The Transcriptome of the Avian Malaria Parasite *Plasmodium ashfordi* Displays Host-Specific Gene Expression

Running title

The Transcriptome of *Plasmodium ashfordi*

Authors

Elin Videvall[1], Charlie K. Cornwallis[1], Dag Ahrén[1,3], Vaidas Palinauskas[2], Gediminas Valkiūnas[2], Olof Hellgren[1]

Affiliation

[1]Department of Biology, Lund University, Lund, Sweden

[2]Institute of Ecology, Nature Research Centre, Vilnius, Lithuania

[3]National Bioinformatics Infrastructure Sweden (NBIS), Lund University, Lund, Sweden

Corresponding authors

Elin Videvall (elin.videvall@biol.lu.se)

Olof Hellgren (olof.hellgren@biol.lu.se)

Abstract

Malaria parasites (*Plasmodium* spp.) include some of the world's most widespread and virulent pathogens, infecting a wide array of vertebrates. Our knowledge of the molecular mechanisms these parasites use to invade and exploit hosts other than mice and primates is, however, extremely limited. How do *Plasmodium* adapt to individual hosts and to the immune response of hosts throughout an infection? To better understand parasite plasticity, and identify genes that are conserved across the phylogeny, it is imperative that we characterize transcriptome-wide gene expression from non-model malaria parasites in multiple host individuals. Here, we used high-throughput Illumina RNA-sequencing on blood from wild-caught Eurasian siskins experimentally infected with a clonal strain of the avian malaria parasite, *Plasmodium ashfordi* (lineage GRW2). By using a multi-step approach to filter out all host transcripts, we successfully assembled the blood-stage transcriptome of *P. ashfordi*. A total of 11 954 expressed parasite transcripts were identified, and 7 860 were annotated with protein information. We further quantified gene expression levels of all parasite transcripts across three hosts during two infection stages – peak and decreasing parasitemia. Interestingly, parasites from the same host during different infection stages displayed remarkably similar expression profiles, but show large differences across hosts. This indicates that *P. ashfordi* adjusts its gene expression to specific host individuals, but contrary to expectation does not markedly change expression across different stages of infection. Finally, we examined genome-wide sequence similarity between *P. ashfordi* and other apicomplexan species, and searched for candidate genes involved in red blood cell invasion. The majority of transcripts were most similar to the human parasite *Plasmodium falciparum*, and a large number of invasion genes were discovered, suggesting conserved red blood cell invasion strategies between mammalian and avian *Plasmodium* spp. The transcriptome of *P. ashfordi* and its host-specific gene expression over two infection stages advances our understanding of *Plasmodium* plasticity and will become a valuable resource as it allows for further studies analyzing gene evolution and comparisons of parasite gene expression.

Keywords

Plasmodium, transcriptome, RNA-seq, malaria, gene expression, host-parasite interaction

Introduction

The apicomplexan parasites of the genus *Plasmodium* (malaria parasites) encompass a worldwide distribution and infect a multitude of vertebrate hosts, including reptiles, birds and mammals (Garnham, 1966). Their virulence can be highly variable between different strains and species. Some induce mild pathogenic effects on hosts and some cause severe disease, leading to high mortality rates (Palinauskas et al., 2008). Host individuals and host species also differ in their resistance and tolerance to malaria, and this interaction between host and parasite ultimately determines disease severity. Furthermore, the molecular response of hosts changes during the course of infection and creates a dynamic environment in which the parasites need to accommodate. Nevertheless, our understanding of how malaria parasites respond molecularly to different host individuals and to changes in the host immune defense over time is very limited.

Parasites of two clades in *Plasmodium* have been extensively studied from a molecular perspective, murine and primate parasites. We have learned a great deal of how malaria parasites of humans evolved and function by studying their rodent-infecting relatives. The majority of studies investigating gene expression in malaria parasites have been conducted using artificially selected cell lines (*in vitro*) or tissue cultures (*ex vivo*). This has provided tremendous insight into the biology of *Plasmodium* life stages (e.g. Bozdech et al., 2003; Hall et al., 2005; Otto et al., 2010; Siegel et al., 2014). However, major discrepancies in parasite expression between cultures and live animals (*in vivo*) have been documented (Lapp et al., 2015). A wide range of host environmental factors are absent in the *in vitro* systems, for example temperature fluctuations, inflammatory and immune effector molecules, hormones, metabolites, microenvironments, and varying levels of oxygen, pH, and glucose (LeRoux et al., 2009). Parasites cultured outside hosts reflect this with different expression patterns, and markedly downregulate important vaccine candidate genes such as cell surface antigens (Daily et al., 2005; Siau et al., 2008). The natural host environment, including genotypic and immunological cues, may therefore strongly affect the transcriptional responses of malaria parasites. To obtain representative transcriptional information from *Plasmodium* parasites, it is therefore valuable to utilize a system of natural host animals.

Alas, almost nothing is known about underlying molecular mechanisms in malaria parasites of hosts other than mice and primates. Which genes are conserved across *Plasmodium* and how do virulence, immune evasion, and host-specificity vary in species infecting non-mammalian animals? In order to investigate these questions, it will be necessary to assemble and characterize genome-wide expression information from malaria parasites throughout their phylogenetic host range. With the recent development of high-throughput sequencing techniques, it has now become possible to generate genomic sequences of non-model parasites (Martinsen and Perkins, 2013). Dual RNA-sequencing of both host and parasite opens up fantastic possibilities of simultaneously studying host-parasite interactions and describing transcriptome expression in both actors. Great care must be taken, however, when assembling parasite sequences *de novo*. Transcripts from the host and/or other sources of contamination may remain even after annotation and blast searches have taken place. Meticulous

filtering of assemblies using bioinformatics is therefore crucial to avoid erroneous conclusions (see e.g. Koutsovoulos et al. 2016). After constructing filtered high-quality parasite transcriptome data, we can start to investigate the aforementioned questions about *Plasmodium*.

In this study, we sequenced and assembled the blood transcriptome of the avian malaria parasite *Plasmodium ashfordi*. Malaria parasites infecting birds are highly suitable for studying transcriptional parasite responses due to their enormous diversity and large variation in host specificity and virulence (Bensch et al., 2004; Križanauskienė et al., 2006; Lachish et al., 2011). Some avian *Plasmodium* are extreme host generalists, successfully infecting birds over several orders, while other parasites are host specialists infecting one species (Drovetski et al., 2014; Pérez-Tris et al., 2007). The avian malaria system allows for the possibility of capturing wild birds in a natural setting, and evaluate their status of malaria infection. Additionally, infection experiments have in the last couple of years showed great potential to study *Plasmodium* in passerines under controlled conditions in laboratories (Cornet et al., 2014; Dimitrov et al., 2015; Ellis et al., 2015; Palinauskas et al., 2008, 2011; Zehtindjiev et al., 2008). We used our assembly of *P. ashfordi* to evaluate transcriptome characteristics, genome-wide sequence similarity, and investigate genes known to be involved in the *Plasmodium* red blood cell invasion process. We analyzed expression levels of parasite genes in three experimentally infected birds during both peak and decreasing parasitemia, two infection stages where the hosts exhibit different transcriptome responses (Videvall et al., 2015). This allowed us for the first time to follow and describe the dynamic transcriptome of an avian malaria parasite over time in individual hosts.

Results

The *Plasmodium ashfordi* transcriptome assembly

We sequenced blood collected from three experimentally infected Eurasian siskins (*Carduelis spinus*) during peak and decreasing parasitemia levels with Illumina dual RNA-sequencing (see Methods for details). The transcriptome of *P. ashfordi* was assembled into two versions in order to make it transparent and as useful as possible for other researchers. The first assembly version, which we refer to as the annotated assembly, contains the transcripts with annotation information from apicomplexa proteins (n = 7 860) (Figure 1; Supplementary Table S1). The second version that we refer to as the total assembly, contains the unannotated contigs (n = 4 094) that were strictly filtered to remove contigs from the avian host (Figure 1) together with the annotated contigs, resulting in a total of 11 954 representative transcripts (Table 1).

The size of the total transcriptome assembly is 9.0 Mbp and the annotated version of the assembly comprise 7.3 Mbp (81.20%) (Table 1). We calculated assembly statistics using the transcriptome-specific measurement E90N50, which represents the contig N50 value based on the set of transcripts representing 90% of the expression data, and is preferable over the original N50 measurement when evaluating transcriptome assemblies (Haas, 2016). The assembly E90N50 is 1 988 bp and the mean transcript length of the annotated transcripts is 930.8 bp (Table 1, Supplementary Figure S1). In comparison, the length of coding sequences in the *Haemoproteus tartakovskyi* (sister parasite to *Plasmodium*) genome is mean = 1 206 and median = 1 809 bp (Bensch et al., 2016). The *P. falciparum* transcriptome contains transcripts of median length 1 320 and mean length 2 197 bp (Gardner et al., 2002). The longest contig in the *P. ashfordi* assembly, consisting of 26 773 bp, is transcribed from the extremely long ubiquitin transferase gene (AK88_05171), which has a transcript length around 27 400 bp in other *Plasmodium* species. The annotated transcriptome has an exceptionally low mean GC content of 21.22% (Figure 1B), which is even lower than the already GC-biased transcriptome of *P. falciparum* (23.80%).

Biological functions of genes expressed in *P. ashfordi*

To evaluate biological and molecular functions of all expressed transcripts in *P. ashfordi*, we analysed gene ontology annotation. Genes were found to primarily belong to the two major molecular groups: binding and catalytic activity, as well as the biological functions: metabolic and cellular processes (Supplementary Figure S5). To further investigate the metabolic processes of *P. ashfordi* in detail, a subsequent analysis of metabolic pathway enrichment compared to *P. falciparum* yielded 19 significant pathways after correcting for multiple testing with the Benjamini and Hochberg false discovery rate (q-value < 0.1). Glycolysis, purine metabolism, and methane metabolism were among the top significant metabolic pathways (Figure 2). An investigation of annotated genes belonging to the gene ontology category “kinase activity” resulted in a total of 95 genes (Supplementary Table S2).

Gene expression is similar across different stages of infection

The *P. ashfordi* transcripts were next analyzed for expression levels within individual hosts across the two parasitemia stages. We accounted for differences in parasitemia levels between hosts and time points, and any variation in sequencing depth between samples, by normalizing overall expression

values according to the DESeq method (Anders and Huber, 2010). We found that the parasites displayed very similar gene expression patterns during peak and decreasing parasitemia stages (Figure 3). No genes were significantly differentially expressed between the time points (q-value > 0.99), and the correlation in gene expression was extremely high (Pearson's product-moment correlation = 0.9983, t = 1 905.2, df = 11 952, p-value < 2.2e-16) (Figure 3A; Supplementary Table S3). Annotated transcripts showing the highest expression fold change (non-significantly) between the two parasitemia stages were derived from the following genes (in order of most observed change): rho-GTPase-activating protein 1, 40S ribosomal protein S3a, two uncharacterized proteins, TATA-box-binding protein, heat shock protein 90, and C50 peptidase (Supplementary Table S1, Supplementary Figure S2).

Gene expression is host-specific

In contrast to the similarities in gene expression between parasitemia stages, the parasite transcriptomes showed much larger differences in expression levels between the different host individuals. A principal component analysis of expression variances clustered parasite samples together within their respective hosts, which showed major similarities in expression profiles (Figure 3C). Samples derived from the same host individual did not separate until the third (15% variance) and fourth (13% variance) principal component dimensions (Figure 3D). The parasite transcriptome from host 4 during decreasing parasitemia showed the largest variation in parasite gene expression among all samples, yet it was still most similar to the transcriptome from the same host during peak parasitemia (Figure 3B; Figure 3C). In fact, all parasite transcriptomes during the decreasing parasitemia stage demonstrated closest distance to the transcriptome sample within the same host ten days earlier (Figure 3B).

To further evaluate if specific transcripts contributed to differences in parasite gene expression levels between individual hosts, we performed a likelihood ratio test over all host individuals while controlling for parasitemia stage. This resulted in 28 significant *P. ashfordi* transcripts (q-value < 0.1) displaying very high expression variation between hosts (Figure 4; Supplementary Table S4). The most significant transcripts were derived from the genes (in order of most significance) cytochrome c oxidase subunit 1, 70 kd heat shock-like protein, M1 family aminopeptidase, and metabolite/drug transporter (Supplementary Table S4).

Transcriptome sequence similarities to other apicomplexan parasites

Almost all annotated contigs (99.59%; n = 7 828) resulted in a best blast hit against a species within the genus *Plasmodium* (Figure 5A). The remaining contigs had matches against species within the genera of *Eimeria* (n = 12), *Cryptosporidium* (n = 6), *Neospora* (n = 5), *Babesia* (n = 4), *Hammondia* (n = 2), *Ascogregarina* (n = 1), *Theileria* (n = 1), and *Toxoplasma* (n = 1) (Supplementary Table S5). The great majority (73.59%) of the contig blast matches were proteins originating from primate parasites, while 25.34% matched rodent parasites, and only 0.92% parasites of birds (Figure 5B).

At the species level, most contigs (29.91%) resulted in best blast hit against *P. falciparum*, followed by *P. reichenowi* (16.88%) and *P. yoelii* (8.59%) (Figure 5C). The significant blast matches to bird parasites consisted of the species *Plasmodium gallinaceum* (n = 56), *Eimeria acervulina* (n = 5), *Eimeria tenella* (n = 4), *Eimeria mitis* (n = 3), *Plasmodium relictum* (n = 3), and *Plasmodium lutzi*

(n = 1). The contigs giving matches to avian *Plasmodium* were primarily derived from commonly sequenced apicomplexa genes and therefore available in public databases, for example cytochrome c oxidase subunit 1 (COX1; *P. lutzi*), merozoite surface protein 1 (MSP1; *P. relictum*), thrombospondin related anonymous protein (TRAP; *P. relictum*), and cytochrome b (CYTB; *P. gallinaceum*) (Supplementary Table S1).

The five contigs with highest GC content in the *P. ashfordi* transcriptome (47.7% – 56.4%) all had matches against the avian parasites *Eimeria*, despite them only comprising 0.15% (n = 12) of the total annotation. *Eimeria* spp. have a very high transcriptome GC content (*E. acervulina*: 55.98%; *E. mitis*: 57.30%), and the *P. ashfordi* transcripts matching this genus consist mostly of ribosomal and transporter genes (Supplementary Table S1).

The *P. ashfordi* contigs with highest expression levels were primarily annotated by uncharacterized protein matches to the rodent parasite *P. yoelii* (Table 2). In fact, the six most highly expressed transcripts that were annotated, all gave significant blast matches to *P. yoelii*. Further investigation revealed that these transcripts are most likely derived from ribosomal RNA.

Identification of conserved *Plasmodium* invasion genes

Finally, to assess molecularly conserved strategies in *P. ashfordi* compared to mammalian malaria parasites, we searched for annotated genes known to be involved in the red blood cell invasion by *Plasmodium* spp. (Beeson et al., 2016; Bozdech et al., 2003). We discovered successfully assembled *P. ashfordi* transcripts from a whole suite of host cell invasion genes (Table 3). This includes for example the genes merozoite surface protein 1 (MSP1), apical membrane antigen 1 (AMA1), merozoite adhesive erythrocytic binding protein (MAEBL), GPI-anchored micronemal antigen (GAMA), and the rhoptry neck binding proteins 2, 4, and 5 (RON2, RON4, and RON5). Interestingly, the *P. ashfordi* RON genes in particular seemed to slightly decrease expression levels in all hosts over the two time points (Figure 6). In general, however, the invasion genes showed a range of expression patterns over time, going in various directions (Figure 6).

All genes known to be involved in the *Plasmodium* motor complex (Baum et al., 2006; Opitz and Soldati, 2002), driving parasite gliding motion and enabling host cell invasion, were discovered in *P. ashfordi*. These include: actin (ACT1), actin-like protein (ALP1), aldolase (FBPA), myosin A (MyoA), myosin A tail interacting protein (MTIP), glideosome-associated protein 45 and 50 (GAP45 and GAP50), and thrombospondin related anonymous protein (TRAP). We also found the bromodomain protein 1 (BDP1), which has been directly linked to erythrocyte invasion by binding to chromatin at transcriptional start sites of invasion-related genes and controlling their expression (Josling et al., 2015).

We found two transcripts matching the low molecular weight rhoptry-associated proteins 1 and 3 (RAP1 and RAP3) that are secreted from the rhoptry organelles during cell invasion. The genomes of human malaria parasites contain a paralog gene called RAP2 as well, whereas rodent malaria parasites contain a single gene copy that is a chimera of RAP2 and RAP3 (RAP2/3) (Counihan et al., 2013). The *P. ashfordi* transcript in question (TR13305|c0_g1_i1) matches *P. falciparum* RAP3 better than the rodent parasite version of RAP2/3. The three high molecular weight rhoptry proteins (RhopH1, RhopH2, RhopH3) which bind to the erythrocyte plasma membrane and transfer to the parasitophorous vacuole membrane upon invasion (Counihan et al., 2013; Vincensini et

al., 2008) were all identified in *P. ashfordi*. RhopH1 encompasses the multigene family of cytoadherence linked asexual proteins (CLAGs), present in varying copy number across *Plasmodium*.

Other assembled *P. ashfordi* orthologs of genes involved in host cell invasion were the rhoptry-associated leucine zipper-like protein 1 (RALP1), rhoptry-associated membrane antigen (RAMA), armadillo-domain containing rhoptry protein (ARO), RH5 interacting protein (RIPR), TRAP-like protein (TLP), merozoite TRAP-like protein (MTRAP), thrombospondin related apical membrane protein (TRAMP), subtilisin proteases 1 and 2, (SUB1 and SUB2), and merozoite surface proteins 8 and 9 (MSP8 and MSP9). MTRAP and TRAMP are proteins that belong to the TRAP-family and are released from the microneme organelles during invasion (Cowman et al., 2012; Green et al., 2006), and the subtilisin proteases SUB1 and SUB2 are heavily involved in the processing and cleavage of immature merozoite antigens, for example MSP1 and AMA1 (Beeson et al., 2016). ARO plays a crucial role in positioning the rhoptry organelles within the apical end of the parasite to enable the release of rhoptry-associated molecules (Mueller et al., 2013) such as RAMA and RALP1, which then bind to the erythrocyte surface.

We furthermore discovered transcripts of several reticulocyte binding proteins (RBP / RH) thought to be absent in the genomes of avian malaria parasites (Lauron et al., 2015). These particular transcripts, together with RAMA, interestingly showed much higher e-values than other invasion genes (Table 3), indicating high differentiation between avian and mammalian *Plasmodium* RH genes. Finally, two rhomboid proteases (ROM1 and ROM4) have been linked to host cell invasion in *P. falciparum* via cleavage of transmembrane adhesins (Baker et al., 2006; Santos et al., 2012). We found both of these genes, together with other rhomboid proteases (ROM2, ROM3, ROM6, ROM8, and ROM10) expressed in *P. ashfordi*. More information about the assembled genes can be found in Supplementary Table S1.

Discussion

Generating and describing genome-wide expression data of parasites from phylogenetically diverse hosts is crucial to better understand the evolution of host-parasite interactions, virulence, and host-specificity. In this study, we assembled and characterized the first transcriptome with quantified gene expression of an avian malaria parasite, *P. ashfordi*. By developing a bioinformatic pipeline capable of dealing with dual RNA-seq data, we successfully filtered away contigs originating from the host and other sources of contamination in a multistep approach. This resulted in a transcriptome with 7 860 annotated transcripts, and an additional 4 094 species-specific unannotated transcripts. Parasite gene expression displayed strikingly similar patterns during two infection stages and within individual hosts. Furthermore, *P. ashfordi* shows large similarities to the human malaria parasite, *P. falciparum*, and the assembly supports several important erythrocyte invasion genes (Table 3) indicating evolutionary conserved invasion strategies across the phylogenetic host range of *Plasmodium* parasites.

P. ashfordi displays host-specific gene expression

Interestingly, and contrary to our expectations, *P. ashfordi* showed exceptionally similar expression profiles inside the same host, despite being sampled ten days apart during two different disease stages. All birds were inoculated with the same clonal malaria strain, which means the parasite probably regulates its gene expression levels to fit the different hosts, although direct modulation of parasite gene expression by the host cannot be excluded. The mechanism behind this host-specific expression pattern is unknown, but could be due to genotype by genotype interactions between the host and the parasite, or plasticity of the parasite to varying host environments. This result has potentially important implications for our understanding of the evolution of host-parasite interactions, and as a result warrants further research extending the limited sample size to more hosts and more timepoints throughout the infection.

Host genotype by parasite genotype interactions are not well documented in malaria parasite systems. Studies with different genotypes of both host and parasite have found effects of host genotype, but not parasite genotype, on factors such as host resistance and parasite virulence (Mackinnon et al., 2002; de Roode et al., 2004; Grech et al., 2006; see also Idaghdour et al., 2012). Even less is known about the transcriptome responses of malaria parasites to different hosts. Daily et al. (2007) discovered host-specific distinct transcriptional states of *P. falciparum* in the blood of Senegalese children, where parasite transcriptomes could be divided into three distinct clusters associated with either starvation responses and invasion, glycolysis, or oxidative stress and heat shock pathways (Daily et al., 2007).

The expression profiles of *P. ashfordi* did not exhibit any significant differences between peak and decreasing parasitemia stages (day 21 and 31 postinfection). The hosts in our study experienced relatively high parasitemia levels also during the decreasing parasitemia stage (see Methods), so it is possible that these specific time points do not provide very different environmental host cues to the parasites. However, the simultaneous transcriptomes of the avian hosts (analyzed in Videvall et al. 2015) displayed large differences in gene expression between these two parasitemia stages, notably with a reduced immune response during decreasing parasitemia. This is important

because it appears that *P. ashfordi* does not adjust its gene expression in response to the decreasing immune levels of the hosts, but instead conforms to the specific environment of individual hosts.

P. falciparum evades the human immune defense via intracellularity, clonal antigenic variation (Guizetti and Scherf, 2013), transcriptional antigenic switches (Recker et al., 2011), splenic avoidance by erythrocytic adherence to the endothelium (Craig and Scherf, 2001) and sequestration in organ microvasculature (Ashley et al., 2014), erythrocytic rosetting (Niang et al., 2014), host immunosuppression (Hisaeda et al., 2004), and manipulation of host gene expression. It is possible that *P. ashfordi* shares several of these evasion strategies with *P. falciparum*, although this remains unknown. One example of immune evasion by manipulation of host gene expression is the parasite gene macrophage migration inhibitory factor homologue (MIF), which contributes to *Plasmodium* parasites' ability to modulate the host immune response by molecular mimicry (Cordery et al., 2007). This gene was discovered transcribed in *P. ashfordi* as well (TR2046|c0_g1_i1), suggesting a similar immune manipulation strategy.

Similarities to *P. falciparum* and other malaria parasites

The majority of all annotated contigs (73.59%) resulted in a best blast hit against primate parasite species (Figure 5A). Curiously, the human malaria parasite *P. falciparum* comprised the majority of all matches with almost a third of the transcripts (29.91%) (Figure 5C). This is likely because *P. falciparum* currently constitutes the organism with most sequence similarities to *P. ashfordi* based on publically available sequences (Bensch et al., 2016). The chimpanzee parasite *P. reichenowi* had the second most blast matches to *P. ashfordi* (Figure 5C), and it is the closest living relative of *P. falciparum* as far as we know (Otto et al., 2014). Furthermore, both parasites share the genome characteristics of being extremely AT-biased, with *P. ashfordi* reaching a remarkably low transcriptomic GC content of 21.22% (Table 1) compared to the already extremely AT-rich *P. falciparum*, which has a transcriptomic GC content of 23.80%. Lastly, because of its role in human disease, *P. falciparum* is the most sequenced *Plasmodium* species (172 official strains in the NCBI taxonomy database as of May 2016) (Gardner et al., 2002), resulting in plenty of opportunities for transcript sequences to find significant blast matches.

Less than one percent of all contigs resulted in a blast hit against avian parasites (0.92%). This is due to the fact that almost no genomic resources are available for avian *Plasmodium*. Despite their enormous diversity, world-wide distribution, and harmful effects on susceptible bird populations, genomic studies of avian malaria parasites have been largely non-existent until now. The genome of *P. gallinaceum*, the malaria parasite of chickens, has been sequenced but not published and is not yet available (Ulrike Boehme, personal communication). A transcriptome assembly of *P. gallinaceum* is available for download (Lauron et al., 2014, 2015), although still contains a large proportion of contigs with significant blast matches to birds, making comparisons with *P. ashfordi* difficult (Supplementary Figure S4). Dual RNA-sequencing of a more distantly related apicomplexan parasite, *Leucocytozoon buteonis*, and its buzzard host was recently described (Pauli et al., 2015), though no publically available transcriptome exists. Finally, 454 RNA-sequencing of the generalist avian malaria parasite *P. relictum* (lineage SGS1) has been previously performed (Bensch et al., 2014; Hellgren et al., 2013), but the low sequencing coverage does not allow for assembly nor any transcriptome

analyses. We hope that future versions of these avian parasite transcriptomes will enable genome-wide comparisons.

The lack of genome-wide sequence data from malaria parasites of hosts other than mice and primates, means that little is known about which genes across *Plasmodium* are conserved and which are unique. As a first step to investigate this, we searched in the *P. ashfordi* assembly for genes known to be involved in the merozoite invasion of host red blood cells. This particular group of genes has shown the greatest diversity in mammalian *Plasmodium* and many are under strong positive selection due to their role in host specificity (Hall et al., 2005; Otto et al., 2014). Previously, only a handful of studies have sequenced candidate invasion genes in avian malaria parasites; these include MAEBL (Martinez et al., 2013), AMA1 and RON2 (Lauron et al., 2014), MSP1 (Hellgren et al., 2013, 2015), RIPR (Lauron et al., 2015), and TRAP (Farias et al., 2012; Templeton and Kaslow, 1997). Due to the evolutionary distance between mammals and saurians, and their inherent blood cell differences (birds/reptiles have erythrocytes with a nucleus and mitochondria while mammalian cells are anucleated), we might expect to find few and highly differentiated gene orthologs. Remarkably, we discovered a large number of red blood cell invasion genes expressed in *P. ashfordi* (Table 3), indicating that most of these specific invasion genes are conserved across both mammalian and avian *Plasmodium*.

The invasion genes that were most differentiated between birds and mammals were the rhoptry-associated membrane antigen (RAMA) and the reticulocyte binding proteins (RBP/RH), which had diverged almost beyond recognition. These RH genes, together with other erythrocyte binding antigens (EBA), are missing in *P. vivax*, and have understandably been assumed absent in the genomes of avian malaria parasites (Lauron et al., 2015). However, our result suggests that several are not only present, but also transcribed, though with high sequence divergence. It is possible that additional erythrocyte binding proteins are present in the *P. ashfordi* assembly, though ortholog searches for these genes will become complicated if they have evolved under especially strong selection pressure in avian *Plasmodium*. Genes that are orthologous between *P. falciparum* and avian malaria parasites but absent from other mammalian parasites are important genomic markers of maintained ancestral states (Bensch et al., 2016). We can therefore advance our understanding about the evolution of *Plasmodium* species by analyzing genomes and transcriptomes of avian malaria parasites.

Conclusion

In this study we have sequenced, assembled, and characterized the transcriptome of *Plasmodium ashfordi*. The results presented here and the associated assembly will help improve our understanding of host-parasite interactions, evolutionary conserved *Plasmodium* strategies, and the phylogenetic relationships between apicomplexans. In addition, we have shown that *P. ashfordi* displays strikingly similar expression profiles within individual hosts during two different stages of the infection – but different expression patterns between individual hosts – indicating host specific parasite gene regulation. The expression information of all transcripts in the *P. ashfordi* transcriptome will ultimately assist researchers studying genes involved in e.g. immune evasion, host-specificity, and parasite plasticity.

399

400

Methods

Experimental setup

We used four wild-caught juvenile Eurasian siskins (*Carduelis spinus*) in an infection experiment. The procedure was carried out 2012 at the Biological Station of the Zoological Institute of the Russian Academy of Sciences on the Curonian Spit in the Baltic Sea. All details regarding the experimental setup have been outlined in Videvall et al. (2015). Briefly, three of the birds were inoculated with a single injection of blood containing the erythrocytic stages of *Plasmodium ashfordi*, subgenus *Novyella*, lineage GRW2. For a description of this parasite, see Valkiūnas et al. (2007). A control bird (bird 1) was simultaneously inoculated with uninfected blood for the evaluation of host transcription (Videvall et al., 2015). Blood samples for RNA-sequencing were taken from birds before infection (day 0), during peak parasitemia (day 21 postinfection) and during the decreasing parasitemia stage (day 31 postinfection).

All birds were thoroughly screened with both microscopic (Palinauskas et al., 2008) and molecular (Hellgren et al., 2004) methods before the experiment to make sure they had no prior haemosporidian infection. The parasitemia intensity varied substantially in infected birds, with bird 2, bird 3, and bird 4 having 24.0%, 48.0%, and 71.3% of their red blood cells infected during peak parasitemia, and later 8.2%, 21.8%, and 33.3%, respectively, during the decreasing parasitemia stage (Videvall et al., 2015). The two parasitemia stages are sometimes referred to as different 'infection stages' in the hosts, but we like to clarify that there is no evidence present suggesting that the parasites have entered a different stage of their life cycle (e.g. tissue merogony). Experimental procedures were approved by the International Research Co-operation Agreement between the Biological Station Rybachy of the Zoological Institute of the Russian Academy of Sciences and Institute of Ecology of Nature Research Centre (25 May 2010). Substantial efforts were made to minimize handling time and potential suffering of birds.

RNA extraction and sequencing

From six infected samples (three treatment birds at days 21 and 31) and six uninfected samples (three treatment birds at day 0 and one control bird at days 0, 21, and 31) total RNA was extracted from 20 µl whole blood. Detailed extraction procedures can be found in Videvall et al. (2015). Total extracted RNA was sent to Beijing Genomics Institute (BGI), China, for RNA quality control, DNAse treatment, rRNA depletion, and amplification using the SMARTer Ultra Low kit (Clontech Laboratories, Inc.). BGI performed library preparation, cDNA synthesis, and paired-end RNA-sequencing using Illumina HiSeq 2000. The blood samples from bird 3 and bird 4 during peak parasitemia were sequenced by BGI an additional time in order to generate more reads from the parasite in preparation for this transcriptome assembly. These resequenced samples were regarded and handled as technical replicates. We quality-screened all the demultiplexed RNA-seq reads using FastQC (v. 0.10.1) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

De novo transcriptome assembly

Quality-filtered RNA-seq reads from all six infected bird samples together with the two re-sequenced samples were used in a *de novo* assembly. This was performed using the transcriptome assembler

Trinity (v. 2.0.6) (Grabherr et al., 2011) with 489 million reads. Mapping of reads to available genomes of human malaria parasites were employed simultaneously, but unfortunately these attempts yielded few hits due to the evolutionary distance between avian *Plasmodium* and human *Plasmodium*, so we continued annotating the assembly using blast searches.

The assembled transcripts were blasted against the NCBI non-redundant protein database using the program DIAMOND BLASTX (v. 0.7.9) (Altschul et al., 1990; Buchfink et al., 2014) with sensitive alignment mode. A total of 47 823 contigs generated significant hits against avian species (Figure 1C). A large number of contigs (n = 260 162) did not produce any significant (e-value < 1e-5) blastx hits (Figure 1D). This is because 1) the host species is a non-model organism without a genome available, leading to a large number of host contigs without blast hits, 2) multiple contigs do not necessarily contain coding sequences, but can be derived from noncoding RNAs, etc. and therefore will not match protein sequences, 3) short, fragmented contigs may not yield sufficient blast hit significance, and 4) an extreme underrepresentation of protein sequences from avian *Plasmodium* species in the NCBI nr database will not result in any significant blast hits to genes unique in avian malaria parasites.

We strictly filtered the initial assembly by only retaining a total of 9 015 transcripts (isoforms) that produced significant blast matches against proteins from species in the Apicomplexa phylum. A previous assembly using an earlier version of Trinity (v. r20140413p1) (Grabherr et al., 2011) performed better when it came to assembling the longest contigs (> 6 kbp). Different versions of assembler software may construct de Bruijn graphs somewhat differently, which is why it can be a good idea to make several assemblies and later combine parts of them (Brian Haas, personal communication). The previous assembly had been blasted and screened for apicomplexa in the same way as described above. In order not to lose these important full-length transcripts, we therefore included the longest contigs from the previous assembly that had 1) not assembled correctly in the current assembly, and 2) significant blastx hits against apicomplexa (n = 10), resulting in a total of 9 025 transcripts. The fact that these contigs contained similar sequences already present in the assembly was dealt with through downstream clustering of the sequences.

Transcriptome cleaning and filtering

Some contigs in the annotated assembly contained poly-A tails which complicated downstream analyses and resulted in biased mapping estimates. We therefore removed all poly-A/T tails using PRINSEQ (v. 0.20.4) (Schmieder and Edwards, 2011) with a minimum prerequisite of 12 continuous A's or T's in the 5' or 3' end of the contigs. A total of 106 202 bases (1.18%) was trimmed and the mean transcript length was reduced from 995.28 to 983.52 bases.

The unknown transcripts that failed to produce significant hits to any organism during the blastx run (n = 260 162) were subsequently cleaned using the following procedure. First, we trimmed them for poly-A tails, resulting in a total of 455 331 bases removed, and a slight decrease of the mean length of the unknown sequences from 555.27 nt before trimming to 553.52 nt after trimming. The majority of these unknown transcripts came from host mRNA, but their GC content displayed a clear bimodal distribution (Figure 1D), where the contigs with very low GC were strongly suspected to originate from the parasite. To avoid any host contigs, we strictly filtered the unknown transcripts to only include sequences with a mean GC content lower than 23% (n = 4 624). This threshold was

485 based on the apicomplexa-matching transcripts which had a mean GC content of 21.22%, and all
486 poly-A trimmed contigs giving significant blast matches to birds (class: Aves) (n = 47 823), where the
487 mean GC content was 47.65%, and the avian contig with the absolute lowest GC content had 23.48%
488 GC (Figure 1C).
489
490 **Transcriptome clustering, further filtering, and validation**
491 To reduce redundancy of multiple isoforms and transcripts derived from the same gene, we first
492 merged together the annotated and the unknown transcripts with a GC content < 23% (n = 13 649).
493 We then clustered these sequences together in order to retain most transcripts but group the highly
494 similar ones based on 97% sequence similarity and a k-mer size of 10, using CD-HIT-EST (v. 4.6) (Li
495 and Godzik, 2006). The most representative (longest) contig in every cluster was selected to represent
496 those transcripts, resulting in 12 266 contigs/clusters. We further filtered all the short sequences (<
497 200 bases), to obtain a set of 12 182 representative transcripts.
498 A second blast filtering step with the trimmed representative contigs against the Refseq
499 genomic database was then employed using BLASTN+ (v. 2.2.29) (Altschul et al., 1990; Camacho et
500 al., 2009) to identify some ambiguous contigs suspected to contain non-coding RNA bird sequences.
501 We removed all contigs that gave significant matches (e-value: 1e-6) against all animals (kingdom
502 Metazoa), so we could be confident that the assembly only consisted of true parasite transcripts. This
503 last filtering step removed 228 contigs.
504 The unannotated transcripts (n = 4 094) of the final assembly were further validated to
505 originate from the parasite by using reads from the six uninfected samples of the same hosts sampled
506 before infection and the control bird. A total of 350 318 482 reads (65 bp and 90 bp) from all
507 uninfected samples were mapped to the unannotated transcripts using Bowtie2 (v. 2.2.5) (Langmead
508 and Salzberg, 2012), resulting in the alignment of 90 read pairs (0.000051%). This extremely low
509 mapping percentage from the uninfected samples greatly supported our conclusion that these
510 transcripts had indeed been transcribed by *P. ashfordi*. These 4 094 representative transcripts with
511 unknown function are referred to throughout the paper as unannotated transcripts.
512 The resulting final transcriptome assembly consists of 11 954 representative annotated and
513 unannotated transcripts. The genomes of *Plasmodium* parasites generally contain around 5-6 000
514 protein-coding genes (Kersey et al., 2015), making it reasonable to assume similar gene numbers for
515 the *P. ashfordi* genome. The larger number of transcripts in the assembly is therefore a result of
516 isoform varieties and fragmented contigs.
517
518 **Estimating expression levels**
519 Poly-A tails of 489 million RNA-seq reads from all samples were trimmed as well using PRINSEQ
520 (v. 0.20.4) (Schmieder and Edwards, 2011). A minimum prerequisite for trimming were 20
521 continuous A's or T's in the 5' or 3' end of each read. Only trimmed reads longer than 40 bp and still
522 in a pair were retained (n = 451 684 626) in order to confidently map high quality reads with good
523 minimum lengths. Bowtie2 (v. 2.2.5) (Langmead and Salzberg, 2012) was used to map the trimmed
524 RNA-seq reads of every sample (n = 8) (six biological and two technical replicates) back to the *P.*
525 *ashfordi* transcriptome consisting of the 11 954 representative sequences. We calculated expression

levels using RSEM (v. 1.2.21) (Li and Dewey, 2011), which produces expected read counts for every contig.

The counts of the 11 954 transcripts were subsequently analyzed inside the R statistical environment (v. 3.2.5) (R Core Team, 2015). We tested for expression differences in the malaria parasites between the two time points, peak and decreasing parasitemia, using DESeq2 (v. 1.10.1) (Love et al., 2014). The two resequenced samples (technical replicates) of bird 3 and bird 4 during peak parasitemia were handled exactly the same as all other samples, and their respective read count were added to their biological samples, according to the DESeq2 manual. Individual host was used as a factor in the analysis, and counts were normalized to account for potential variation in sequencing depth as well as the large differences in number of parasites present in the blood (parasitemia levels). Regularized log transformation of counts was performed in order to represent the data without any prior knowledge of sampling design in the principal component analysis and sample distance calculations. This way of presenting counts without bias is preferred over variance stabilizing of counts when normalization size factors varies greatly between the samples, as they naturally do in our data.

Transcriptome evaluation

Transcriptome statistics such as GC content, contig length, and assembled bases were calculated using bash scripts and in the R statistical environment (v. 3.2.5) (Pages et al., 2015; R Core Team, 2015). P-values were corrected for multiple testing with the Benjamini and Hochberg false discovery rate (Benjamini and Hochberg, 1995) and corrected values have been labeled as q-values throughout. We calculated the GC content of two *Eimeria* transcriptomes downloaded from ToxoDB (v. 25) (Gajria et al., 2007), initially sequenced by Reid et al. (2014). Transcriptome E90N50 was calculated using RSEM (v. 1.2.21) and Trinity (v. 2.0.6) (Grabherr et al., 2011; Haas, 2016; Li and Dewey, 2011). Plots were made with the R package ggplot2 (v. 2.1.0) (Wickham, 2009). Metabolic pathway and gene ontology enrichment analyses were conducted in PlasmoDB (Aurrecoechea et al., 2009) using orthologs of annotated gene names to compare against the gene ontology and metabolic pathway background information from *P. falciparum* 3D7. The red blood cell invasion genes were searched for in the transcriptome annotation of *P. ashfordi* (Supplementary Table S1). Only genes with documented involvement in *Plasmodium* erythrocyte invasion were included.

Data availability

The sequence reads of both host and parasite have been deposited at the NCBI Sequence Read Archive under the accession number PRJNA311546. The assembled *P. ashfordi* transcripts are available for download at http://mbio-serv2.mbioekol.lu.se/Malavi/Downloads.

Author Contributions

The study design was initially conceived by OH, VP and GV, and further developed together with EV and CKC. VP performed the experiments. OH performed the RNA extractions. Assembly and bioinformatic and statistical analyses were performed by EV. DA advised in the trimming step of the assembly and in the mapping of sequence reads. OH, CKC, and EV planned the paper. EV drafted the paper with extensive input from all authors.

Funding

This work was supported by the Swedish Research Council (grant 621-2011-3548 to OH and 2010-5641 to CKC), the Crafoord Foundation (grant 20120630 to OH), European Social Fund under the global grant measure (grant VPI-3.1.-ŠMM-07-K-01-047 to GV), Research Council of Lithuania (grant MIP-045/2015 to GV) and a Wallenberg Academy Fellowship to CKC.

Acknowledgments

We would like to thank Staffan Bensch for stimulating discussions and comments on this paper. We are also grateful to Brian Haas for advice on transcriptome assemblies, and the director of the Biological Station “Rybachy”, Casimir V. Bolshakov, for generously providing facilities for the experimental research. The assembly and blastn computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) (Lampa et al., 2013) under project b2014134; all other computational analyses were performed by EV on a local machine.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi:10.1016/S0022-2836(05)80360-2.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11, R106. doi:10.1186/gb-2010-11-10-r106.

Ashley, E. a., Dhorda, M., Fairhurst, R. M., Amaratunga, C., Lim, P., Suon, S., et al. (2014). Spread of Artemisinin Resistance in *Plasmodium falciparum* Malaria. *N. Engl. J. Med.* 371, 411–423. doi:10.1056/NEJMoa1314981.

Aurrecoechea, C., Brestelli, J., Brunk, B. P., Dommer, J., Fischer, S., Gajria, B., et al. (2009). PlasmoDB: A functional genomic database for malaria parasites. *Nucleic Acids Res.* 37, 539–543. doi:10.1093/nar/gkn814.

Baker, R. P., Wijetilaka, R., and Urban, S. (2006). Two Plasmodium rhomboid proteases preferentially cleave different adhesins implicated in all invasive stages of malaria. *PLoS Pathog.* 2, 0922–0932. doi:10.1371/journal.ppat.0020113.

Baum, J., Richard, D., Healer, J., Rug, M., Krnajski, Z., Gilberger, T. W., et al. (2006). A conserved molecular motor drives cell invasion and gliding motility across malaria life cycle stages and other apicomplexan parasites. *J. Biol. Chem.* 281, 5197–5208. doi:10.1074/jbc.M509807200.

Beeson, J. G., Drew, D. R., Boyle, M. J., Feng, G., Fowkes, F. J. I., and Richards, J. S. (2016). Merozoite surface proteins in red blood cell invasion, immunity and vaccines against malaria. *FEMS Microbiol. Rev.*, fuw001. doi:10.1093/femsre/fuw001.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* 57, 289–300.

Bensch, S., Canbäck, B., DeBarry, J. D., Johansson, T., Hellgren, O., Kissinger, J. C., et al. (2016). The Genome of *Haemoproteus tartakovskyi* and Its Relationship to Human Malaria Parasites. *Genome Biol. Evol.* 8, 1361–1373. doi:10.1093/gbe/evw081.

Bensch, S., Coltman, D. W., Davis, C. S., Hellgren, O., Johansson, T., Malenfant, R. M., et al. (2014). Genomic Resources Notes accepted 1 June 2013–31 July 2013. *Mol. Ecol. Resour.* 14, 218. doi:10.1111/1755-0998.12166.

Bensch, S., Pérez-Tris, J., Waldenström, J., and Hellgren, O. (2004). Linkage between nuclear and mitochondrial DNA sequences in avian malaria parasites: multiple cases of cryptic speciation? *Evolution (N. Y).* 58, 1617–1621. doi:10.1111/j.0014-3820.2004.tb01742.x.

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. doi:10.1371/journal.pbio.0000005.

Buchfink, B., Xie, C., and Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi:10.1038/nmeth.3176.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST plus: architecture and applications. *BMC Bioinformatics* 10, 1. doi:Artn 421\nDoi 10.1186/1471-2105-10-421.

Cordery, D. V., Kishore, U., Kyes, S., Shafi, M. J., Watkins, K. R., Williams, T. N., et al. (2007). Characterization of a Plasmodium falciparum Macrophage-Migration Inhibitory Factor Homologue. *J. Infect. Dis.* 195, 905–912. doi:10.1086/511309.

Cornet, S., Bichet, C., Larcombe, S., Faivre, B., and Sorci, G. (2014). Impact of host nutritional status on infection dynamics and parasite virulence in a bird-malaria system. *J. Anim. Ecol.* 83, 256–65. doi:10.1111/1365-2656.12113.

Counihan, N. A., Kalanon, M., Coppel, R. L., and De Koning-Ward, T. F. (2013). Plasmodium

640 rhoptry proteins: Why order is important. *Trends Parasitol.* 29, 228–236.
641 doi:10.1016/j.pt.2013.03.003.

642 Cowman, A. F., Berry, D., and Baum, J. (2012). The cellular and molecular basis for malaria parasite
643 invasion of the human red blood cell. *J. Cell Biol.* 198, 961–971. doi:10.1083/jcb.201206112.

644 Craig, A., and Scherf, A. (2001). Molecules on the surface of the Plasmodium falciparum infected
645 erythrocyte and their role in malaria pathogenesis and immune evasion. *Mol. Biochem.*
646 *Parasitol.* 115, 129–143. doi:10.1016/S0166-6851(01)00275-4.

647 Daily, J. P., Le Roch, K. G., Sarr, O., Ndiaye, D., Lukens, A., Zhou, Y., et al. (2005). In vivo
648 transcriptome of Plasmodium falciparum reveals overexpression of transcripts that encode
649 surface proteins. *J. Infect. Dis.* 191, 1196–1203. doi:10.1086/428289.

650 Daily, J. P., Scanfeld, D., Pochet, N., Le Roch, K., Plouffe, D., Kamal, M., et al. (2007). Distinct
651 physiological states of Plasmodium falciparum in malaria-infected patients. *Nature* 450, 1091–
652 1095. doi:10.1038/nature06311.

653 Dimitrov, D., Palinauskas, V., Iezhova, T. a, Bernotienė, R., Ilgūnas, M., Bukauskaitė, D., et al.
654 (2015). Plasmodium spp.: an experimental study on vertebrate host susceptibility to avian
655 malaria. *Exp. Parasitol.* 148, 1–16. doi:10.1016/j.exppara.2014.11.005.

656 Drovetski, S. V., Aghayan, S. A., Mata, V. A., Lopes, R. J., Mode, N. A., Harvey, J. A., et al. (2014).
657 Does the niche breadth or trade-off hypothesis explain the abundance-occupancy relationship in
658 avian Haemosporidia? *Mol. Ecol.* 23, 3322–3329. doi:10.1111/mec.12744.

659 Ellis, V. A., Cornet, S., Merrill, L., Kunkel, M. R., Tsunekage, T., and Ricklefs, R. E. (2015). Host
660 immune responses to experimental infection of Plasmodium relictum (lineage SGS1) in
661 domestic canaries (Serinus canaria). *Parasitol. Res.* 114, 3627–3636. doi:10.1007/s00436-015-
662 4588-7.

663 Farias, M. E. M., Atkinson, C. T., LaPointe, D. A., and Jarvi, S. I. (2012). Analysis of the trap gene
664 provides evidence for the role of elevation and vector abundance in the genetic diversity of
665 Plasmodium relictum in Hawaii. *Malar. J.* 11, 1–8. doi:http://dx.doi.org/10.1186/1475-2875-11-
666 305.

667 Gajria, B., Bahl, A., Brestelli, J., Dommer, J., Fischer, S., Gao, X., et al. (2007). ToxoDB: an
668 integrated Toxoplasma gondii database resource. *Nucleic Acids Res.* 36, D553–D556.
669 doi:10.1093/nar/gkm981.

670 Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., et al. (2002). Genome
671 sequence of the human malaria parasite Plasmodium falciparum. *Nature* 419, 498–511.
672 doi:10.1038/nature01097.

673 Garnham, P. C. C. (1966). *Malaria parasites and other Haemosporidia.* Oxford, UK: Blackwell
674 Scientific Publications Ltd.

675 Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. a, Amit, I., et al. (2011). Full-
676 length transcriptome assembly from RNA-Seq data without a reference genome. *Nat.*
677 *Biotechnol.* 29, 644–52. doi:10.1038/nbt.1883.

678 Grech, K., Watt, K., and Read, A. F. (2006). Host-parasite interactions for virulence and resistance in
679 a malaria model system. *J. Evol. Biol.* 19, 1620–1630. doi:10.1111/j.1420-9101.2006.01116.x.

680 Green, J. L., Hinds, L., Grainger, M., Knuepfer, E., and Holder, A. A. (2006). Plasmodium
681 thrombospondin related apical merozoite protein (PTRAMP) is shed from the surface of
682 merozoites by PfSUB2 upon invasion of erythrocytes. *Mol. Biochem. Parasitol.* 150, 114–117.
683 doi:10.1016/j.molbiopara.2006.06.010.

684 Guizetti, J., and Scherf, A. (2013). Silence, activate, poise and switch! Mechanisms of antigenic
685 variation in Plasmodium falciparum. *Cell. Microbiol.* 15, 718–726. doi:10.1111/cmi.12115.

686 Haas, B. (2016). Transcriptome Contig Nx and ExN50 stats. Available at:
687 https://github.com/trinityrnaseq/trinityrnaseq/wiki/Transcriptome-Contig-Nx-and-ExN50-stats

[Accessed May 4, 2016].

Hall, N., Karras, M., Raine, J. D., Carlton, J. M., Kooij, T. W. A., Berriman, M., et al. (2005). A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and proteomic analyses. *Science* 307, 82–6. doi:10.1126/science.1103717.

Hellgren, O., Atkinson, C. T., Bensch, S., Albayrak, T., Dimitrov, D., Ewen, J. G., et al. (2015). Global phylogeography of the avian malaria pathogen Plasmodium relictum based on MSP1 allelic diversity. *Ecography (Cop.).* 38, 842–850. doi:10.1111/ecog.01158.

Hellgren, O., Kutzer, M., Bensch, S., Valkiūnas, G., and Palinauskas, V. (2013). Identification and characterization of the merozoite surface protein 1 (msp1) gene in a host-generalist avian malaria parasite, Plasmodium relictum (lineages SGS1 and GRW4) with the use of blood transcriptome. *Malar. J.* 12, 381. doi:10.1186/1475-2875-12-381.

Hellgren, O., Waldenström, J., and Bensch, S. (2004). A new PCR assay for simultaneous studies of Leucocytozoon, Plasmodium, and Haemoproteus from avian blood. *J. Parasitol.* 90, 797–802. doi:10.1645/GE-184R1.

Hisaeda, H., Maekawa, Y., Iwakawa, D., Okada, H., Himeno, K., Kishihara, K., et al. (2004). Escape of malaria parasites from host immunity requires CD4+ CD25+ regulatory T cells. *Nat. Med.* 10, 29–30. doi:10.1038/nm975.

Idaghdour, Y., Quinlan, J., Goulet, J.-P., Berghout, J., Gbeha, E., Bruat, V., et al. (2012). Evidence for additive and interaction effects of host genotype and infection in malaria. *Proc. Natl. Acad. Sci. U. S. A.* 109, 16786–93. doi:10.1073/pnas.1204945109.

Josling, G. A. a, Petter, M., Oehring, S. C. C., Gupta, A. P. P., Dietz, O., Wilson, D. W. W., et al. (2015). A Plasmodium Falciparum Bromodomain Protein Regulates Invasion Gene Expression. *Cell Host Microbe* 17, 741–751. doi:10.1016/j.chom.2015.05.009.

Kersey, P. J., Allen, J. E., Armean, I., Boddu, S., Bolt, B. J., Carvalho-Silva, D., et al. (2015). Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Res.* 44, D574–80. doi:10.1093/nar/gkv1209.

Koutsovoulos, G., Kumar, S., Laetsch, D. R., Stevens, L., Daub, J., and Conlon, C. (2016). No evidence for extensive horizontal gene transfer in the genome of the tardigrade Hypsibius dujardini. 1–6. doi:10.1073/pnas.1600338113.

Križanauskienė, A., Hellgren, O., Kosarev, V., Sokolov, L., Bensch, S., and Valkiūnas, G. (2006). Variation in host specificity between species of avian hemosporidian parasites: evidence from parasite morphology and cytochrome B gene sequences. *J. Parasitol.* 92, 1319–24. doi:10.1645/GE-873R.1.

Lachish, S., Knowles, S. C. L., Alves, R., Wood, M. J., and Sheldon, B. C. (2011). Fitness effects of endemic malaria infections in a wild bird population: The importance of ecological structure. *J. Anim. Ecol.* 80, 1196–1206. doi:10.1111/j.1365-2656.2011.01836.x.

Lampa, S., Dahlo, M., Olason, P., Hagberg, J., and Spjuth, O. (2013). Lessons learned from implementing a national infrastructure in Sweden for storage and analysis of next-generation sequencing data. *Gigascience* 2, 9. doi:10.1186/2047-217X-2-9.

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–9. doi:10.1038/nmeth.1923.

Lapp, S. A., Mok, S., Zhu, L., Wu, H., Preiser, P. R., Bozdech, Z., et al. (2015). Plasmodium knowlesi gene expression differs in ex vivo compared to in vitro blood-stage cultures. *Malar. J.* 14, 110. doi:10.1186/s12936-015-0612-8.

Lauron, E. J., Aw Yeang, H. X., Taffner, S. M., and Sehgal, R. N. M. (2015). De novo assembly and transcriptome analysis of Plasmodium gallinaceum identifies the Rh5 interacting protein (ripr), and reveals a lack of EBL and RH gene family diversification. *Malar. J.* 14, 296. doi:10.1186/s12936-015-0814-0.

Lauron, E. J., Oakgrove, K. S., Tell, L. a, Biskar, K., Roy, S. W., and Sehgal, R. N. M. (2014). Transcriptome sequencing and analysis of Plasmodium gallinaceum reveals polymorphisms and selection on the apical membrane antigen-1. *Malar. J.* 13, 382. doi:10.1186/1475-2875-13-382.

LeRoux, M., Lakshmanan, V., and Daily, J. P. (2009). Plasmodium falciparum biology: analysis of in vitro versus in vivo growth conditions. *Trends Parasitol.* 25, 474–481. doi:10.1016/j.pt.2009.07.005.

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323. doi:10.1186/1471-2105-12-323.

Li, W., and Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi:10.1093/bioinformatics/btl158.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi:10.1101/002832.

Mackinnon, M. J., Gaffney, D. J., and Read, A. F. (2002). Virulence in rodent malaria: Host genotype by parasite genotype interactions. *Infect. Genet. Evol.* 1, 287–296. doi:10.1016/S1567-1348(02)00039-4.

Martinez, C., Marzec, T., Smith, C. D., Tell, L. A., and Sehgal, R. N. M. (2013). Identification and expression of maebl, an erythrocyte-binding gene, in Plasmodium gallinaceum. *Parasitol. Res.* 112, 945–954. doi:10.1007/s00436-012-3211-4.

Martinsen, E. S., and Perkins, S. L. (2013). "The diversity of Plasmodium and other haemosporidians: The intersection of taxonomy, phylogenetics and genomics," in *Malaria parasites: comparative genomics, evolution and molecular biology.* (Caister Academic Press, Norfolk), 1–15.

Mueller, C., Klages, N., Jacot, D., Santos, J. M., Cabrera, A., Gilberger, T. W., et al. (2013). The toxoplasma protein ARO mediates the apical positioning of rhoptry organelles, a prerequisite for host cell invasion. *Cell Host Microbe* 13, 289–301. doi:10.1016/j.chom.2013.02.001.

Niang, M., Bei, A. K., Madnani, K. G., Pelly, S., Dankwa, S., Kanjee, U., et al. (2014). STEVOR is a plasmodium falciparum erythrocyte binding protein that mediates merozoite invasion and rosetting. *Cell Host Microbe* 16, 81–93. doi:10.1016/j.chom.2014.06.004.

Opitz, C., and Soldati, D. (2002). "The glideosome": A dynamic complex powering gliding motion and host cell invasion by Toxoplasma gondii. *Mol. Microbiol.* 45, 597–604. doi:10.1046/j.1365-2958.2002.03056.x.

Otto, T. D., Rayner, J. C., Böhme, U., Pain, A., Spottiswoode, N., Sanders, M., et al. (2014). Genome sequencing of chimpanzee malaria parasites reveals possible pathways of adaptation to human hosts. *Nat. Commun.* 5, 4754. doi:10.1038/ncomms5754.

Otto, T. D., Wilinski, D., Assefa, S., Keane, T. M., Sarry, L. R., Böhme, U., et al. (2010). New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq. *Mol. Microbiol.* 76, 12–24. doi:10.1111/j.1365-2958.2009.07026.x.

Pages, H., Aboyoun, P., Gentleman, R., and DebRoy, S. (2015). Biostrings: String objects representing biological sequences, and matching algorithms. *R Packag. version 2.36.1.*

Palinauskas, V., Valkiūnas, G., Bolshakov, C. V, and Bensch, S. (2008). Plasmodium relictum (lineage P-SGS1): effects on experimentally infected passerine birds. *Exp. Parasitol.* 120, 372–80. doi:10.1016/j.exppara.2008.09.001.

Palinauskas, V., Valkiūnas, G., Bolshakov, C. V, and Bensch, S. (2011). Plasmodium relictum (lineage SGS1) and Plasmodium ashfordi (lineage GRW2): the effects of the co-infection on experimentally infected passerine birds. *Exp. Parasitol.* 127, 527–33. doi:10.1016/j.exppara.2010.10.007.

Pauli, M., Chakarov, N., Rupp, O., Kalinowski, J., Goesmann, A., Sorenson, M. D., et al. (2015). De novo assembly of the dual transcriptomes of a polymorphic raptor species and its malarial

784 parasite. *BMC Genomics* 16, 1038. doi:10.1186/s12864-015-2254-1.

785 Pérez-Tris, J., Hellgren, O., Krizanauskiene, A., Waldenström, J., Secondi, J., Bonneaud, C., et al.
786 (2007). Within-host speciation of malaria parasites. *PLoS One* 2, e235.
787 doi:10.1371/journal.pone.0000235.

788 R Core Team (2015). R: A language and environment for statistical computing. *R Found. Stat.*
789 *Comput. Vienna, Austria.* Available at: https://www.r-project.org/.

790 Recker, M., Buckee, C. O., Serazin, A., Kyes, S., Pinches, R., Christodoulou, Z., et al. (2011).
791 Antigenic variation in Plasmodium falciparum malaria involves a highly structured switching
792 pattern. *PLoS Pathog.* 7. doi:10.1371/journal.ppat.1001306.

793 Reid, A. J., Blake, D. P., Ansari, H. R., Billington, K., Browne, H. P., Bryant, J., et al. (2014).
794 Genomic analysis of the causative agents of coccidiosis in domestic chickens. *Genome Res.* 24,
795 1676–1685. doi:10.1101/gr.168955.113.

796 de Roode, J. C., Culleton, R., Cheesman, S. J., Carter, R., and Read, A. F. (2004). Host heterogeneity
797 is a determinant of competitive exclusion or coexistence in genetically diverse malaria
798 infections. *Proc. Biol. Sci.* 271, 1073–1080. doi:10.1098/rspb.2004.2695.

799 Santos, J. M., Graindorge, A., and Soldati-Favre, D. (2012). New insights into parasite rhomboid
800 proteases. *Mol. Biochem. Parasitol.* 182, 27–36. doi:10.1016/j.molbiopara.2011.11.010.

801 Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets.
802 *Bioinformatics* 27, 863–864. doi:10.1093/bioinformatics/btr026.

803 Siau, A., Silvie, O., Franetich, J. F., Yalaoui, S., Marinach, C., Hannoun, L., et al. (2008).
804 Temperature shift and host cell contact up-regulate sporozoite expression of Plasmodium
805 falciparum genes involved in hepatocyte infection. *PLoS Pathog.* 4.
806 doi:10.1371/journal.ppat.1000121.

807 Siegel, T. N., Hon, C.-C., Zhang, Q., Lopez-Rubio, J.-J., Scheidig-Benatar, C., Martins, R. M., et al.
808 (2014). Strand-specific RNA-Seq reveals widespread and developmentally regulated
809 transcription of natural antisense transcripts in Plasmodium falciparum. *BMC Genomics* 15, 150.
810 doi:10.1186/1471-2164-15-150.

811 Templeton, T. J., and Kaslow, D. C. (1997). Cloning and cross-species comparison of the
812 thrombospondin-related anonymous protein (TRAP) gene from Plasmodium knowlesi,
813 Plasmodium vivax and Plasmodium gallinaceum. *Mol. Biochem. Parasitol.* 84, 13–24.
814 doi:10.1016/S0166-6851(96)02775-2.

815 Valkiūnas, G., Zehtindjiev, P., Hellgren, O., Ilieva, M., Iezhova, T. A., and Bensch, S. (2007).
816 Linkage between mitochondrial cytochrome b lineages and morphospecies of two avian malaria
817 parasites, with a description of Plasmodium (Novyella) ashfordi sp. nov. *Parasitol. Res.* 100,
818 1311–22. doi:10.1007/s00436-006-0409-3.

819 Videvall, E., Cornwallis, C. K., Palinauskas, V., Valkiūnas, G., and Hellgren, O. (2015). The Avian
820 Transcriptome Response to Malaria Infection. *Mol. Biol. Evol.* 32, 1255–1267.
821 doi:10.1093/molbev/msv016.

822 Vincensini, L., Fall, G., Berry, L., Blisnick, T., and Braun Breton, C. (2008). The RhopH complex is
823 transferred to the host cell cytoplasm following red blood cell invasion by Plasmodium
824 falciparum. *Mol. Biochem. Parasitol.* 160, 81–89. doi:10.1016/j.molbiopara.2008.04.002.

825 Wickham, H. (2009). *ggplot2: elegant graphics for data analysis.* New York: Springer.

826 Zehtindjiev, P., Ilieva, M., Westerdahl, H., Hansson, B., Valkiūnas, G., and Bensch, S. (2008).
827 Dynamics of parasitemia of malaria parasites in a naturally and experimentally infected
828 migratory songbird, the great reed warbler Acrocephalus arundinaceus. *Exp. Parasitol.* 119, 99–
829 110. doi:10.1016/j.exppara.2007.12.018.

830

Tables

Table 1. Assembly statistics of the *Plasmodium ashfordi* transcriptome.

	Annotated transcripts	Unannotated transcripts	Complete assembly
Number of contigs	7 860	4 094	11 954
Number of bases	7 316 007	1 694 373	9 010 380
Contig length min-max (bp)	200 – 26 773	200 – 4 171	200 – 26 773
Median contig length (bp)	681.0	321.0	498.0
Mean contig length (bp)	930.8	413.9	753.8
GC content (%)	21.22	17.26	20.48

839 **Table 2. Most highly expressed transcripts in the *Plasmodium ashfordi* transcriptome.**

840

P. ashfordi transcript	Protein description	Species match (blastx)	Mean FPKM[1]
TR71765|c7_g6_i1	Uncharacterized protein (PY04653)	*P. yoelii*	218 188.9
TR71765|c7_g7_i1	Uncharacterized protein (PY01927)	*P. yoelii*	189 356.5
TR87146|c0_g1_i1	Uncharacterized protein (PY04656)	*P. yoelii*	31 677.2
TR55310|c0_g1_i1	Unannotated *P. ashfordi* transcript	NA	13 753.1
TR24310|c0_g1_i1	70 kd Heat shock-like protein	*P. yoelii*	6 528.5
TR87393|c0_g5_i1	Histone H3	*P. yoelii*	5 917.4
TR52464|c0_g1_i1	Uncharacterized protein (PY00146)	*P. yoelii*	5 756.5
TR69136|c5_g1_i1	Uncharacterized protein (PFTANZ_01287)	*P. falciparum*	5 448.1
TR73260|c2_g1_i1	Uncharacterized protein (PFFCH_03197)	*P. falciparum*	4 980.4
TR162674|c0_g1_i1	Histone H4	*P. falciparum*	3 769.6
TR55306|c1_g1_i1	Heat shock protein 86	*P. vivax*	3 665.4
TR16018|c0_g1_i1	Histone H2B	*Babesia equi*	2 716.3
TR213315|c0_g1_i1	Cytochrome c oxidase subunit 1	*P. lutzi*	2 453.2
TR59776|c0_g2_i1	Uncharacterized protein (PY07298)	*P. yoelii*	2 236.1
TR113917|c0_g1_i1	Uncharacterized protein (PRCDC_1024200)	*P. reichenowi*	1 994.0

841

842 [1]Fragments Per Kilobase of transcript per Million mapped reads (FPKM) is a relative expression
843 measure normalized for library size and gene length.

844

845

846 **Table 3. Assembled transcripts of genes involved in *Plasmodium* invasion of red blood cells.**

847

Gene name	Gene product	Represented by transcripts	Species match (blastx)	Bit score	e-value
ACT1	Actin 1	TR215622|c0_g1_i1	*P. vivax*	758.4	5.5E-216
ALP1	Actin-like protein 1	TR55613|c0_g1_i1	*P. vivax*	619	5.1E-174
AMA-1	Apical membrane antigen 1	TR118224|c0_g1_i1	*P. gallinaceum*	624	2.1E-175
ARO	Armadillo-domain containing rhoptry protein	TR18124|c0_g1_i1	*P. knowlesi*	520	1.9E-144
		TR18124|c0_g2_i1		501.9	6.0E-139
BDP1	Bromodomain protein 1	TR66841|c0_g2_i1	*P. falciparum*	364.8	1.9E-97
		TR66841|c0_g2_i2		357.1	3.9E-95
FBPA	Fructose-bisphosphate aldolase	TR12911|c0_g1_i1	*P. cynomolgi*	223.8	2.5E-55
		TR12911|c0_g2_i1	*P. berghei*	353.2	2.7E-94
GAMA	GPI-anchored micronemal antigen	TR16322|c0_g2_i1	*P. reichenowi*	567.4	2.5E-158
GAP45	Glideosome-associated protein 45	TR34884|c0_g1_i1	*P. vivax*	209.9	3.9E-51
GAP50	Glideosome-associated protein 50	TR144721|c0_g1_i1	*P. knowlesi*	577.8	1.2E-161
MAEBL	Merozoite adhesive erythrocytic binding protein	TR234951|c0_g1_i1	*P. yoelii*	171.8	6.3E-40
		TR99718|c0_g1_i1	*P. gallinaceum*	146.4	2.4E-32
		TR125315|c0_g1_i1	*P. gallinaceum*	84.3	7.2E-14
MCP1	Merozoite capping protein 1	TR122100|c0_g1_i1	*P. berghei*	164.5	1.7E-37
MSP1	Merozoite surface protein 1	TR112241|c0_g2_i1	*P. relictum*	368.6	4.2E-98
		TR176579|c0_g1_i1		134.4	1.4E-28
MTIP (MLC1)	Myosin A tail domain interacting protein	TR8937|c0_g1_i1	*P. vivax*	295	8.6E-77
MTRAP	Merozoite TRAP-like protein	TR188691|c0_g2_i1	*P. vinckei*	98.6	1.3E-17
MYOA	Myosin A	TR198550|c0_g1_i1	*Babesia microti*	199.1	1.5E-47
RAMA	Rhoptry-associated membrane antigen	TR144799|c0_g1_i1	*P. knowlesi*	70.1	5.0E-09
RALP1	Rhoptry-associated leucine zipper-like protein 1	TR7446|c0_g1_i1	*P. falciparum*	102.1	1.2E-18
RAP1	Rhoptry-associated protein 1	TR115690|c2_g1_i1	*P. coatneyi*	236.1	1.3E-58
RAP3	Rhoptry-associated protein 3	TR13305|c0_g1_i1	*P. falciparum*	202.6	9.9E-49
RBP1 (RH1)	Reticulocyte-binding protein 1	TR53756|c1_g1_i1	*P. falciparum*	84.7	8.0E-13
RBP2 (RH2)	Reticulocyte-binding protein 2	TR208311|c0_g1_i1	*P. vivax*	76.6	5.9E-11
		TR66282|c1_g2_i1		146.7	2.1E-31
RBP2b (RH2b)	Reticulocyte-binding protein 2 homolog B	TR87235|c1_g1_i1	*P. vivax*	97.8	8.7E-18
		TR35437|c1_g1_i1	*P. falciparum*	86.7	1.9E-14
RHOPH1	High molecular weight rhoptry protein 1 (CLAG9)	TR7367|c0_g1_i1	*P. reichenowi*	272.3	7.9E-70
		TR200056|c0_g1_i1	*P. reichenowi*	82.8:	2.3E-13
		TR233854|c0_g1_i1	*P. falciparum*	61.6	4.8E-07
RHOPH2	High molecular weight rhoptry protein 2	TR25858|c0_g2_i1	*P. vivax*	542.3	7.8E-151
		TR25858|c0_g3_i1	*P. fragile*	335.9	5.7E-89
		TR175810|c0_g1_i1	*P. vivax*	180.6	3.4E-42
RHOPH3	High molecular weight rhoptry protein 3	TR83052|c6_g1_i1	*P. falciparum*	612.8	7.3E-172
RIPR	RH5 interacting protein	TR49727|c0_g2_i1	*P. reichenowi*	502.3	6.4E-139
		TR49727|c0_g2_i2	*P. fragile*	299.3	1.0E-77
ROM1	Rhomboid protease 1	TR178976|c0_g1_i1	*P. yoelii*	102.4	2.4E-19
ROM4	Rhomboid protease 4	TR92176|c1_g1_i1	*P. reichenowi*	783.5	2.1E-223

RON2	Rhoptry neck protein 2	TR62454|c0_g1_i1	*P. cynomolgi*	659.8	8.5E-186
RON4	Rhoptry neck protein 4	TR145809|c0_g1_i1	*P. reichenowi*	349.4	8.9E-93
RON5	Rhoptry neck protein 5	TR67313|c1_g1_i1	*P. reichenowi*	1105.9	0
SUB1	Subtilisin 1	TR164664|c0_g1_i1	*P. inui*	636.7	3.0E-179
SUB2	Subtilisin 2	TR169322|c0_g1_i1	*P. inui*	563.9	2.7E-157
TLP	TRAP-like protein	TR153597|c0_g1_i1	*P. falciparum*	440.7	3.2E-120
TRAMP	Thrombospondin related apical membrane protein	TR179194|c0_g1_i1	*P. falciparum*	231.1	2.1E-57
TRAP	Thrombospondin related anonymous protein	TR16495|c0_g1_i1	*P. relictum*	229.2	6.4E-57

Figures

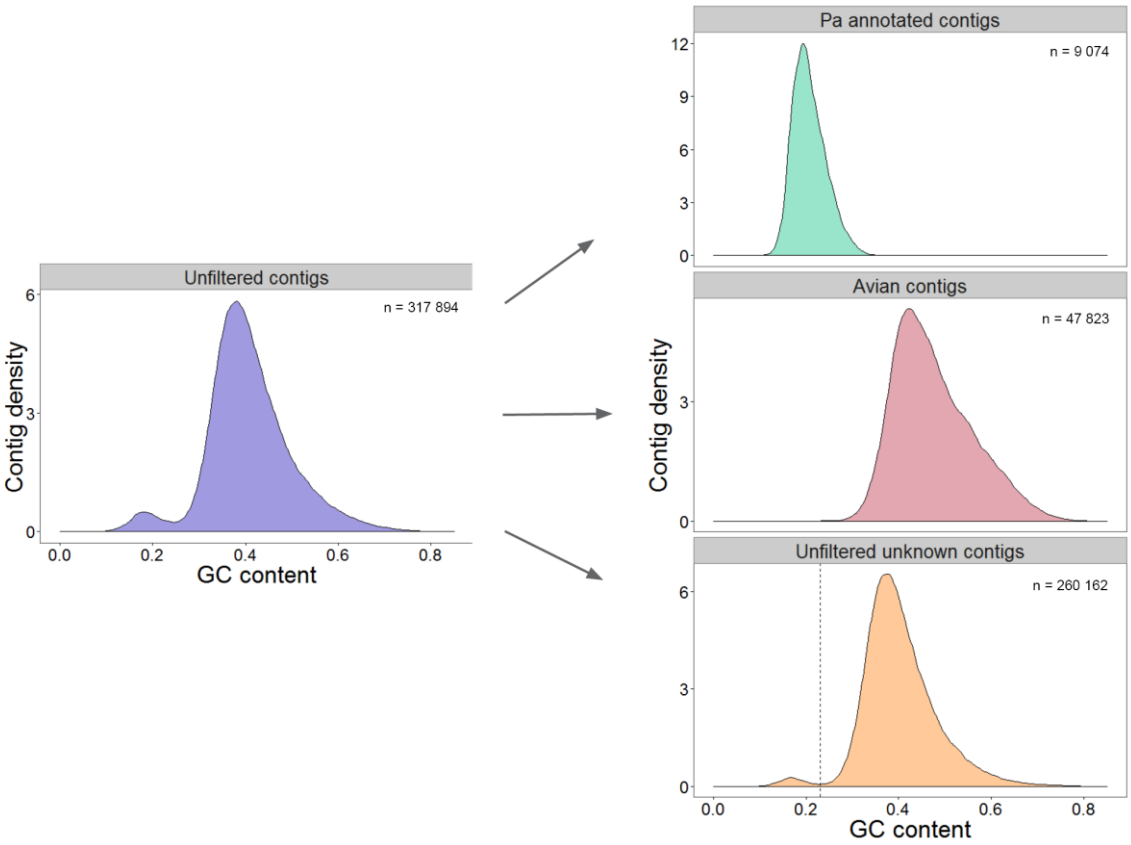


Figure 1. **Density of contig GC content in the *Plasmodium ashfordi* transcriptome.** In **(A)** the initial, unfiltered assembly, **(B)** the annotated *P. ashfordi* transcriptome assembly, **(C)** all contigs giving significant blastx matches to birds, and **(D)** all unknown contigs before GC % filtering. The arrows indicate assembly versions before and after initial filtering and cleaning steps. Both the initial, unfiltered assembly and the assembly with unknown, unfiltered contigs display a bimodal curve, incorporating both avian and *P. ashfordi* transcripts. The dashed straight line in (D) indicates the 23% GC cut-off where the unknown transcripts with lower GC content were extracted, filtered, and later included in the final *P. ashfordi* assembly as unannotated parasite transcripts.

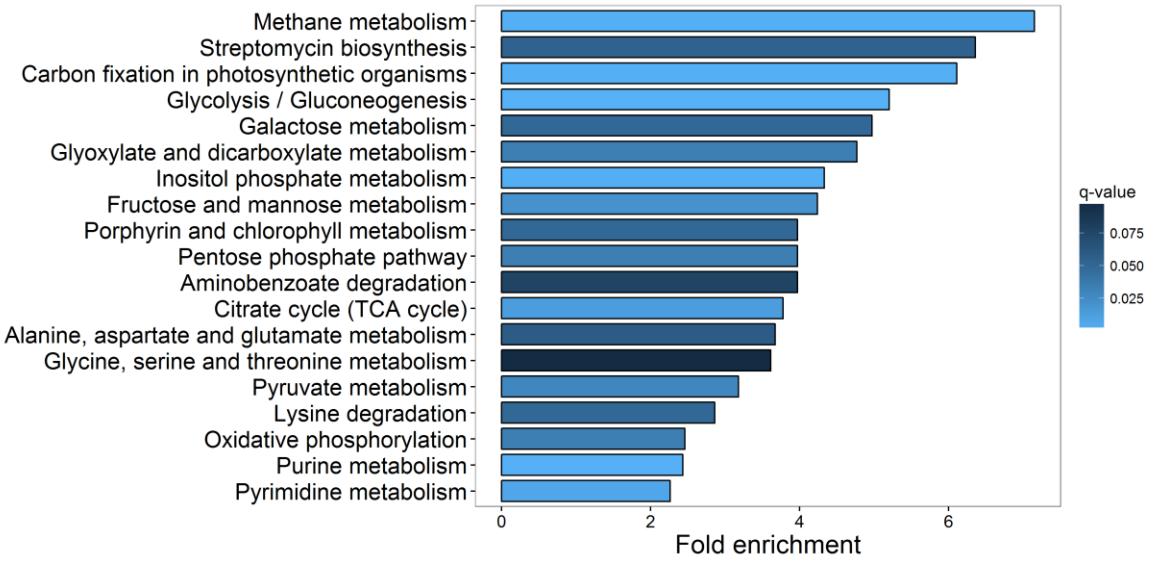


Figure 2. Enriched metabolic pathways in the *Plasmodium ashfordi* transcriptome. Significantly enriched metabolic pathways based on all transcripts in the annotated assembly. The x-axis denotes fold enrichment of genes in respective pathway compared to the annotation background of *P. falciparum*. Color gradient indicates p-values after Benjamini and Hochberg correction (q-value), where a lighter blue color signifies increased statistical significance.

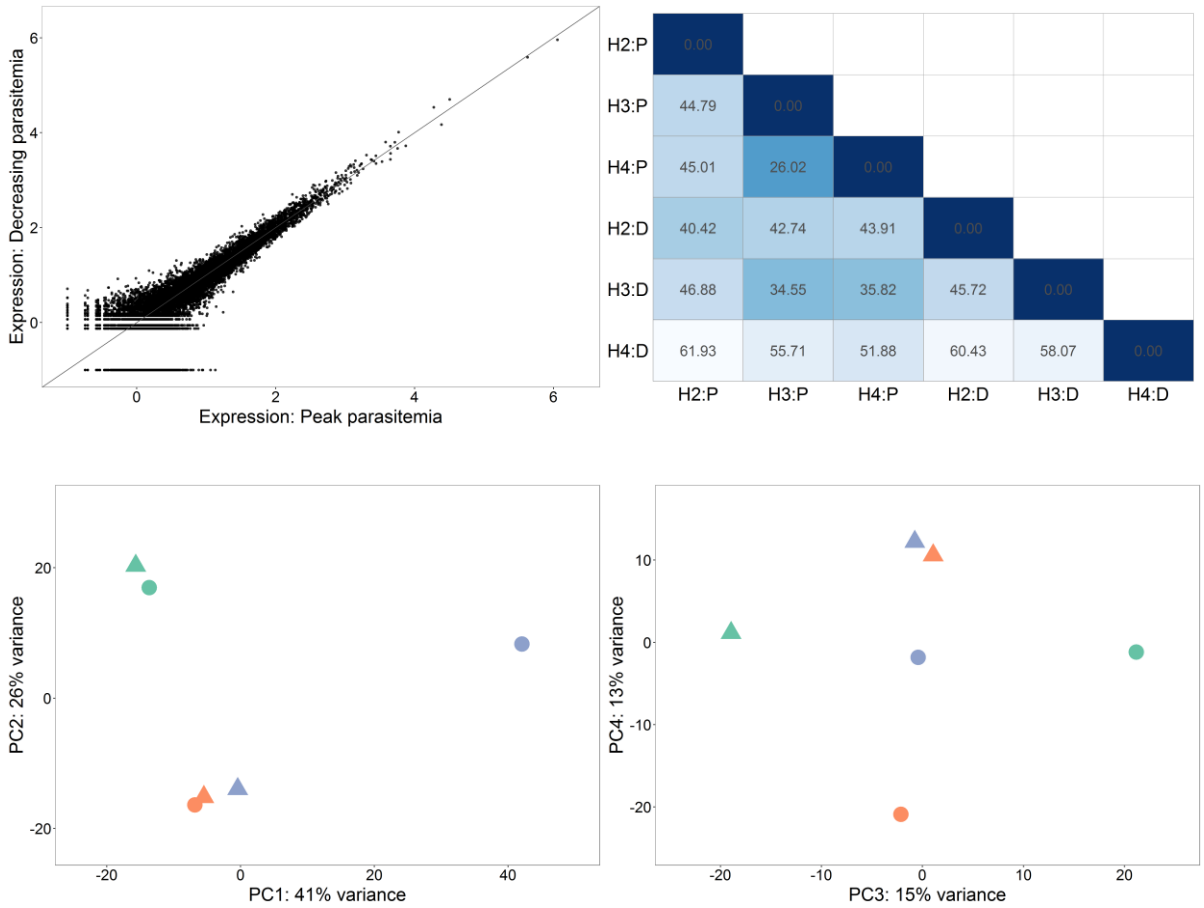


Figure 3. *Plasmodium ashfordi* gene expression in individual hosts during two infection stages. **(A)** Scatterplot displaying expression levels of all transcripts in the *P. ashfordi* transcriptome (n = 11 954). The axes show log-transformed normalized mean expression values + 0.1 during peak parasitemia stage (x-axis; n = 3) and decreasing parasitemia stage (y-axis; n = 3). **(B)** Heatmap portraying Euclidian distance measures between parasite expression profiles in different hosts during two time points. Lighter color signifies greater distance. H = host, P = peak parasitemia, and D = decreasing parasitemia. The distances between parasite transcriptomes in the decreasing parasitemia stage and to the samples from same host sampled during peak parasitemia stage constitute the shortest distances for all the samples during decreasing parasitemia. **(C-D)** Principal component analysis (PCA) plots show clustering of samples based on variation in regularized log-transformed normalized gene expression. **(C)** shows principal component 1 and 2, and **(D)** shows principal component 3 and 4. Colors of parasite transcriptomes illustrate which host they are sampled from: host 2 = green, host 3 = orange, host 4 = purple. Triangles and circles, respectively indicate transcriptomes sampled during peak and decreasing parasitemia stages.

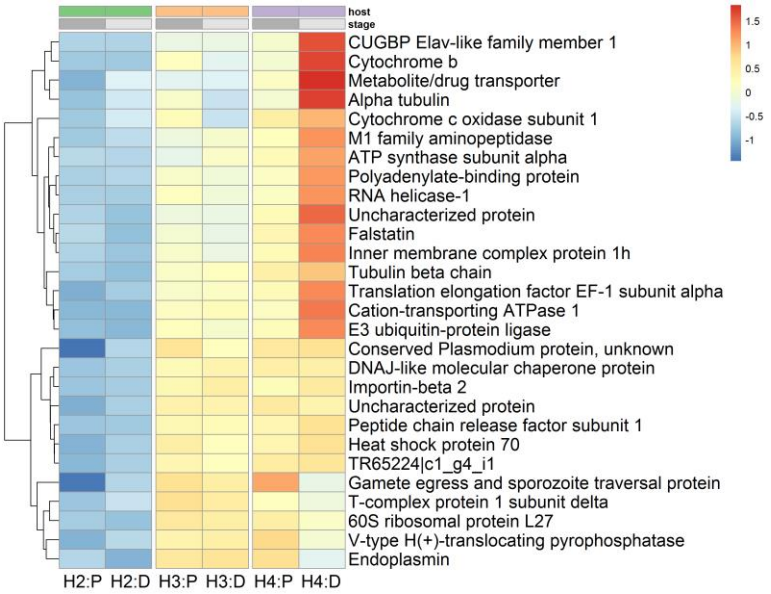


Figure 4. Expression of *Plasmodium ashfordi* transcripts differentially expressed between individual hosts. Heatmap of relative gene expression levels of 28 significant transcripts (rows) between parasites in different hosts (columns). Warmer color signifies higher expression, and blue color indicates lower transcript expression. Hierarchical clustering of genes is portrayed with a dendrogram. H = host, P = peak parasitemia (dark grey), and D = decreasing parasitemia (light grey). Same color scheme for hosts as in Figure 3C-D applies (host 2 = green, host 3 = orange, host 4 = purple). To compare across genes, expression levels have been normalized with respect to library size, regularized log-transformed, and scaled and centered around zero to give Z-scores.

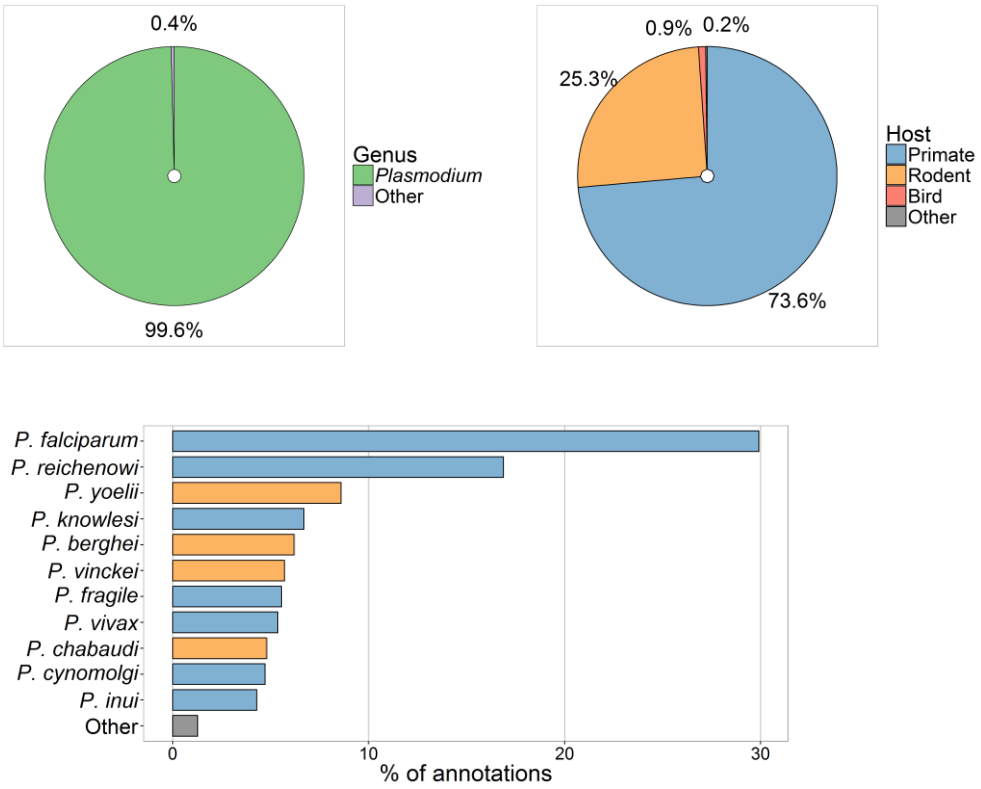
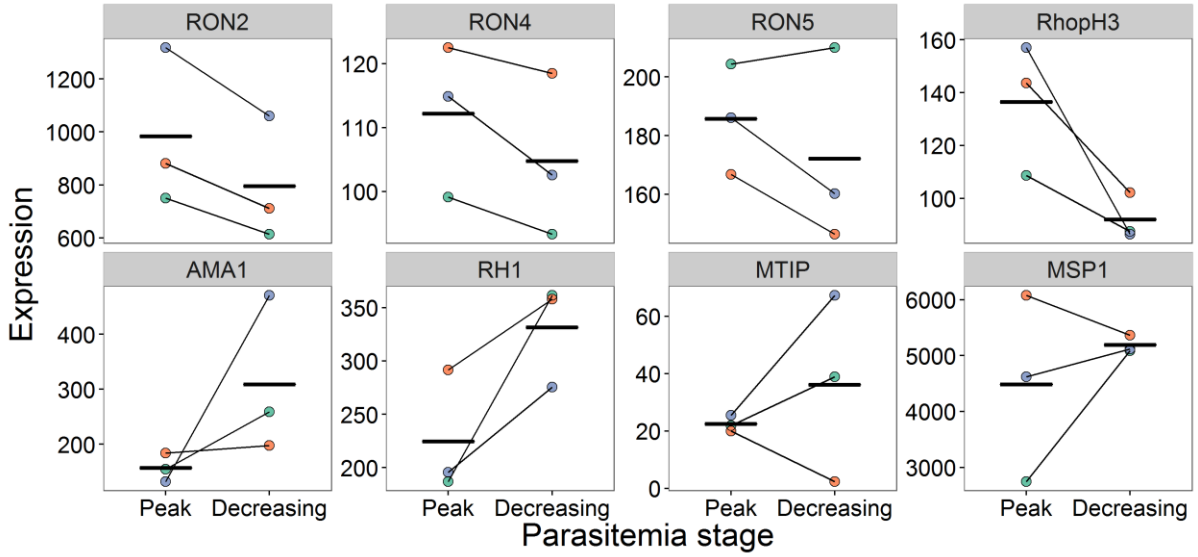


Figure 5. Distribution of apicomplexan parasites significantly matching the *Plasmodium ashfordi* transcripts. (A) Pie chart showing the distribution of *P. ashfordi* annotations giving significant blast matches to *Plasmodium* spp. and parasites of other genera. **(B)** Pie chart showing the distribution of annotations giving significant matches to parasites infecting primates, rodents, birds, and other hosts. **(C)** Bar plot displaying the percentage of annotated contigs giving significant matches to parasite species. Similar color scheme as in (B) applies, i.e. blue signifies a parasite of primates, orange a parasite of rodents. The "other" category contains here all other apicomplexan species, including bird parasites, comprising a total of 20 different species (complete list can be found in Supplementary Table S5).

925



926

927 **Figure 6**. **Individual gene expression plots for some of the *Plasmodium ashfordi* transcripts**
928 **involved in host cell invasion.** Line plots displaying normalized parasite gene expression in each
929 individual host over the two sampled parasitemia stages. Host 2 is depicted in green, host 3 in orange,
930 and host 4 in purple. Thick horizontal lines indicate mean expression levels in each stage.

931

932

Supplementary Figures

Supplementary Figures S1 – S5.
See attached supplementary pdf file.

Supplementary Tables

Please contact EV for details regarding unpublished supplementary tables.

Supplementary Table S1. Annotation information of all annotated *P. ashfordi* transcripts (n = 7 860).
[csv file]

Supplementary Table S2. *P. ashfordi* genes annotated with the gene ontology term "kinase activity" (n = 95).
[csv file]

Supplementary Table S3. Normalized expression levels of all *P. ashfordi* transcripts (n = 11 954) in individual hosts during peak and decreasing parasitemia stages.
[csv file]

Supplementary Table S4. *P. ashfordi* transcripts that were significantly differentially expressed between host individuals (n = 28).
[csv file]

Supplementary Table S5. Species distribution matches from the annotated transcripts of *P. ashfordi*.
[csv file]