28-dimensional vectors were used as inputs to train a neural network with 3 hidden layers (Figure 4A and Table S5). Six-fold cross-validation was used for training, wherein the training set was divided into 6 equal parts (folds) to train 6 separate models. Each time, 5 folds were used to train the model and 1 fold to measure the performance of the trained model.

We named this model PlastNN (ApicoPLAST Neural Network). PlastNN recognized apicoplast proteins with a cross-validation accuracy of 96 ± 3% (mean ± s.d. across 6 models), along with sensitivity of 95 ± 5%, and PPV of 94 ± 4% (Figure 4B). This performance was higher than logistic regression on the same dataset (average accuracy = 91%). Combining the transcriptome features and the amino acid frequencies improves performance: the same neural network architecture with amino acid frequencies alone as input resulted in a lower average accuracy of 91%, while using transcriptome data alone resulted in an average accuracy of 90% (Table S6). Comparison of the performance of PlastNN to existing prediction algorithms indicates that PlastNN distinguishes apicoplast and non-apicoplast proteins with higher accuracy than any previous prediction method (Figure 4C). To identify new apicoplast proteins, PlastNN was used to predict the apicoplast status of 450 predicted signal peptide-containing proteins that were not in our positive or negative training sets. Since PlastNN is composed of 6 models, we designated proteins as "apicoplast" if plastid localization was predicted by ≥4 of the 6 models. PlastNN predicts 118 out of the 450 proteins to be targeted to the apicoplast (Table S7). Combining these results with those from apicoplast BioID (Table S1) and with experimental

localization of proteins from the literature (Table S2) yielded a compiled proteome of 346 putative nuclear-encoded apicoplast proteins (Table S8).

The apicoplast proteome contains a multitude of novel and essential proteins

To determine whether candidate apicoplast proteins from this study have the potential to reveal unexplored parasite biology or are candidate antimalarial drug targets, we assessed the novelty and essentiality of the identified proteins. We found that substantial fractions of the BioID and PlastNN proteomes (49% and 71%, respectively) and 50% of the compiled apicoplast proteome represented proteins that could not be assigned to an established apicoplast pathway and therefore might be involved in novel organellar processes (Figure 5A). Furthermore, we identified orthologs of identified genes in the 150 genomes present in the OrthoMCL database (Chen et al., 2006): 39% of the compiled apicoplast proteome were unique to apicomplexan parasites, with 58% of these proteins found only in *Plasmodium* spp. (Figure 5B). This analysis indicates that many of the proteins identified are significantly divergent from proteins in their metazoan hosts.

Consistent with the critical role of the apicoplast in parasite biology, a recent genome-scale functional analysis of genes in the rodent malaria parasite *P. berghei* showed that numerous apicoplast proteins are essential for blood-stage survival (Bushell et al., 2017). Using this dataset, we found that 77% of those proteins in the compiled apicoplast proteome that had *P. berghei* homologs analyzed by PlasmoGEM were important for normal blood-stage parasite growth (Figure 5C). Notably, of 49 proteins that were annotated explicitly with “unknown function” in their gene description and for which essentiality data are available, 38 are important for normal parasite growth, indicating that the high rate of essentiality for apicoplast proteins is

248 true of both previously known and newly discovered proteins. Overall, these data suggest that we
249 have identified dozens of novel proteins that are likely critical for apicoplast biology.
250
251 **Localization of candidate proteins confirms accuracy of protein identification**
252 To confirm the utility of our approaches, we experimentally determined the localization
253 of several candidate apicoplast proteins. A rhomboid protease homolog ROM7 and 3 conserved
254 *Plasmodium* proteins of unknown function (PF3D7_0521400, PF3D7_1472800, and
255 PF3D7_0721100) were each overexpressed as a *C*-terminal GFP fusions and tested in apicoplast
256 localization assays. First, we detected the apicoplast-dependent cleavage of each candidate as a
257 marker of its import. Most nuclear-encoded apicoplast proteins are proteolytically processed to
258 remove *N*-terminal targeting sequences following successful import into the apicoplast (Waller et
259 al., 1998; van Dooren et al., 2002). This processing is abolished in parasites rendered
260 "apicoplast-minus" by treatment with an inhibitor (actinonin) to cause apicoplast loss (Yeh and
261 DeRisi, 2011; Amberg-Johnson et al., 2017). Comparison of protein molecular weight in
262 apicoplast-intact and -minus parasites showed that ROM7, PF3D7_0521400, and
263 PF3D7_1472800 (but not PF3D7_0721100) were cleaved in an apicoplast-dependent manner
264 (Figure 6A).
265 Second, we localized the candidate-GFP fusions by live fluorescence microscopy and
266 assessed their mislocalization in apicoplast-minus parasites. Consistent with apicoplast
267 localization, ROM7-GFP, PF3D7_0521400-GFP, and PF3D7_1472800-GFP localized to
268 branched structures characteristic of the apicoplast (Figure 6B). In apicoplast-minus parasites,
269 these proteins mislocalized to diffuse puncta (Figure 6B), as previously observed for apicoplast
270 proteins (Yeh and DeRisi, 2011). Interestingly, while in untreated parasites PF3D7_0721100-
271 GFP localized to a few large bright puncta not previously described for any apicoplast protein,

this protein also relocalized to the typical numerous diffuse puncta seen for genuine apicoplast proteins in apicoplast-minus parasites (Figure 6B). Taken together, these data validate the apicoplast localization of ROM7, PF3D7_0521400, and PF3D7_1472800. Though targeting peptide cleavage and the characteristic branched structure were not detected for PF3D7_0721100, the mislocalization of PF3D7_0721100-GFP to puncta characteristic of apicoplast-minus parasites indicates that this protein may also be a true apicoplast protein.

We next focused on four ATP-binding cassette (ABC) proteins in our dataset: ABCB3, ABCB4, ABCB7, and ABCF1. ABCB-family proteins are small molecule membrane transporters, while ABCF-family proteins do not contain transmembrane domains and are typically involved in translation regulation (Kerr, 2004; Dean and Annilo, 2005). While this manuscript was in preparation, the apicoplast localization of ABCB3 and ABCB4 were confirmed in *P. falciparum* and *P. berghei*, respectively (Sayers et al., 2018). To assess localization and function of ABCB7 and ABCF1, we modified their endogenous loci to contain a *C*-terminal triple HA tag and tandem copies of a tetracycline repressor (TetR)-binding RNA aptamer in the 3′ UTR of either gene (Figure S3) (Goldfless et al., 2014; Ganesan et al., 2016). We performed immunofluorescence analysis (IFA) to determine whether ABCB7-HA and ABCF1-HA colocalized with the apicoplast marker ACP. ABCF1-HA exhibited clear co-localization with ACP, confirming its apicoplast localization (Figure 6C). ABCB7-HA localized to elongated structures that may be indicative of an intracellular organelle but rarely co-localized with ACP, indicating a primarily non-apicoplast localization (Figure 6C). Overall, of 8 candidates of unknown localization at the start of this study, we identified 6 confirmed apicoplast proteins, 1 likely apicoplast protein, and 1 potential false positive.

A novel apicoplast protein ABCF1 is essential and required for organelle biogenesis

We determined the essentiality and knockdown phenotype of a newly identified apicoplast protein, ABCF1, taking advantage of the TetR-binding aptamers inserted into its 3′ UTR as described above. In the presence of anhydrotetracycline (ATc), binding of the aptamer by a TetR-DOZI repressor is inhibited and ABCF1 is expressed. Upon removal of ATc, repressor binding blocks gene expression (Goldfless et al., 2014; Ganesan et al., 2016). Knockdown of ABCF1 caused robust parasite growth inhibition (Figure 7A-B). Significantly, growth inhibition of ABCF1-deficient parasites was reversed in the presence of isopentenyl pyrophosphate (IPP) (Figure 7B), which bypasses the need for a functional apicoplast (Yeh and DeRisi, 2011), indicating that ABCF1 has an essential apicoplast function. Essential apicoplast functions can be placed into two broad categories: those involved in organelle biogenesis, and those involved solely in IPP production. Disruption of proteins required for organelle biogenesis causes apicoplast loss, while disruption of proteins involved in IPP production does not (Yeh and DeRisi, 2011; Wu et al., 2015; Amberg-Johnson et al., 2017). We determined whether knockdown of ABCF1 caused apicoplast loss by assessing 1) absence of the apicoplast genome, 2) loss of transit peptide processing of nuclear-encoded apicoplast proteins, and 3) relocalization of apicoplast proteins to puncta. Indeed, the apicoplast:nuclear genome ratio drastically decreased in ABCF1 knockdown parasites beginning 1 cycle after knockdown (Figure 7C), and western blot showed that the apicoplast protein ClpP was not processed in ABCF1 knockdown parasites (Figure 7D). Furthermore, IFA of the apicoplast marker ACP confirmed redistribution from an intact plastid to diffuse cytosolic puncta (Figure 7E). In contrast to ABCF1, a similar knockdown of ABCB7 caused no observable growth defect after four growth cycles despite significant reduction in protein levels (Figure S4). Together, these results show that ABCF1 is a

318 novel and essential apicoplast protein with a previously unknown function in organelle
319 biogenesis.
320
321 **Discussion**
322 Since the discovery of the apicoplast, identification of its proteome has been a pressing
323 priority. We report the first large-scale proteomic analysis of the apicoplast in blood-stage
324 malaria parasites, which identified 187 candidate proteins with 52% sensitivity and 92% PPV. A
325 number of groups have also profiled parasite-specific membrane compartments using proximity
326 biotinylation but observed contamination with proteins in or trafficking through the ER,
327 preventing accurate identification of these proteomes without substantial manual curation and
328 validation (Chen et al., 2015; Nadipuram et al., 2016; Chen et al., 2017; Khosh-Naucke et al.,
329 2018; Schnider et al., 2018). This background labeling is expected since proteins traffic through
330 the ER en route to several parasite-specific compartments, including the parasitophorous
331 vacuole, host cytoplasm, food vacuole, and invasion organelles. The high specificity of our
332 apicoplast BioID proteome depended on 1) the use of a control cell line expressing ER-localized
333 GFP-BirA* to detect enrichment of apicoplast proteins from background ER labeling and 2)
334 strong positive and negative controls to set an accurate threshold. We suspect a similar strategy
335 to detect nonspecific ER background may also improve the specificity of proteomic datasets for
336 other parasite-specific, endomembrane-derived compartments.
337 Leveraging our successful proteomic analysis, we used these empirical data as an updated
338 training set to also improve computational predictions of apicoplast proteins. PlastNN identified
339 an additional 118 proteins with 95% sensitivity and 94% PPV. Although two previous prediction
340 algorithms, PATS and ApicoAP, also applied machine learning to the problem of transit peptide

prediction, we reasoned that their low accuracy arose from the small training sets used (ApicoAP) and the use of cytosolic as well as endomembrane proteins in the negative training set (PATS). By using an expanded positive training set based on proteomic data and limiting our training sets to only signal peptide-containing proteins, we developed an algorithm with higher sensitivity than BioID and higher accuracy than previous apicoplast protein prediction models. Moreover, PlastNN suggests testable hypotheses regarding the contribution of sequence-based and temporal regulation to protein trafficking in the ER.

Overall, we have compiled a high-confidence apicoplast proteome of 346 proteins that are rich in novel and essential functions (Figure 5A and 5C). This proteome likely represents a majority of soluble apicoplast proteins, since 1) our bait for proximity biotinylation targeted to the lumen and 2) most soluble proteins use canonical targeting sequences that can be predicted. Further improvements to the apicoplast proteome will focus on expanding the coverage of membrane proteins, which more often traffic via distinctive routes (Mullin et al., 2006; Parsons et al., 2007). Performing proximity biotinylation with additional bait proteins may identify such atypical apicoplast proteins. In the current study, our bait was an inert fluorescent protein targeted to the apicoplast lumen to minimize potential toxicity of the construct. The success of this apicoplast GFP bait gives us confidence to attempt more challenging baits, including proteins localized to sub-organellar membrane compartments or components of the protein import machinery. Performing apicoplast BioID in liver and mosquito stages may also define apicoplast functions in these stages.

The apicoplast proteome will be a valuable resource for uncovering cryptic pathways required for malaria pathogenesis and prioritizing new antimalarial drug targets. Already several candidates of biological interest based on their biochemical function annotations were validated.

We demonstrated an unexpected role for the ATP-binding cassette protein *Pf*ABCF1 in apicoplast biogenesis. An *E. coli* homolog, EttA, regulates translation initiation in response to cellular ATP levels (Boel et al., 2014; Chen et al., 2014). Mammalian and yeast ABCF1 homologs also interact with ribosomes and regulate translation (Vazquez de Aldana et al., 1995; Marton et al., 1997; Tyzack et al., 2000; Paytubi et al., 2009). By analogy, *Pf*ABCF1 may regulate the prokaryotic translation machinery in the apicoplast, although the mechanistic basis for the severe defect in parasite replication upon loss of *Pf*ABCF1 is unclear. We also validated *Pf*ROM7 as an apicoplast-localized rhomboid protease. This protein may have a role in apicoplast protein import, as a rhomboid protease was recently identified as a component of symbiont-derived ERAD-like machinery (SELMA) that transports proteins across a novel secondary plastid membrane in diatoms (Lau et al., 2016). Neither *Pf*ABCF1 nor *Pf*ROM7 had known roles in the apicoplast prior to their identification in this study, underscoring the utility of unbiased approaches to identify new organellar proteins.

A recent study aimed at identifying apicoplast membrane transporters highlights the difficulty in identifying novel apicoplast functions in the absence of a high-confidence proteome (Sayers et al., 2018). Taking advantage of the tractable genetics in murine *Plasmodium* species, Sayers et al. screened 27 candidates in *P. berghei* for essentiality and apicoplast localization. Following >50 transfections, 3 essential and 4 non-essential apicoplast membrane proteins were identified. One newly-identified essential apicoplast membrane protein was then validated to be required for apicoplast biogenesis in *P. falciparum*. In contrast, even though our study was not optimized to identify membrane proteins, the combination of BioID and PlastNN identified 2 known apicoplast transporters, 4 of the new apicoplast membrane protein homologs, and 56 additional proteins predicted to contain at least one transmembrane domain. A focused screen of

higher quality candidates in *P. falciparum* is likely to be more rapid and yield the most relevant biology. Our high-confidence apicoplast proteome will streamline these labor-intensive screens, focusing on strong candidates for downstream biological function elucidation. As methods for analyzing gene function in *P. falciparum* parasites continue to improve, this resource will become increasingly valuable for characterizing unknown organellar pathways.

Materials and Methods

Parasite growth

Plasmodium falciparum Dd2^{attB} (Nkrumah et al., 2006) (MRA-843) were obtained from MR4. NF54$^{Cas9+T7\ Polymerase}$ parasites (Sidik et al., 2016) were a gift from Jacquin Niles. Parasites were grown in human erythrocytes (2% hematocrit) obtained from the Stanford Blood Center in RPMI 1640 media (Gibco) supplemented with 0.25% Albumax II (Gibco), 2 g/L sodium bicarbonate, 0.1 mM hypoxanthine (Sigma), 25 mM HEPES, pH 7.4 (Sigma), and 50 µg/L gentamicin (Gold Biotechnology) at 37°C, 5% O_2, and 5% CO_2.

Vector construction

Oligonucleotides were purchased from the Stanford Protein and Nucleic Acid facility or IDT. gBlocks were ordered from IDT. Molecular cloning was performed using In-Fusion cloning (Clontech) or Gibson Assembly (NEB). Primer and gBlock sequences are available in Table S9.

To generate the plasmid pRL2-ACP$_L$-GFP for targeting transgenes to the apicoplast, the first 55 amino acids from ACP were PCR amplified with primers MB015 and MB016 and were inserted in front of the GFP in the pRL2 backbone (Balabaskaran Nina et al., 2011) via the AvrII/BsiWI sites. To generate pRL2-ACP$_L$-GFP-BirA* for targeting a GFP-BirA* fusion to the

410 apicoplast, GFP was amplified from pLN-ENR-GFP using primers MB087 and MB088 and
411 BirA* was amplified from pcDNA3.1 mycBioID (Addgene 35700) (Roux et al., 2012) using
412 primers MB089 and MB090. These inserts were simultaneously cloned into BsiWI/AflII-
413 digested pRL2-ACP$_L$-GFP to generate pRL2-ACP$_L$-GFP-BirA*. To generate pRL2-SS-GFP-
414 BirA*-SDEL for targeting GFP-BirA* to the ER, SS-GFP-BirA*-SDEL was PCR amplified
415 from pRL2-ACP$_L$-GFP-BirA* using primers MB093 and MB094 and was cloned into
416 AvrII/AflII-digested pRL2-ACP$_L$-GFP. For GFP-tagging to confirm localization of proteins
417 identified by apicoplast BioID, full-length genes were amplified from parasite cDNA with
418 primers as described in Table S9 and were cloned into the AvrII/BsiWI sites of pRL2-ACP$_L$-
419 GFP.

420 For CRISPR-Cas9-based editing of endogenous ABCB7 and ABCF1 loci, sgRNAs were
421 designed using the eukaryotic CRISPR guide RNA/DNA design tool (http://grna.ctegd.uga.edu/).
422 To generate a linear plasmid for CRISPR-Cas9-based editing, left homology regions were
423 amplified with primers MB256 and MB257 (ABCB7) or MB260 and MB261 (ABCF1) and right
424 homology regions were amplified with MB258 and MB259 (ABCB7) or MB262 and MB263
425 (ABCF1). For each gene, a gBlock containing the recoded coding sequence *C*-terminal of the
426 CRISPR cut site and a triple HA tag was synthesized with appropriate overhangs for Gibson
427 Assembly. This fragment was simultaneously cloned into the FseI/ApaI sites of the linear
428 plasmid pSN054-V5. Next, the appropriate right homology region and a gBlock containing the
429 sgRNA expression cassette were simultaneously cloned into the AscI/I-SceI sites of the resultant
430 vector to generate the plasmids pSN054-ABCB7-TetR-DOZI and pSN054-ABCF1-TetR-DOZI.
431

432 **Parasite transfection**

Transfections were carried out using variations on the spontaneous uptake method (Deitsch et al., 2001; Wagner et al., 2014). In the first variation, 100 µg of each plasmid was ethanol precipitated and resuspended in 30 µL sterile TE buffer and was added to 150 µL packed RBCs resuspended to a final volume of 400 µL in cytomix. The mixture was transferred to a 0.2 cm electroporation cuvette (Bio-Rad) and was electroporated at 310 V, 950 µF, infinity resistance in a Gene Pulser Xcell electroporation system (Bio-Rad) before allowing parasites to invade. Drug selection was initiated 3 days after transfection. Alternatively, 50 µg of each plasmid was ethanol precipitated and resuspended in 0.2 cm electroporation cuvettes in 100 µL TE buffer, 100 µL RPMI containing 10 mM HEPES-NaOH, pH 7.4, and 200 µL packed uninfected RBCs. RBCs were pulsed with 8 square wave pulses of 365 V x 1 ms separated by 0.1 s. RBCs were allowed to reseal for 1 hour in a 37°C water bath before allowing parasites to invade. Drug selection was initiated 4 days after transfection. All transfectants were selected with 2.5 µg/mL Blasticidin S (Research Products International). Additionally, BioID-ER parasites were selected with 125 µg/mL G418 sulfate (Corning) and ABCB7 and ABCF1 TetR-DOZI parasites were grown in the presence of 500 nM ATc. Transfections for generating BioID constructs (Figure 1) and expression of GFP-tagged candidates (Figure 6) were performed in the Dd2^{attB} background. Transfections for CRISPR editing were performed with the NF54$^{Cas9+T7\ Polymerase}$ background and clonal parasite lines were obtained by limiting dilution.

Correct modification of transfectant genomes was confirmed by PCR. Briefly, 200 µL of 2% hematocrit culture was pelleted and resuspended in water, and 2 µL of the resulting lysate was used as template for PCR with Phusion polymerase (NEB). PCR targets and their corresponding primer pairs are as follows: integrated *attL* site, p1 + p2; integrated *attR* site, MW001 + MW003; unintegrated *attB* site, MW004 + MW003; ABCB7 unintegrated left

456 homology region (LHR), MB269 + MB270; ABCB7 integrated LHR, MB269 + MB255;
457 ABCB7 unintegrated right homology region (RHR), MB281 + MB278; ABCB7 integrated RHR,
458 MB276 + MB 278; ABCF1 unintegrated LHR, MB271 + MB272; ABCF1 integrated LHR, MB
459 271 + MB255; ABCF1 unintegrated RHR, MB282 + MB283; ABCF1 integrated RHR, MB276
460 + MB283.
461
462 **Biotin labeling**
463 To label parasites for analysis by streptavidin blot, fixed imaging, or mass spectrometry, cultures
464 of majority ring stage parasites were treated with 50 µM biotin or with a DMSO vehicle only
465 control. Cultures were harvested for analysis 16 hours later as majority trophozoites and
466 schizonts.
467
468 **Actinonin treatment and IPP rescue**
469 To generate apicoplast-minus parasites, ring-stage cultures were treated with 10 µM actinonin
470 (Sigma) and 200 µM IPP (Isoprenoids, LLC) and cultured for 3 days before analysis.
471
472 **Western blotting**
473 Parasites were separated from RBCs by lysis in 0.1% saponin and were washed in PBS. Parasite
474 pellets were resuspended in PBS containing 1X NuPAGE LDS sample buffer with 50 mM DTT
475 and were boiled at 95°C for 10 minutes before separation on NuPAGE or Bolt Bis-Tris gels and
476 transfer to nitrocellulose. Membranes were blocked in 0.1% Hammarsten casein (Affymetrix) in
477 0.2X PBS with 0.01% sodium azide. Antibody incubations were performed in a 1:1 mixture of
478 blocking buffer and TBST (Tris-buffered saline with Tween-20; 10 mM Tris, pH 8.0, 150 mM

NaCl, 0.25 mM EDTA, 0.05% Tween 20). Blots were incubated with primary antibody for either 1 hour at room temperature or at 4°C overnight at the following dilutions: 1:20,000 mouse-α-GFP JL-8 (Clontech 632381); 1:20,000 rabbit-α-*Plasmodium* aldolase (Abcam ab207494); 1:1000 rat-α-HA 3F10 (Sigma 11867423001); 1:4000 rabbit-α-*Pf*ClpP (El Bakkouri et al., 2010). Blots were washed once in TBST and were incubated for 1 hour at room temperature in a 1:10,000 dilution of the appropriate secondary antibody: IRDye 800CW donkey-α-rabbit; IRDye 680LT goat-α-mouse; IRDye 680LT goat-α-rat (LI-COR Biosciences). For detection of biotinylated proteins, blots were incubated with 1:1000 IRDye 680RD streptavidin for one hour at room temperature. Blots were washed three times in TBST and once in PBS before imaging on a LI-COR Odyssey imager.

Microscopy

For live imaging, parasites were settled onto glass-bottom microwell dishes (MatTek P35G-1.5-14-C) or Lab-Tek II chambered coverglass (ThermoFisher 155409) in PBS containing 0.4% glucose and 2 µg/mL Hoechst 33342 stain (ThermoFisher H3570).

For fixed imaging of biotinylated proteins in cells, biotin-labeled parasites were processed as in Tonkin et al. (Tonkin et al., 2004) with modifications. Briefly, parasites were washed in PBS and were fixed in 4% paraformaldehyde (Electron Microscopy Science 15710) and 0.015% glutaraldehyde (Electron Microscopy Sciences 16019) in PBS for 30 minutes. Cells were washed once in PBS, resuspended in PBS, and allowed to settle onto poly-L-lysine-coated coverslips (Corning) for 60 minutes. Coverslips were then washed once with PBS, permeabilized in 0.1% Triton X-100 in PBS for 10 minutes, and washed twice more in PBS. Cells were treated with 0.1 mg/mL sodium borohydride in PBS for 10 minutes, washed once in PBS, and blocked

in 3% BSA in PBS. To visualize biotin-labeled proteins, coverslips were incubated with 1:1000 AlexaFluor 546-conjugated streptavidin (ThermoFisher S11225) for one hour followed by three washes in PBS. No labeling of GFP was necessary, as these fixation conditions preserve intrinsic GFP fluorescence (Tonkin et al., 2004). Coverslips were mounted onto slides with ProLong Gold antifade reagent with DAPI (ThermoFisher) and were sealed with nail polish prior to imaging.

For immunofluorescence analysis, parasites were processed as above except that fixation was performed with 4% paraformaldehyde and 0.0075% glutaraldehyde in PBS for 20 minutes and blocking was performed with 5% BSA in PBS. Following blocking, primary antibodies were used in 5% BSA in PBS at the following concentrations: 1:500 rabbit-α-ACP (Gallagher and Prigge, 2010); 1:100 rat-α-HA 3F10 (Sigma 11867423001). Coverslips were washed three times in PBS, incubated with goat-α-rat 488 (ThermoFisher A-11006) and donkey-α-rabbit 568 (ThermoFisher A10042) secondary antibodies at 1:3000, and washed three times in PBS prior to mounting as above.

Live and fixed cells were imaged with a 100X, 1.4 NA objective on an Olympus IX70 microscope with a DeltaVision system (Applied Precision) controlled with SoftWorx version 4.1.0 and equipped with a CoolSnap-HQ CCD camera (Photometrics). With the exception of images presented in Figure 1B, which were taken in a single z-plane, images were captured as a series of z-stacks separated by 0.2 µm intervals, deconvolved, and displayed as maximum intensity projections. Brightness and contrast were adjusted in Fiji (ImageJ) for display purposes.

Biotin pulldowns, mass spectrometry, and data analysis

Biotin-labeled parasites were harvested by centrifugation and were released from the host RBC by treatment with 0.1% saponin/PBS. Parasites were washed twice more with 0.1% saponin/PBS

followed by PBS and were either used immediately for analysis or were stored at -80°C. Parasite pellets were resuspended in RIPA buffer [50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% Triton X-100, 1 mM EDTA] containing a protease inhibitor cocktail (Pierce) and were lysed on ice for 30 minutes with occasional pipetting. Insoluble debris was removed by centrifugation at 16,000 xg for 15 minutes at 4°C. Biotinylated proteins were captured using High Capacity Streptavidin Agarose beads (Pierce) for 2 hours at room temperature. Beads were then washed three times with RIPA buffer, three times with SDS wash buffer [50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 2% SDS], six times with urea wash buffer [50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 8 M urea], and three times with 100 mM ammonium bicarbonate. Proteins were reduced with 5 mM DTT for 60 minutes at 37°C followed by treatment with 14 mM iodoacetamide (Pierce) at room temperature for 45 minutes. Beads were washed once with 100 mM ammonium bicarbonate and were digested with 10 µg/mL trypsin (Promega) at 37°C overnight. The following day, samples were digested with an additional 5 µg/mL trypsin for 3-4 hours. Digested peptides were separated from beads by addition of either 35% or 50% final concentration acetonitrile, and peptides were dried on a SpeedVac prior to desalting with C18 stage tips.

Desalted peptides were resuspended in 0.1% formic acid and analyzed by online capillary nanoLC-MS/MS. Samples were separated on an in-house made 20 cm reversed phase column (100 µm inner diameter, packed with ReproSil-Pur C18-AQ 3.0 µm resin (Dr. Maisch GmbH)) equipped with a laser-pulled nanoelectrospray emitter tip. Peptides were eluted at a flow rate of 400 nL/min using a two-step linear gradient including 2-25% buffer B in 70 min and 25-40% B in 20 min (buffer A: 0.2% formic acid and 5% DMSO in water; buffer B: 0.2% formic acid and 5% DMSO in acetonitrile) in an Eksigent ekspert nanoLC-425 system (AB Sciex). Peptides were

548 then analyzed using a LTQ Orbitrap Elite mass spectrometer (Thermo Scientific). Data
549 acquisition was executed in data dependent mode with full MS scans acquired in the Orbitrap
550 mass analyzer with a resolution of 60000 and m/z scan range of 340-1600. The top 20 most
551 abundant ions with intensity threshold above 500 counts and charge states 2 and above were
552 selected for fragmentation using collision- induced dissociation (CID) with isolation window of
553 2 m/z, normalized collision energy of 35%, activation Q of 0.25 and activation time of 5 ms. The
554 CID fragments were analyzed in the ion trap with rapid scan rate. In additional runs, the top 10
555 most abundant ions with intensity threshold above 500 counts and charge states 2 and above
556 were selected for fragmentation using higher energy collisional dissociation (HCD) with
557 isolation window of 2 m/z, normalized collision energy of 35%, and activation time of 25 ms.
558 The HCD fragments were analyzed in the Orbitrap with a resolution of 15000. Dynamic
559 exclusion was enabled with repeat count of 1 and exclusion duration of 30 s. The AGC target
560 was set to 1000000, 50000, and 5000 for full FTMS scans, FTMSn scans and ITMSn scans,
561 respectively. The maximum injection time was set to 250 ms, 250 ms, and 100 ms for full FTMS
562 scans, FTMSn scans and ITMSn scans, respectively.

563 The resulting spectra were searched against a “target-decoy” sequence database (Elias
564 and Gygi, 2007) consisting of the PlasmoDB protein database (release 32, released April 19,
565 2017), the Uniprot human database (released February 2, 2015), and the corresponding reversed
566 sequences using the SEQUEST algorithm (version 28, revision 12). The parent mass tolerance
567 was set to 50 ppm and the fragment mass tolerance to 0.6 Da for CID scans, 0.02 Da for HCD
568 scans. Enzyme specificity was set to trypsin. Oxidation of methionines was set as variable
569 modification and carbamidomethylation of cysteines was set as static modification. Peptide

identifications were filtered to a 1% peptide false discovery rate using a linear discriminator analysis (Huttlin et al., 2010). Precursor peak areas were calculated for protein quantification.

Apicoplast protein prediction algorithms and positive/negative control apicoplast proteins

To generate updated lists of PATS-predicted apicoplast proteins, nuclear-encoded *P. falciparum* 3D7 proteins (excluding pseudogenes) from PlasmoDB version 28 (released March 30, 2016) were used to check for existence of a putative bipartite apicoplast targeting presequence using the artificial neural network predictor PATS (Zuegge et al., 2001).

Updated PlasmoAP-predicted apicoplast proteins were identified using the PlasmoDB version 32 proteome (released April 19, 2017) by first checking for the presequence of a predicted signal peptide using the neural network version of SignalP version 3.0 (Bendtsen et al., 2004), and were considered positive if they had a *D-score* above the default cutoff. The SignalP *C-score* was used to predict the signal peptide cleavage position, and the remaining portion of the protein was inspected for presence of a putative apicoplast transit peptide using the rules described for PlasmoAP (Foth et al., 2003), implemented in a Perl script.

P. falciparum proteins predicted to localize to the apicoplast by ApicoAP were accessed from the original paper (Cilingir et al., 2012). Genes predicted to encode pseudogenes were excluded.

A positive control list of 96 high-confidence apicoplast proteins (Table S2) was generated based on either (1) published localization of that protein in *Plasmodium* parasites or *Toxoplasma gondii* or (2) presence of that protein in either the isoprenoid biosynthesis or fatty acid biosynthesis/utilization pathways. To generate a negative control list of potential false positives, nuclear-encoded proteins (excluding pseudogenes) predicted to contain a signal peptide were

identified as above and 451 of these proteins were designated as negative controls based on GO terms, annotations, and the published literature.

Feature extraction for neural network

To generate the positive training set for PlastNN, we took the combined list of previously known apicoplast proteins (Table S2) and apicoplast proteins identified by BioID (Table S1) and removed proteins that (1) were likely false positives based on manual inspection; (2) were likely targeted to the apicoplast without the canonical bipartite *N*-terminal leader sequence; or (3) did not contain a predicted signal peptide based on the SignalP 3.0 *D*-score. This yielded a final positive training set of 205 proteins (Table S4). The negative training set was the previously generated list of known non-apicoplast proteins (Table S2). The test set for PlastNN consisted of 450 proteins predicted to have a signal peptide by the SignalP 3.0 *D*-score that were not in the positive or negative training sets.

For each protein in our training and test sets, we took the 50 amino acids immediately after the end of the predicted signal peptide (according to the SignalP 3.0 *C*-score) and calculated the frequency of each of the 20 amino acids in this sequence. The length of 50 amino acids was chosen empirically by trying lengths from 20-100; highest accuracy was obtained using 50. Scaled FPKM values at 8 time points during intraerythrocytic development were obtained from published RNA-Seq (Bartfai et al., 2010). By combining the amino acid frequencies with the 8 transcriptome values, we represented each protein in our training and test sets by a feature vector of length 28.

Neural network training and cross-validation

To train the model, the 205 positive and 451 negative training examples were combined and randomly shuffled. The training set was divided into 6 equal folds, each containing 109 or 110 examples. We trained models using 6-fold cross-validation; that is, we trained 6 separate models with the same architecture, each using 5 of the 6 folds for training and then using the one remaining fold as a cross-validation set to evaluate performance. Accuracy, sensitivity, and PPV are calculated on this cross-validation set. The final reported values of accuracy, sensitivity, and PPV are the average and standard deviation over all 6 models. When predicting on the test set, the final predictions are generated by a majority vote of all 6 models.

Neural networks were trained using the RMSProp optimization algorithm with a learning rate of 0.0001. Tensorflow version 1.4.1 was used to build and train the neural network. Logistic regression on the same dataset was carried out using the caret package (version 6.0-77) in R version 3.3.3.

Accuracy, sensitivity, and positive predictive value (PPV) calculations

The BioID apicoplast proteome and the predicted proteomes from PATS, PlasmoAP, ApicoAP, and PlastNN were analyzed for accuracy, sensitivity, and PPV according to the following formulae:

$$\text{Accuracy} = (TP + TN)/(TP + FP + TN + FN)$$

$$\text{Sensitivity} = TP/(TP + FN)$$

$$\text{PPV} = TP/(TP + FP)$$

Abbreviations: TP, true positive; TN, true negative; FP, false positive; FN, false negative.

637 Because none of the 451 negative control proteins were identified in our 187-protein BioID
638 proteome, for the purposes of calculating PPV in Figure 2C we identified 5 likely false positives
639 by manual inspection of these 187 proteins and added these to the negative control list.
640
641 **Protein Novelty Analysis**
642 Proteins in the apicoplast proteome were manually categorized for having a potentially novel
643 function based on PlasmoDB version 33 (released June 30, 2017) gene product annotations.
644 Gene products with annotations that could clearly assign a given protein to an established
645 cellular pathway were labeled as "Known Pathway;" gene products with a descriptive annotation
646 that did not clearly suggest a cellular pathway were labeled as "Annotated Gene Product,
647 Unknown Function;" and gene products that explicitly contained the words "unknown function"
648 were labeled as "Unknown Function."
649
650 **OrthoMCL Orthology Analysis**
651 To analyze the conservation of candidate apicoplast proteins identified by apicoplast BioID,
652 OrthoMCL ortholog group IDs were obtained from PlasmoDB. Based on OrthoMCL version 5
653 (released July 23, 2015), each ortholog group was then categorized as being present only in
654 *Plasmodium* spp., only in Apicomplexa, or present in at least one organism outside of the
655 Apicomplexa.
656
657 **Gene Essentiality Analysis**
658 Essentiality data for *P. berghei* orthologs of *P. falciparum* genes encoding apicoplast proteins
659 were accessed from the original PlasmoGEM manuscript (Bushell et al., 2017).

660

661 **Parasite Growth Time Courses**

662 Sorbitol-synchronized ABCB7 and ABCF1 TetR-DOZI parasites were washed multiple times to
663 remove residual ATc and were returned to culture medium containing 500 nM ATc, 200 µM IPP
664 (Isoprenoids, LLC), or no supplements. Samples for growth assays, DNA isolation, or western
665 blotting were harvested every other day when the majority of parasites were trophozoites and
666 schizonts. For growth assays, parasites were fixed in 1% paraformaldehyde in PBS and were
667 stored at 4°C until completion of the time course. Samples were then stained with 50 nM
668 YOYO-1 and parasitemia was analyzed on a BD Accuri C6 flow cytometer. Samples for DNA
669 isolation and western blotting were treated with 0.1% saponin in PBS to release parasites from
670 the erythrocyte, washed in PBS, and stored at -80°C until analysis.

671

672 **Quantitative PCR**

673 Total parasite DNA was isolated from time course samples using the DNeasy Blood & Tissue
674 Kit (Qiagen). Quantitative PCR was performed using Power SYBR Green PCR Master Mix
675 (Thermo Fisher) with primers CHT1 F and CHT1 R targeting the nuclear gene chitinase or TufA
676 F and TufA R targeting the apicoplast gene elongation factor Tu (0.15 µM final concentration
677 each primer) (Yeh and DeRisi, 2011). Quantitative PCR was performed on an Applied
678 Biosystems 7900HT Real-Time PCR System with the following thermocycling conditions:
679 Initial denaturation 95°C/10 minutes; 35 cycles of 95°C/1 minute, 56°C/1 minute; 65°C/1
680 minute; final extension 65°C/10 minutes. Relative quantification of each target was performed
681 using the $\Delta\Delta C_t$ method.

682

Data and Software Availability

Raw mass spectrometry data are available via the Chorus repository (https://chorusproject.org/) with project identifier 1440. Code related to the development of the PlastNN algorithm is available at https://github.com/sjang92/plastNN.

Acknowledgements

We thank Jacquin Niles for providing the NF54$^{Cas9+T7\ Polymerase}$ cell line and pSN054-V5 plasmid, Sean Prigge for α-ACP antibody, and Walid Houry for α-ClpP antibody. We also thank Julian Lutze for assistance with molecular cloning of candidate apicoplast genes.

Financial Disclosure

Funding for this work was provided by National Institutes of Health grants K08 AI097239 and DP5 OD012119 (E.Y.), a Burroughs Wellcome Fund Career Award for Medical Scientists (E.Y.), the Chan Zuckerberg Biohub Investigator Program (E.Y. and J.Z.), a Stanford Bio-X Interdisciplinary Initiatives Seed Grant (E.Y. and J.E.E.), an NSF CRII grant (J.Z.), a National Health and Medical Research Council RD Wright Biomedical fellowship (S.A.R.), and a William R. and Sara Hart Kimball Stanford Graduate Fellowship (M.J.B.).

Author Contributions

Conceptualization, M.J.B. and E.Y.; Software, A.L., S.J.W., A.J., S.Z., X.W., J.Z. and S.A.R.; Investigation, M.J.B., and S.G.; Resources, L.Z., J.E.E., and S.A.R.; Writing – Original Draft, M.J.B. and E.Y.; Writing – Review & Editing, M.J.B., S.G., L.Z., A.L., S.W.J., A.J., S.Z., X.W.,

S.A.R., J.Z., J.E.E., and E.Y.; Supervision, E.Y., J.E.E., and J.Z.; Funding Acquisition, E.Y., J.E.E., J.Z., and S.A.R.

Figures and Legends

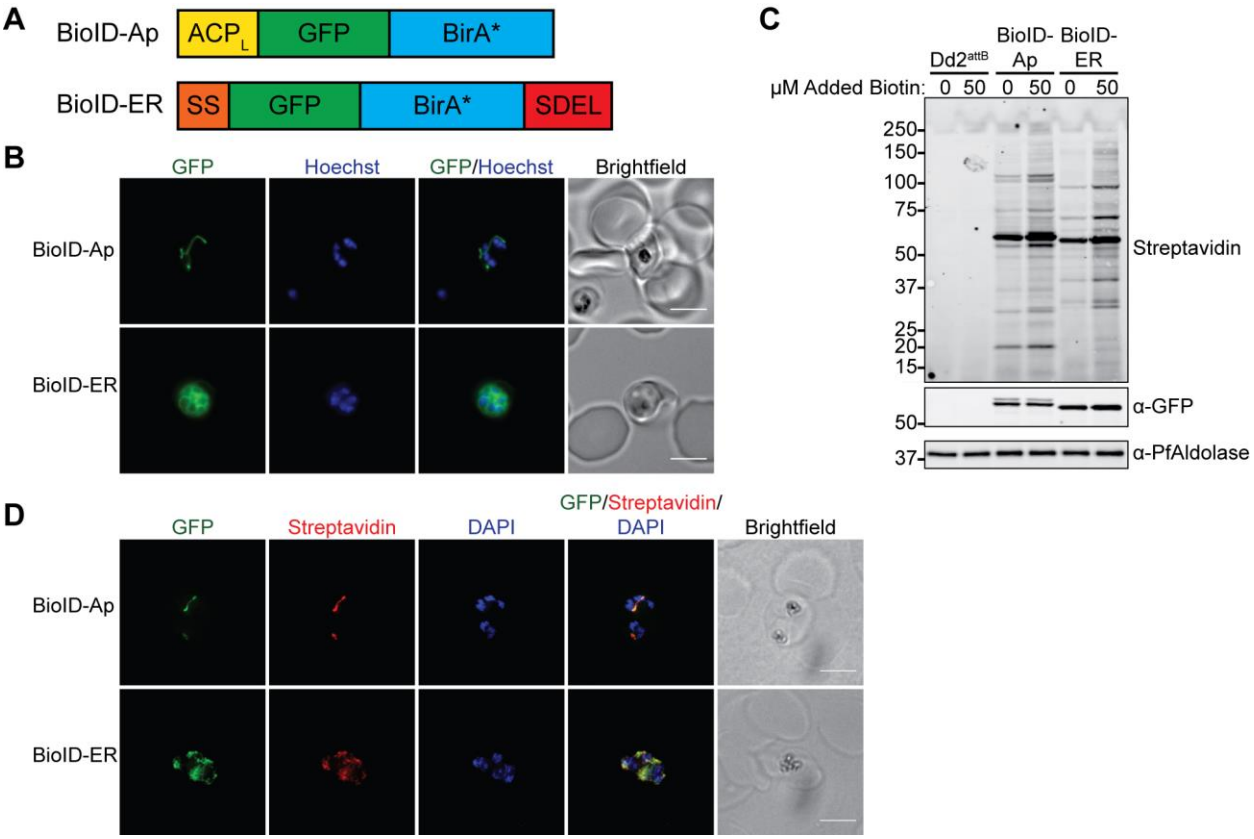
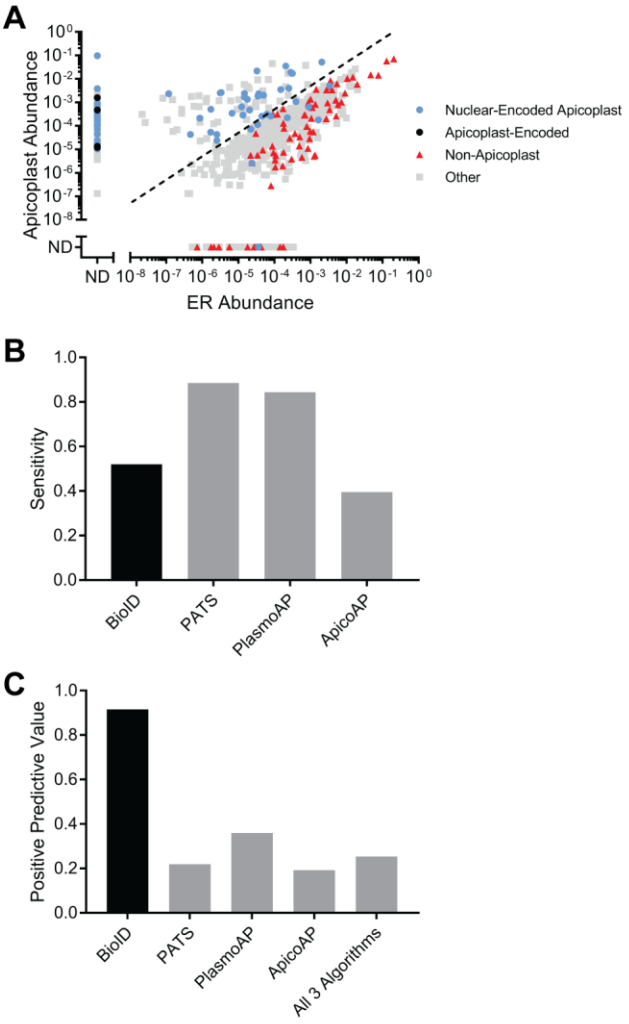


Figure 1. The promiscuous biotin ligase BirA* biotinylates proteins in the *P. falciparum* apicoplast and ER. (A) Schematic (not to scale) of constructs for apicoplast- and ER-targeting of GFP-BirA*. ACP_L, ACP leader sequence; SS, signal sequence; SDEL, ER-retention motif. (B) Live cell imaging of Hoechst-stained BioID-Ap and BioID-ER parasites. Scale bars, 5 μm. (C) Western blot of untreated and biotin-labeled $Dd2^{attB}$, BioID-Ap, and BioID-ER parasites. (D) Fixed cell imaging of biotinylated proteins in biotin-labeled BioID-Ap and BioID-ER parasites. Scale bars, 5 μm. See also Figure S1.



718

719 **Figure 2. Accurate, unbiased identification of apicoplast proteins using BioID.** (A)
720 Abundances of 728 proteins identified by mass spectrometry in BioID-Ap and BioID-ER
721 parasites. Protein abundances were calculated by summing the total MS1 area of all matched
722 peptides and normalized by the total summed intensity of all *P. falciparum* peptides matched.
723 Dotted line represents 5-fold apicoplast:ER enrichment. ND, not detected. (B) Sensitivities of
724 BioID, PATS, PlasmoAP, and ApicoAP based on identification of 96 known apicoplast proteins.
725 (C) PPV of BioID, PATS, PlasmoAP, ApicoAP, and a dataset consisting of proteins predicted to
726 localize to the apicoplast by all three bioinformatic algorithms. Calculated as the number of true

positives divided by the total number of true positives and false positives. See also Figure S2, Table S1, Table S2, and Table S3.

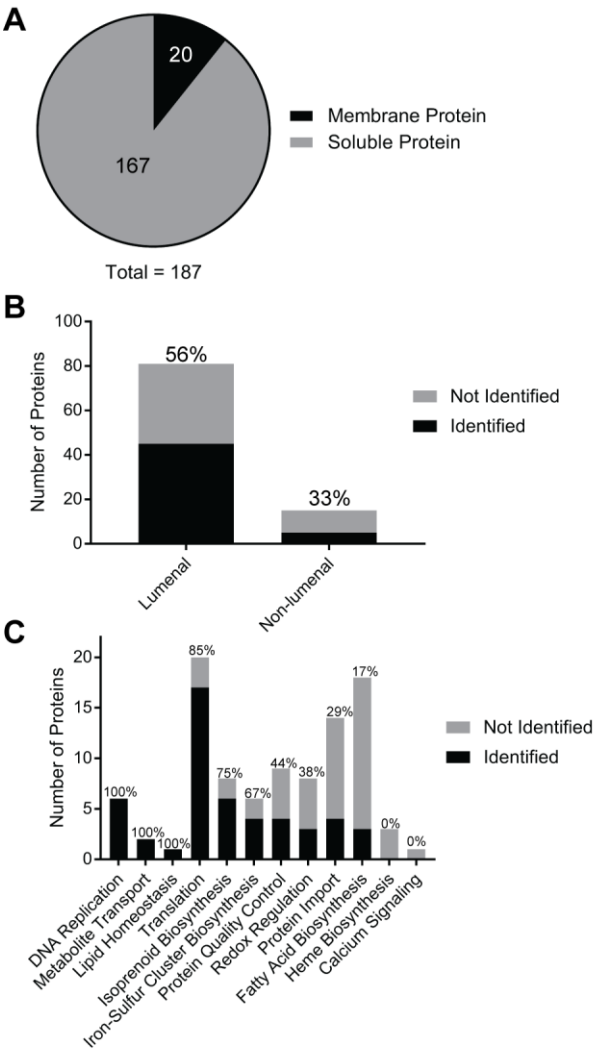
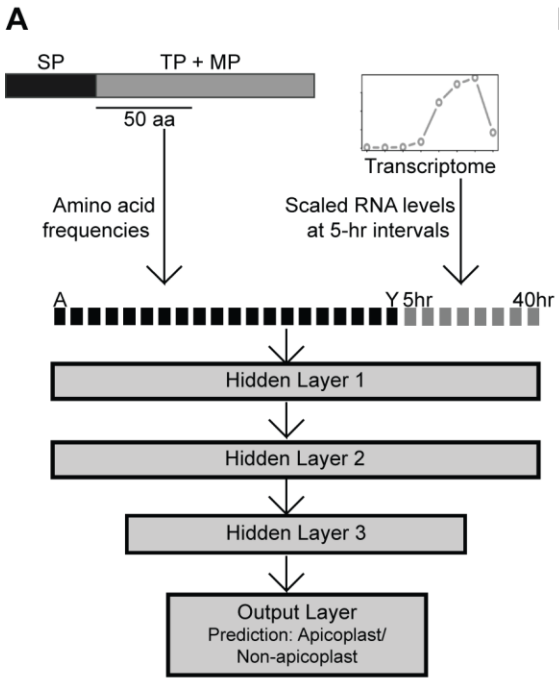


Figure 3. Diversity of protein labeling by apicoplast BioID. (A) Fraction of proteins identified by apicoplast BioID that are predicted to localize to a membrane. Proteins were considered "membrane" if they had at least one transmembrane domain annotated in PlasmoDB ending >80 amino acids from the annotated *N*-terminus. (B) Number of lumenal and non-lumenal positive controls identified. Percentages above bars indicate the percentage of known proteins from each

736 category identified. (C) Number of proteins from established apicoplast pathways identified.
737 Percentages above bars indicate the percentage of known proteins from each pathway identified.
738 See also Table S1 and Table S2.
739



B

Model	1	2	3	4	5	6	Mean	SD
Accuracy	98%	94%	94%	95%	99%	99%	**96%**	**3%**
Sensitivity	100%	91%	89%	90%	100%	97%	**95%**	**5%**
PPV	91%	89%	91%	97%	97%	100%	**94%**	**4%**

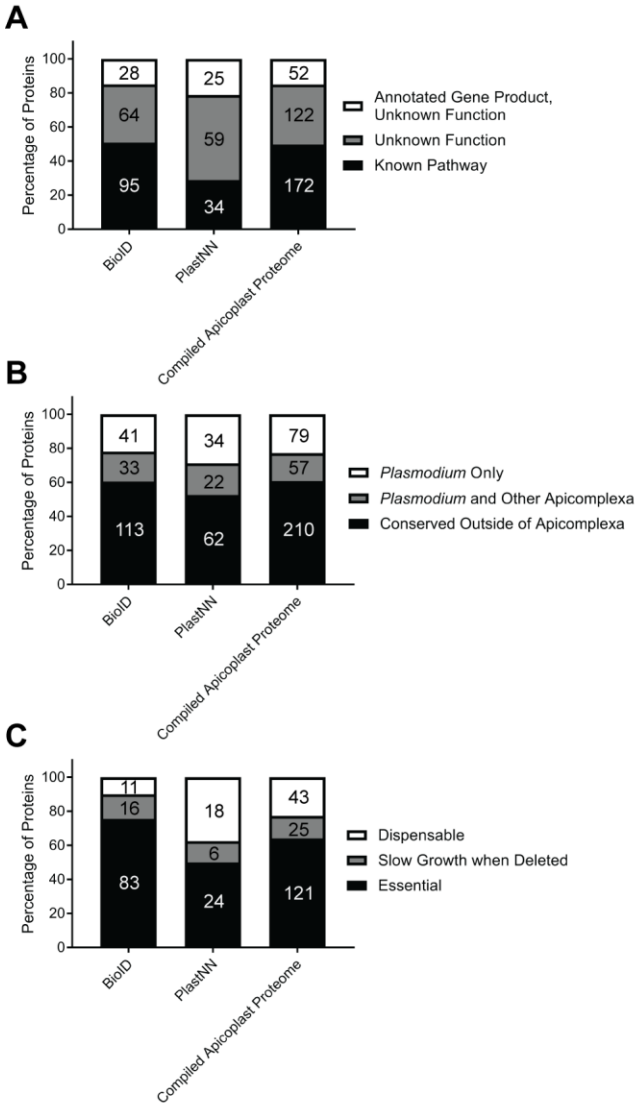
C

Algorithm	Accuracy	Sensitivity	PPV
PATS	53%	97%	40%
PlasmoAP	77%	95%	58%
ApicoAP	66%	70%	48%
PlastNN	**96%**	**95%**	**94%**

740

741 **Figure 4. Improved prediction of apicoplast proteins using the PlastNN algorithm.** (A)
742 Schematic of the PlastNN algorithm. For each signal peptide-containing protein, a region of 50
743 amino acids immediately following the signal peptide cleavage site was selected and the
744 frequencies of the 20 canonical amino acids in this region were calculated, resulting in a vector
745 of length 20. Scaled RNA levels of the gene encoding the protein at 8 time points were added,
746 resulting in a 28-dimensional vector representing each protein. This was used as input to train a
747 neural network with 3 hidden layers, resulting in a prediction of whether the protein is targeted to
748 the apicoplast or not. (B) Table showing the performance of the 6 models in PlastNN. Each
749 model was trained on 5/6th of the training set and cross-validated on the remaining 1/6th. Values
750 shown are accuracy, sensitivity, and PPV on the cross-validation set. The final values reported

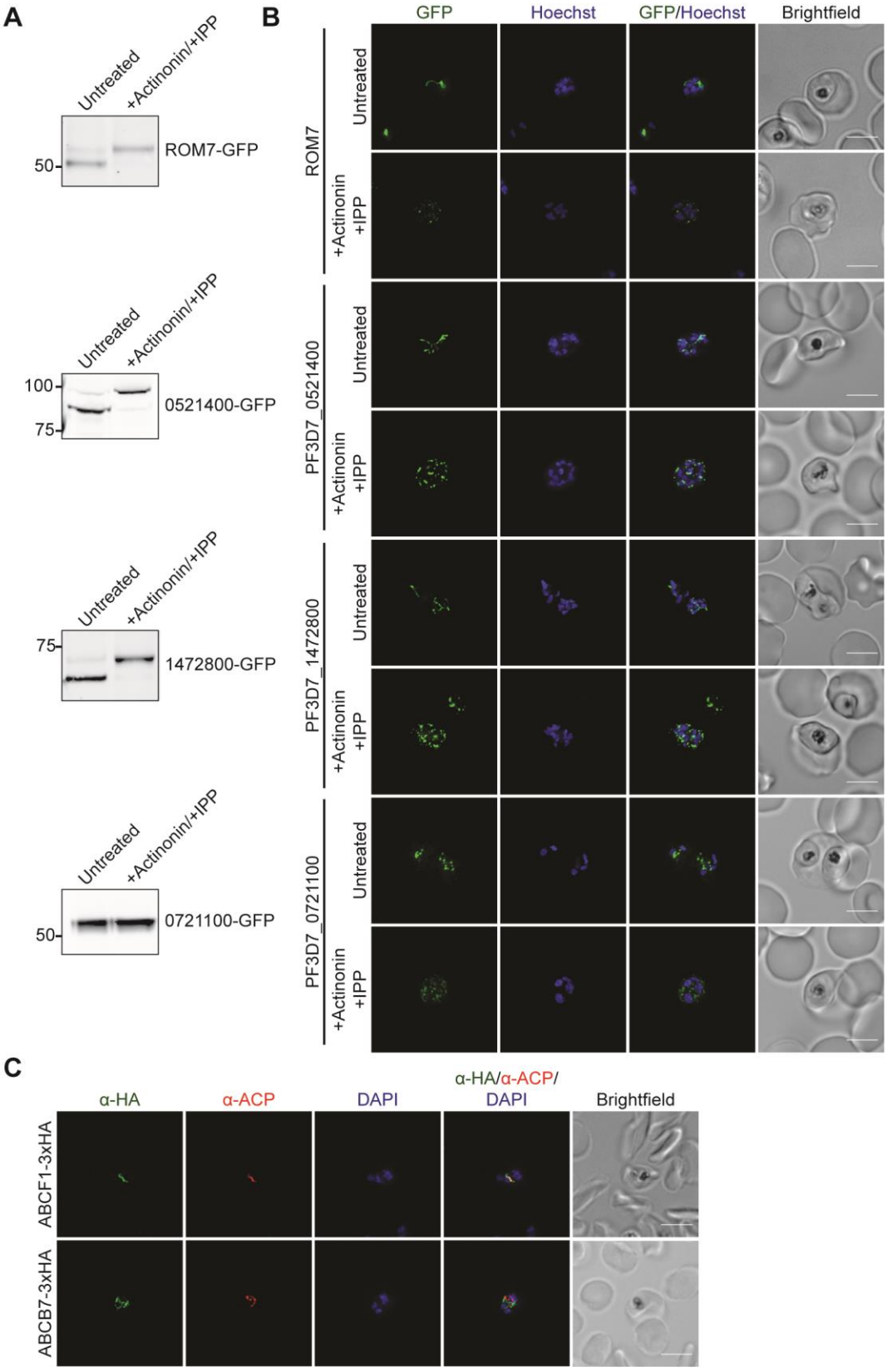
751 are the average and standard deviation over all 6 models. (C) Comparison of accuracy,
752 sensitivity, and PPV for three previous algorithms and PlastNN. See also Table S2, Table S3,
753 Table S4, Table S5, Table S6, Table S7.
754



755

756 **Figure 5. Apicoplast BioID identifies novel and essential proteins.** (A) Percentage of proteins
757 identified that have annotated gene products but unknown function, that have gene products
758 annotated explicitly with "unknown function," or that have annotated gene products and function
759 in a known cellular pathway. (B) Percentage of proteins identified that are *Plasmodium*- or

760 Apicomplexa-specific based on OrthoMCL orthology analysis. (C) Percentage of proteins
761 identified that are essential, cause slow growth when deleted, or are dispensable based on
762 PlasmoGEM essentiality data of *P. berghei* orthologs. Absolute number of proteins identified as
763 indicated. See also Table S1, Table S7, and Table S8.
764



765

766 **Figure 6. Localization of candidate apicoplast proteins identified in this study.** (A) Transit
767 peptide processing assay for *C*-terminally GFP-tagged candidates. Ring-stage parasites were
768 either untreated or treated with 10 µM actinonin/200 µM IPP for 3 days and protein processing
769 was assessed by western blot. (B) Live cell localization of GFP-tagged candidates in apicoplast-
770 intact and -minus parasites. (C) IFAs of endogenously triple HA-tagged ABCB7 and ABCF1.
771 See also Figure S3.
772

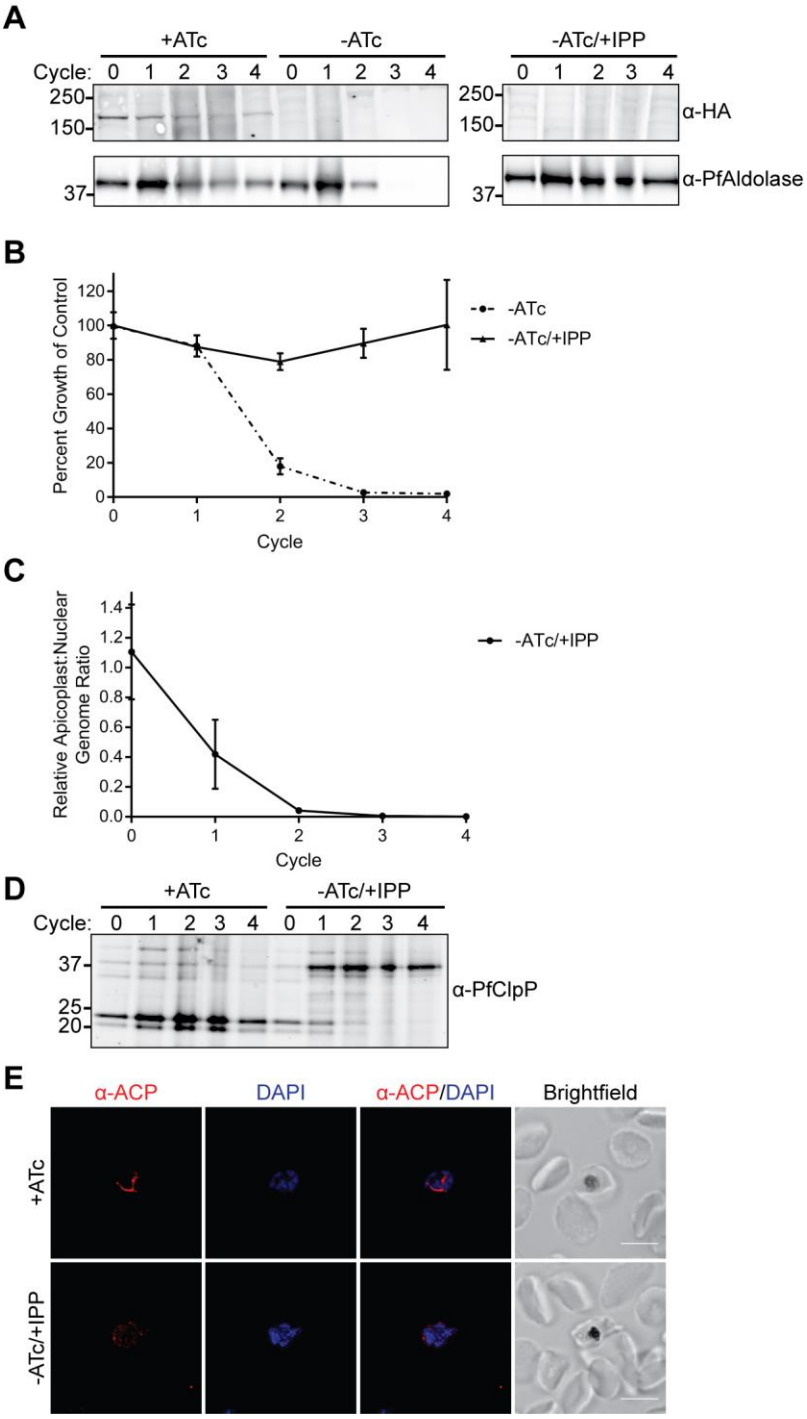
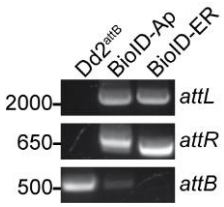


Figure 7. ABCF1 is an essential apicoplast protein required for organelle biogenesis.

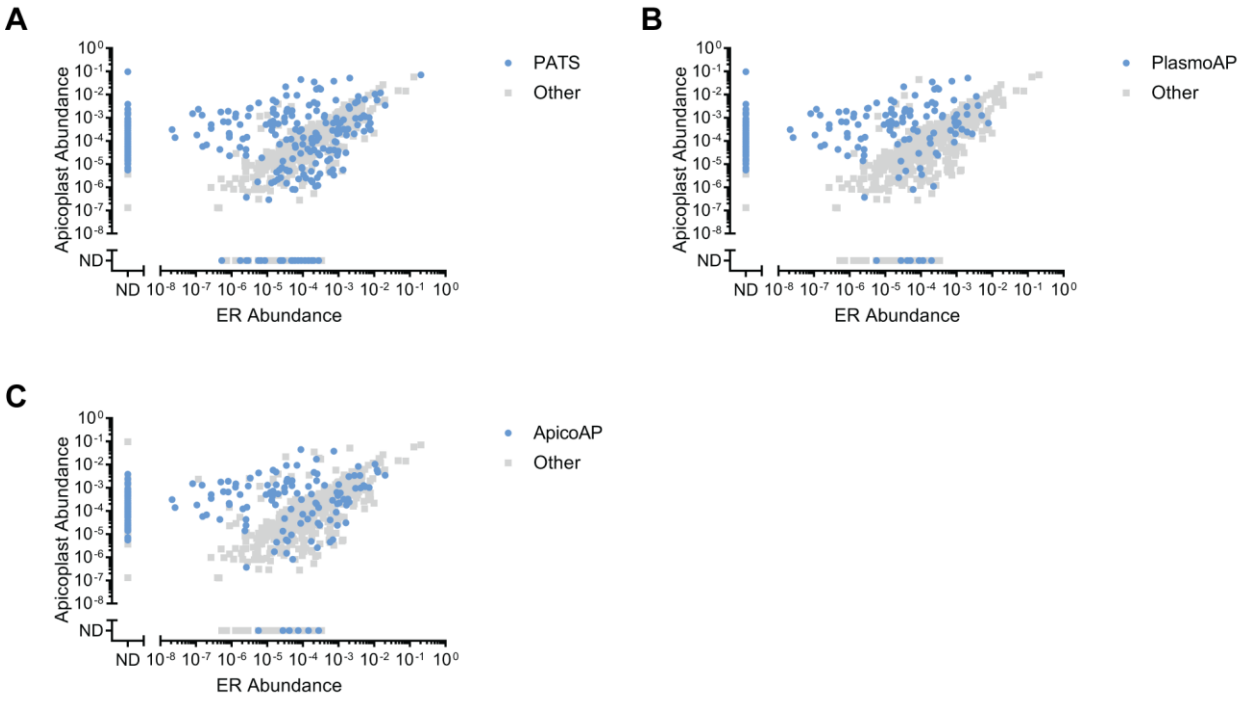
ABCF1-3xHA knockdown parasites were grown in the presence of ATc (+ATc), the absence of ATc (-ATc), or the absence of ATc with IPP supplementation (-ATc/+IPP) for 4 growth cycles. (A) ABCF1-HA expression. (B) Parasite growth. At each time point, data are normalized to the

778 untreated (+ATc) control. Error bars represent standard deviation of the mean of two biological
779 replicates. (C) Apicoplast:nuclear genome ratio. At each time point, data are normalized to the
780 untreated (+ATc) control. Error bars represent standard deviation of the mean of two biological
781 replicates, each performed in technical triplicate. (D) ClpP processing. (E) ACP localization after
782 2 cycles of knockdown. See also Figure S4.
783



784

785 **Figure S1, Related to Figure 1. Integration of BioID-Ap and BioID-ER constructs into**
786 **Dd2^{attB} parasites.** PCR products showing integrated *attL* and *attR* sites or unintegrated *attB* site.
787



788

Figure S2, Related to Figure 2. Bioinformatically predicted apicoplast proteins are not clearly distinguishable based on apicoplast:ER abundance ratio. Proteins predicted to localize to the apicoplast by (A) PATS, (B) PlasmoAP, or (C) ApicoAP are highlighted in each graph. Data points are identical to those in Figure 2A.

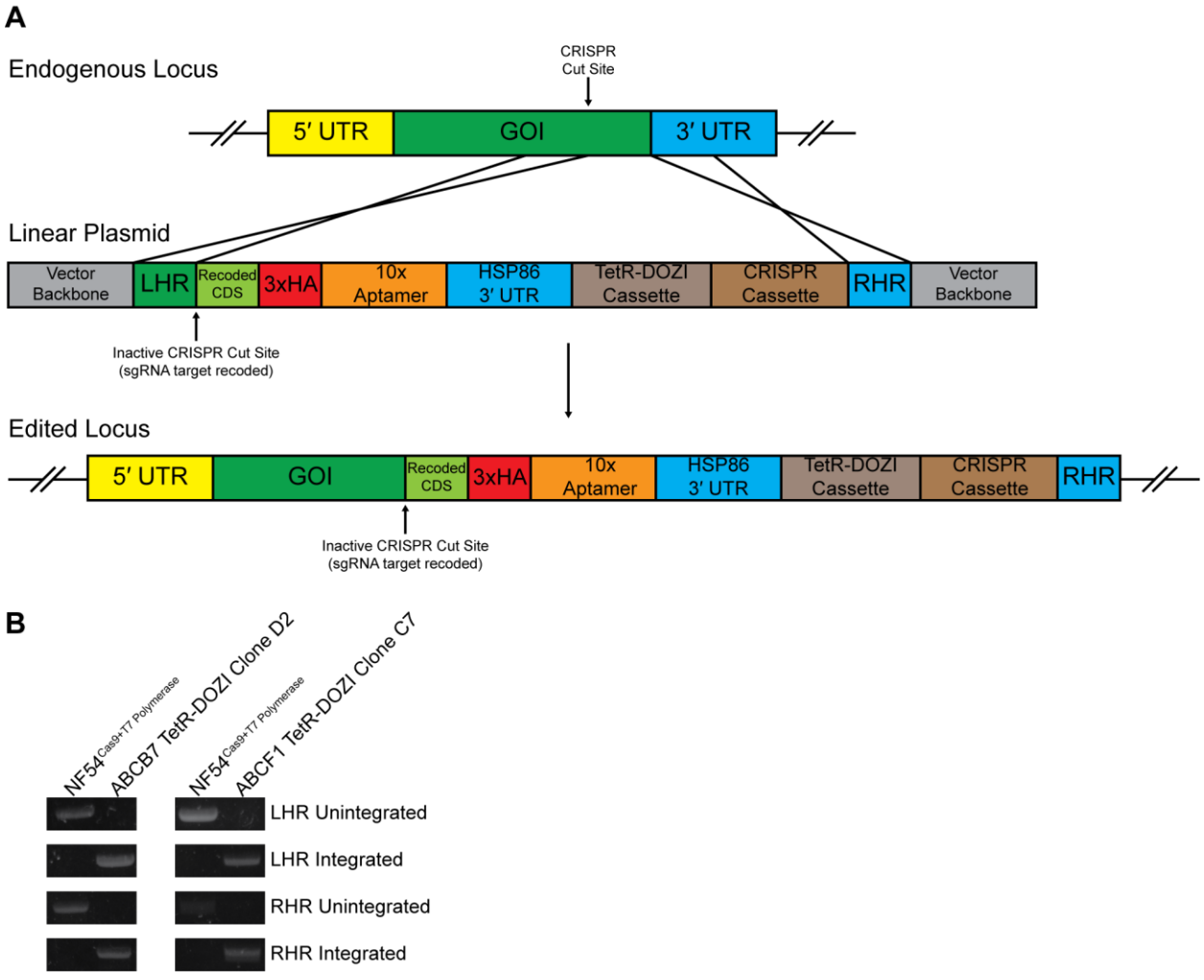


Figure S3, Related to Figure 6. Generation of ABCB7 and ABCF1 TetR-DOZI conditional knockdown cell lines. (A) Schematic of CRISPR-Cas9-based endogenous editing to generate conditional knockdown cell lines. GOI, gene of interest; LHR, left homology region; RHR, right

homology region. (B) PCR products showing integrated or unintegrated LHR and RHR sites in parental NF54$^{Cas9+T7\ Polymerase}$ or clonal genome-edited parasites.

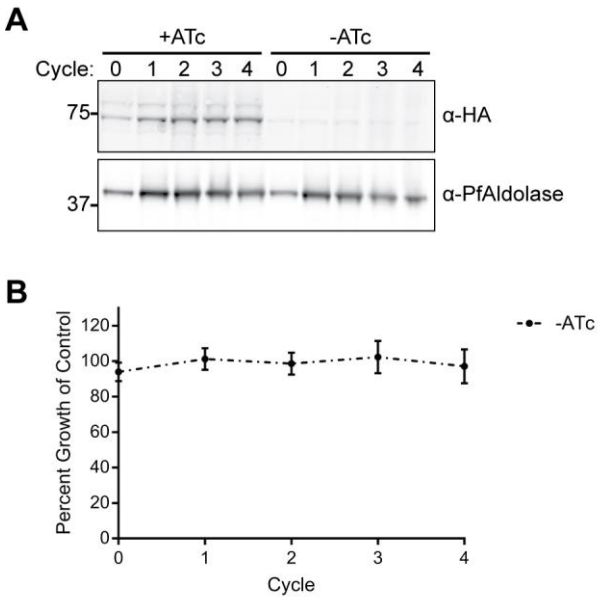


Figure S4, Related to Figure 7. Knockdown of ABCB7 does not cause growth inhibition. ABCB7-3xHA knockdown parasites were grown in the presence of ATc (+ATc) or the absence of ATc (-ATc) for 4 growth cycles. (A) ABCB7-HA expression. (B) Parasite growth. At each time point, data are normalized to the untreated (+ATc) control. Error bars represent standard deviation of the mean of two biological replicates.

Table S1, Related to Figures 2, 3, and 5. Abundances of 728 *P. falciparum* proteins identified by mass spectrometry in ≥2 biological replicates and with ≥2 unique peptides in at least one mass spectrometry run.

Table S2, Related to Figures 2-4. Positive and negative control apicoplast proteins used in this study.

Table S3, Related to Figures 2 and 4. Proteins predicted to localize to the apicoplast by PATS, PlasmoAP, and ApicoAP.

Table S4, Related to Figure 4. Positive training set used to develop PlastNN.

Table S5, Related to Figure 4. Layer dimensions for PlastNN neural network.

Layer	Input	Hidden 1	Hidden 2	Hidden 3	Output
Dimension	28	64	64	16	2
Non-linearity	-	ReLU	ReLU	ReLU	Softmax

Table S6, Related to Figure 4. Performance of different models in cross-validation.

Model	**Features**	**Hyperparameters**	**Accuracy**		**Sensitivity**		**PPV**	
			Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Logistic regression	Amino acid frequencies		0.88	0.04	0.79	0.09	0.84	0.08
Logistic regression	Amino acid frequencies + transcriptome	LASSO penalty; lambda = 0.005	0.91	0.03	0.85	0.06	0.87	0.05
Neural network	Amino acid frequencies	L2 regularization; 0.05 Learning rate = 0.001	0.91	0.02	0.83	0.09	0.87	0.08
Neural network	Amino acid frequencies + transcriptome		0.96	0.03	0.95	0.05	0.94	0.04

Table S7, Related to Figures 4 and 5. Results of PlastNN prediction algorithm.

Table S8, Related to Figure 5. Compiled list of 346 candidate apicoplast proteins based on localization in the published literature, BioID, and PlastNN.

Table S9, Related to Figures 1 and 6. Primer and gBlock sequences used in this study.

References

825 Agrawal, S., Chung, D.W., Ponts, N., van Dooren, G.G., Prudhomme, J., Brooks, C.F.,
826 Rodrigues, E.M., Tan, J.C., Ferdig, M.T., Striepen, B., *et al.* (2013). An apicoplast localized
827 ubiquitylation system is required for the import of nuclear-encoded plastid proteins. PLOS
828 Pathog. *9*, e1003426.
829
830 Agrawal, S., van Dooren, G.G., Beatty, W.L., and Striepen, B. (2009). Genetic evidence that an
831 endosymbiont-derived endoplasmic reticulum-associated protein degradation (ERAD) system
832 functions in import of apicoplast proteins. J. Biol. Chem. *284*, 33683-33691.
833
834 Amberg-Johnson, K., Hari, S.B., Ganesan, S.M., Lorenzi, H.A., Sauer, R.T., Niles, J.C., and
835 Yeh, E. (2017). Small molecule inhibition of apicomplexan FtsH1 disrupts plastid biogenesis in
836 human pathogens. eLife *6*, e29865.
837
838 Aurrecoechea, C., Barreto, A., Basenko, E.Y., Brestelli, J., Brunk, B.P., Cade, S., Crouch, K.,
839 Doherty, R., Falke, D., Fischer, S., *et al.* (2017). EuPathDB: the eukaryotic pathogen genomics
840 database resource. Nucleic Acids Res. *45*, D581-D591.
841
842 Balabaskaran Nina, P., Morrisey, J.M., Ganesan, S.M., Ke, H., Pershing, A.M., Mather, M.W.,
843 and Vaidya, A.B. (2011). ATP synthase complex of *Plasmodium falciparum*: dimeric assembly
844 in mitochondrial membranes and resistance to genetic disruption. J. Biol. Chem. *286*, 41312-
845 41322.
846
847 Bartfai, R., Hoeijmakers, W.A., Salcedo-Amaya, A.M., Smits, A.H., Janssen-Megens, E., Kaan,
848 A., Treeck, M., Gilberger, T.W., Francoijs, K.J., and Stunnenberg, H.G. (2010). H2A.Z
849 demarcates intergenic regions of the *Plasmodium falciparum* epigenome that are dynamically
850 marked by H3K9ac and H3K4me3. PLOS Pathog. *6*, e1001223.
851
852 Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved prediction of
853 signal peptides: SignalP 3.0. J. Mol. Biol. *340*, 783-795.
854
855 Boel, G., Smith, P.C., Ning, W., Englander, M.T., Chen, B., Hashem, Y., Testa, A.J., Fischer,
856 J.J., Wieden, H.J., Frank, J., *et al.* (2014). The ABC-F protein EttA gates ribosome entry into the
857 translation elongation cycle. Nat. Struct. Mol. Biol. *21*, 143-151.
858
859 Bozdech, Z., Llinas, M., Pulliam, B.L., Wong, E.D., Zhu, J., and DeRisi, J.L. (2003). The
860 transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. PLOS
861 Biol. *1*, E5.
862
863 Bushell, E., Gomes, A.R., Sanderson, T., Anar, B., Girling, G., Herd, C., Metcalf, T.,
864 Modrzynska, K., Schwach, F., Martin, R.E., *et al.* (2017). Functional profiling of a *Plasmodium*
865 genome reveals an abundance of essential genes. Cell *170*, 260-272.
866
867 Chen, A.L., Kim, E.W., Toh, J.Y., Vashisht, A.A., Rashoff, A.Q., Van, C., Huang, A.S., Moon,
868 A.S., Bell, H.N., Bentolila, L.A., *et al.* (2015). Novel components of the *Toxoplasma* inner
869 membrane complex revealed by BioID. mBio *6*, e02357-14.
870

Chen, A.L., Moon, A.S., Bell, H.N., Huang, A.S., Vashisht, A.A., Toh, J.Y., Lin, A.H., Nadipuram, S.M., Kim, E.W., Choi, C.P., *et al.* (2017). Novel insights into the composition and function of the *Toxoplasma* IMC sutures. Cell. Microbiol. *19*, e12678.

Chen, B., Boel, G., Hashem, Y., Ning, W., Fei, J., Wang, C., Gonzalez, R.L., Jr., Hunt, J.F., and Frank, J. (2014). EttA regulates translation by binding the ribosomal E site and restricting ribosome-tRNA dynamics. Nat. Struct. Mol. Biol. *21*, 152-159.

Chen, F., Mackey, A.J., Stoeckert, C.J., Jr., and Roos, D.S. (2006). OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res. *34*, D363-D368.

Cilingir, G., Broschat, S.L., and Lau, A.O. (2012). ApicoAP: the first computational model for identifying apicoplast-targeted proteins in multiple species of Apicomplexa. PLOS ONE *7*, e36598.

Dahl, E.L., and Rosenthal, P.J. (2007). Multiple antibiotics exert delayed effects against the *Plasmodium falciparum* apicoplast. Antimicrob. Agents Chemother. *51*, 3485-3490.

Dahl, E.L., Shock, J.L., Shenai, B.R., Gut, J., DeRisi, J.L., and Rosenthal, P.J. (2006). Tetracyclines specifically target the apicoplast of the malaria parasite *Plasmodium falciparum*. Antimicrob. Agents Chemother. *50*, 3124-3131.

Dang, H.Q., Zhou, Q., Rowlett, V.W., Hu, H., Lee, K.J., Margolin, W., and Li, Z. (2017). Proximity interactions among basal body components in *Trypanosoma brucei* identify novel regulators of basal body biogenesis and inheritance. mBio *8*, e02120-16.

Dean, M., and Annilo, T. (2005). Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates. Annu. Rev. Genomics Hum. Genet. *6*, 123-142.

Deitsch, K., Driskill, C., and Wellems, T. (2001). Transformation of malaria parasites by the spontaneous uptake and expression of DNA from human erythrocytes. Nucleic Acids Res. *29*, 850-853.

El Bakkouri, M., Pow, A., Mulichak, A., Cheung, K.L., Artz, J.D., Amani, M., Fell, S., de Koning-Ward, T.F., Goodman, C.D., McFadden, G.I., *et al.* (2010). The Clp chaperones and proteases of the human malaria parasite *Plasmodium falciparum*. J. Mol. Biol. *404*, 456-477.

Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat. Methods *4*, 207-214.

Foth, B.J., Ralph, S.A., Tonkin, C.J., Struck, N.S., Fraunholz, M., Roos, D.S., Cowman, A.F., and McFadden, G.I. (2003). Dissecting apicoplast targeting in the malaria parasite *Plasmodium falciparum*. Science *299*, 705-708.

915 Gallagher, J.R., and Prigge, S.T. (2010). *Plasmodium falciparum* acyl carrier protein crystal
916 structures in disulfide-linked and reduced states and their prevalence during blood stage growth.
917 Proteins *78*, 575-588.
918
919 Ganesan, S.M., Falla, A., Goldfless, S.J., Nasamu, A.S., and Niles, J.C. (2016). Synthetic RNA-
920 protein modules integrated with native translation mechanisms to control gene expression in
921 malaria parasites. Nat. Commun. *7*, 10727.
922
923 Goldfless, S.J., Wagner, J.C., and Niles, J.C. (2014). Versatile control of *Plasmodium falciparum*
924 gene expression with an inducible protein-RNA interaction. Nat. Commun. *5*, 5329.
925
926 Goodman, C.D., Su, V., and McFadden, G.I. (2007). The effects of anti-bacterials on the malaria
927 parasite *Plasmodium falciparum*. Mol. Biochem. Parasitol. *152*, 181-191.
928
929 Gupta, G.D., Coyaud, E., Goncalves, J., Mojarad, B.A., Liu, Y., Wu, Q., Gheiratmand, L.,
930 Comartin, D., Tkach, J.M., Cheung, S.W., *et al.* (2015). A dynamic protein interaction landscape
931 of the human centrosome-cilium interface. Cell *163*, 1484-1499.
932
933 Hung, V., Lam, S.S., Udeshi, N.D., Svinkina, T., Guzman, G., Mootha, V.K., Carr, S.A., and
934 Ting, A.Y. (2017). Proteomic mapping of cytosol-facing outer mitochondrial and ER membranes
935 in living human cells by proximity biotinylation. eLife *6*, e24463.
936
937 Hung, V., Zou, P., Rhee, H.W., Udeshi, N.D., Cracan, V., Svinkina, T., Carr, S.A., Mootha,
938 V.K., and Ting, A.Y. (2014). Proteomic mapping of the human mitochondrial intermembrane
939 space in live cells via ratiometric APEX tagging. Mol. Cell *55*, 332-341.
940
941 Huttlin, E.L., Jedrychowski, M.P., Elias, J.E., Goswami, T., Rad, R., Beausoleil, S.A., Villen, J.,
942 Haas, W., Sowa, M.E., and Gygi, S.P. (2010). A tissue-specific atlas of mouse protein
943 phosphorylation and expression. Cell *143*, 1174-1189.
944
945 Jomaa, H., Wiesner, J., Sanderbrand, S., Altincicek, B., Weidemeyer, C., Hintz, M., Turbachova,
946 I., Eberl, M., Zeidler, J., Lichtenthaler, H.K., *et al.* (1999). Inhibitors of the nonmevalonate
947 pathway of isoprenoid biosynthesis as antimalarial drugs. Science *285*, 1573-1576.
948
949 Kalanon, M., Tonkin, C.J., and McFadden, G.I. (2009). Characterization of two putative protein
950 translocation components in the apicoplast of *Plasmodium falciparum*. Eukaryot. Cell *8*, 1146-
951 1154.
952
953 Ke, H., Sigala, P.A., Miura, K., Morrisey, J.M., Mather, M.W., Crowley, J.R., Henderson, J.P.,
954 Goldberg, D.E., Long, C.A., and Vaidya, A.B. (2014). The heme biosynthesis pathway is
955 essential for *Plasmodium falciparum* development in mosquito stage but not in blood stages. J.
956 Biol. Chem. *289*, 34827-34837.
957
958 Kerr, I.D. (2004). Sequence analysis of twin ATP binding cassette proteins involved in
959 translational control, antibiotic resistance, and ribonuclease L inhibition. Biochem. Biophys. Res.
960 Commun. *315*, 166-173.

Khosh-Naucke, M., Becker, J., Mesen-Ramirez, P., Kiani, P., Birnbaum, J., Frohlke, U., Jonscher, E., Schluter, H., and Spielmann, T. (2018). Identification of novel parasitophorous vacuole proteins in *P. falciparum* parasites using BioID. Int. J. Med. Microbiol. *308*, 13-24.

Kim, D.I., Birendra, K.C., Zhu, W., Motamedchaboki, K., Doye, V., and Roux, K.J. (2014). Probing nuclear pore complex architecture with proximity-dependent biotinylation. Proc. Natl. Acad. Sci. U.S.A. *111*, E2453-E2461.

Kohler, S., Delwiche, C.F., Denny, P.W., Tilney, L.G., Webster, P., Wilson, R.J., Palmer, J.D., and Roos, D.S. (1997). A plastid of probable green algal origin in Apicomplexan parasites. Science *275*, 1485-1489.

Lau, J.B., Stork, S., Moog, D., Schulz, J., and Maier, U.G. (2016). Protein-protein interactions indicate composition of a 480 kDa SELMA complex in the second outermost membrane of diatom complex plastids. Mol. Microbiol. *100*, 76-89.

Loh, K.H., Stawski, P.S., Draycott, A.S., Udeshi, N.D., Lehrman, E.K., Wilton, D.K., Svinkina, T., Deerinck, T.J., Ellisman, M.H., Stevens, B., *et al.* (2016). Proteomic analysis of unbounded cellular compartments: synaptic clefts. Cell *166*, 1295-1307.

Marton, M.J., Vazquez de Aldana, C.R., Qiu, H., Chakraburtty, K., and Hinnebusch, A.G. (1997). Evidence that GCN1 and GCN20, translational regulators of *GCN4*, function on elongating ribosomes in activation of eIF2alpha kinase GCN2. Mol. Cell. Biol. *17*, 4474-4489.

McFadden, G.I., Reith, M.E., Munholland, J., and Lang-Unnasch, N. (1996). Plastid in human parasites. Nature *381*, 482.

Morriswood, B., Havlicek, K., Demmel, L., Yavuz, S., Sealey-Cardona, M., Vidilaseris, K., Anrather, D., Kostan, J., Djinovic-Carugo, K., Roux, K.J., *et al.* (2013). Novel bilobe components in *Trypanosoma brucei* identified using proximity-dependent biotinylation. Eukaryot. Cell *12*, 356-367.

Mullin, K.A., Lim, L., Ralph, S.A., Spurck, T.P., Handman, E., and McFadden, G.I. (2006). Membrane transporters in the relict plastid of malaria parasites. Proc. Natl. Acad. Sci. U.S.A. *103*, 9572-9577.

Nadipuram, S.M., Kim, E.W., Vashisht, A.A., Lin, A.H., Bell, H.N., Coppens, I., Wohlschlegel, J.A., and Bradley, P.J. (2016). *In vivo* biotinylation of the *Toxoplasma* parasitophorous vacuole reveals novel dense granule proteins important for parasite growth and pathogenesis. mBio *7*, e00808-16.

Nagaraj, V.A., Sundaram, B., Varadarajan, N.M., Subramani, P.A., Kalappa, D.M., Ghosh, S.K., and Padmanaban, G. (2013). Malaria parasite-synthesized heme is essential in the mosquito and liver stages and complements host heme in the blood stages of infection. PLOS Pathog. *9*, e1003522.

Nkrumah, L.J., Muhle, R.A., Moura, P.A., Ghosh, P., Hatfull, G.F., Jacobs, W.R., Jr., and Fidock, D.A. (2006). Efficient site-specific integration in *Plasmodium falciparum* chromosomes mediated by mycobacteriophage Bxb1 integrase. Nat. Methods *3*, 615-621.

Parsons, M., Karnataki, A., Feagin, J.E., and DeRocher, A. (2007). Protein trafficking to the apicoplast: deciphering the apicomplexan solution to secondary endosymbiosis. Eukaryot. Cell *6*, 1081-1088.

Paytubi, S., Wang, X., Lam, Y.W., Izquierdo, L., Hunter, M.J., Jan, E., Hundal, H.S., and Proud, C.G. (2009). ABC50 promotes translation initiation in mammalian cells. J. Biol. Chem. *284*, 24061-24073.

Pei, Y., Tarun, A.S., Vaughan, A.M., Herman, R.W., Soliman, J.M., Erickson-Wayman, A., and Kappe, S.H. (2010). *Plasmodium* pyruvate dehydrogenase activity is only essential for the parasite's progression from liver infection to blood infection. Mol. Microbiol. *75*, 957-971.

Ralph, S.A., van Dooren, G.G., Waller, R.F., Crawford, M.J., Fraunholz, M.J., Foth, B.J., Tonkin, C.J., Roos, D.S., and McFadden, G.I. (2004). Tropical infectious diseases: metabolic maps and functions of the *Plasmodium falciparum* apicoplast. Nat. Rev. Microbiol. *2*, 203-216.

Rhee, H.W., Zou, P., Udeshi, N.D., Martell, J.D., Mootha, V.K., Carr, S.A., and Ting, A.Y. (2013). Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. Science *339*, 1328-1331.

Roux, K.J., Kim, D.I., Raida, M., and Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J. Cell Biol. *196*, 801-810.

Sayers, C.P., Mollard, V., Buchanan, H.D., McFadden, G.I., and Goodman, C.D. (2018). A genetic screen in rodent malaria parasites identifies five new apicoplast putative membrane transporters, one of which is essential in human malaria parasites. Cell. Microbiol. *20*, e12789.

Schnider, C.B., Bausch-Fluck, D., Bruhlmann, F., Heussler, V.T., and Burda, P.C. (2018). BioID reveals novel proteins of the *Plasmodium* parasitophorous vacuole membrane. mSphere *3*, e00522-17.

Sheiner, L., Demerly, J.L., Poulsen, N., Beatty, W.L., Lucas, O., Behnke, M.S., White, M.W., and Striepen, B. (2011). A systematic screen to discover and analyze apicoplast proteins identifies a conserved and essential protein import factor. PLOS Pathog. *7*, e1002392.

Sheiner, L., Vaidya, A.B., and McFadden, G.I. (2013). The metabolic roles of the endosymbiotic organelles of *Toxoplasma* and *Plasmodium* spp. Curr. Opin. Microbiol. *16*, 452-458.

Sidik, S.M., Huet, D., Ganesan, S.M., Huynh, M.H., Wang, T., Nasamu, A.S., Thiru, P., Saeij, J.P., Carruthers, V.B., Niles, J.C., *et al.* (2016). A genome-wide CRISPR screen in *Toxoplasma* identifies essential apicomplexan genes. Cell *166*, 1423-1435.

Spork, S., Hiss, J.A., Mandel, K., Sommer, M., Kooij, T.W., Chu, T., Schneider, G., Maier, U.G., and Przyborski, J.M. (2009). An unusual ERAD-like complex is targeted to the apicoplast of *Plasmodium falciparum*. Eukaryot. Cell *8*, 1134-1145.

Stanway, R.R., Witt, T., Zobiak, B., Aepfelbacher, M., and Heussler, V.T. (2009). GFP-targeting allows visualization of the apicoplast throughout the life cycle of live malaria parasites. Biol. Cell *101*, 415-430.

Tonkin, C.J., van Dooren, G.G., Spurck, T.P., Struck, N.S., Good, R.T., Handman, E., Cowman, A.F., and McFadden, G.I. (2004). Localization of organellar proteins in *Plasmodium falciparum* using a novel set of transfection vectors and a new immunofluorescence fixation method. Mol. Biochem. Parasitol. *137*, 13-21.

Tyzack, J.K., Wang, X., Belsham, G.J., and Proud, C.G. (2000). ABC50 interacts with eukaryotic initiation factor 2 and associates with the ribosome in an ATP-dependent manner. J. Biol. Chem. *275*, 34131-34139.

van Dooren, G.G., and Striepen, B. (2013). The algal past and parasite present of the apicoplast. Annu. Rev. Microbiol. *67*, 271-289.

van Dooren, G.G., Su, V., D'Ombrain, M.C., and McFadden, G.I. (2002). Processing of an apicoplast leader sequence in *Plasmodium falciparum* and the identification of a putative leader cleavage enzyme. J. Biol. Chem. *277*, 23612-23619.

Vaughan, A.M., O'Neill, M.T., Tarun, A.S., Camargo, N., Phuong, T.M., Aly, A.S., Cowman, A.F., and Kappe, S.H. (2009). Type II fatty acid synthesis is essential only for malaria parasite late liver stage development. Cell. Microbiol. *11*, 506-520.

Vazquez de Aldana, C.R., Marton, M.J., and Hinnebusch, A.G. (1995). GCN20, a novel ATP binding cassette protein, and GCN1 reside in a complex that mediates activation of the eIF-2 alpha kinase GCN2 in amino acid-starved cells. EMBO J. *14*, 3184-3199.

Wagner, J.C., Platt, R.J., Goldfless, S.J., Zhang, F., and Niles, J.C. (2014). Efficient CRISPR-Cas9-mediated genome editing in *Plasmodium falciparum*. Nat. Methods *11*, 915-918.

Waller, R.F., Keeling, P.J., Donald, R.G., Striepen, B., Handman, E., Lang-Unnasch, N., Cowman, A.F., Besra, G.S., Roos, D.S., and McFadden, G.I. (1998). Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. Proc. Natl. Acad. Sci. U.S.A. *95*, 12352-12357.

Waller, R.F., Reed, M.B., Cowman, A.F., and McFadden, G.I. (2000). Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway. EMBO J. *19*, 1794-1802.

Wu, W., Herrera, Z., Ebert, D., Baska, K., Cho, S.H., DeRisi, J.L., and Yeh, E. (2015). A chemical rescue screen identifies a *Plasmodium falciparum* apicoplast inhibitor targeting MEP isoprenoid precursor biosynthesis. Antimicrob. Agents Chemother. *59*, 356-364.

Yeh, E., and DeRisi, J.L. (2011). Chemical rescue of malaria parasites lacking an apicoplast defines organelle function in blood-stage *Plasmodium falciparum*. PLOS Biol. *9*, e1001138.

Yu, M., Kumar, T.R., Nkrumah, L.J., Coppi, A., Retzlaff, S., Li, C.D., Kelly, B.J., Moura, P.A., Lakshmanan, V., Freundlich, J.S., *et al.* (2008). The fatty acid biosynthesis enzyme FabI plays a key role in the development of liver-stage malarial parasites. Cell Host Microbe *4*, 567-578.

Zuegge, J., Ralph, S., Schmuker, M., McFadden, G.I., and Schneider, G. (2001). Deciphering apicoplast targeting signals--feature extraction from nuclear-encoded precursors of *Plasmodium falciparum apicoplast proteins*. Gene *280*, 19-26.