Gene 792 (2021) 145723

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Shifts in gene expression variability in the blood-stage of *Plasmodium relictum*

Victor Kalbskopf [a,*], Dag Ahrén [b], Gediminas Valkiūnas [c], Vaidas Palinauskas [c], Olof Hellgren [a]

[a] MEMEG, Department of Biology, Lund University, Sweden
[b] National Bioinformatics Infrastructure Sweden (NBIS), SciLifeLab, Department of Biology, Lund University, Sweden
[c] Nature Research Centre, Akademijos 2, 08412 Vilnius, Lithuania

ARTICLE INFO

Keywords:
Avian malaria
En vivo expression
Parasitemia
Plasmodium

ABSTRACT

Avian malaria is a common and widespread disease of birds caused by a diverse group of pathogens of the genera *Plasmodium*. We investigated the transcriptomal profiles of one of the most common species, *Plasmodium relictum*, lineage SGS1, at multiple timepoints during the blood stages of the infection under experimental settings. The parasite showed well separated overall transcriptome profiles between day 8 and 20 after the infection, shown by well separated PCA profiles. Moreover, gene expression becomes more heterogenous within the experimental group late in the infection, either due to adaptations to individual differences between the experimental hosts, or due to desynchronisation of the life-cycle of the parasite. Overall, this study shows how the avian malaria system can be used to study gene expression of the avian *Plasmodium* parasite under controlled experimental settings, thus allowing for future comparative analysis of gene responses of parasite with different life-history traits and host effects.

1. Introduction

Host-parasite interactions might be one of the most common interaction taking place in nature as parasitism is considered to be the most common consumer strategy in life (Butlin and Thompson, 1995). In order to persist and transmit, the parasite must be reasonably adapted to its environment, where the host environments might change depending on the transmission strategies of the parasites. Some parasites infect a single host species whereas other parasites utilize several hosts or vector species, thus facing many different environments for which they have to be adapted to. Different hosts, either individuals within a species or different species for parasites that are host generalists, will show variation at the physiological and molecular level. In order to cope with these different environments, parasites might use different strategies (Thomas et al., 2002), either using a general strategy (non-plastic) or environmental dependent (plastic) strategy in order to effectively exploit its host. However, by having a plastic response, the parasite must have the tools to change its response based on the environment. The finest scale of non-plastic or plastic interactions between hosts and parasites take place at the molecular level. This is where the genetic composition and gene expression of the host and the parasite is what ultimately determines the outcome of the interaction (Penczykowski et al., 2016). In order for the parasite to survive and reproduce it must have the molecular "keys" to invade the host and then evade and/or suppress the immune system in order to proliferate. However, how static is the gene responses of parasites when infecting different individuals? One host-parasite association where this can be explored is the avian malaria system.

To date, fifty-five species of avian malaria parasites, *Plasmodium* spp., have been morphologically identified (Valkiunas and Iezhova, 2018), but with the use of molecular barcoding this number might be considerably higher, even in the magnitude of several thousand evolutionary independent lineages or species (Bensch et al., 2004). Avian malarial parasites are transmitted by mosquitoes and belong to the same genus as human and other mammalian species with which it shares the main features of the infection cycle. As in mammalian malaria, the sporogony and sexual process occur in the mosquitoes and only asexual replication takes place in the vertebrate host. The asexual reproduction (exoerythrocytic merogony) starts in different internal organs that might differ depending on species, followed by invasion of the red blood cells (RBC). Replication in the RBC produce both merozoites which can reinfect RBC, or the sexual stages, gametocytes, which develop into male

Abbreviations: dpi, days post infection; GO, gene ontology; CoV, co-efficient of variation; BLAST, Basic local alignment search tool; TBP-1, Tat binding protein.

* Corresponding author at: Ekologihuset, Naturvetarvägen 6A, 22362 Lund, Sweden.
E-mail address: victor.kalbskopf@biol.lu.se (V. Kalbskopf).

https://doi.org/10.1016/j.gene.2021.145723
Received 19 February 2021; Received in revised form 12 May 2021; Accepted 14 May 2021
Available online 18 May 2021
0378-1119/© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

and female gametes when ingested by mosquitoes (Baker, 2010).

Within the avian malaria parasites, *Plasmodium relictum* is one of the most common and cosmopolitan species observed to date (Hellgren et al., 2015). It has a diverse host repertoire, infecting over 300 species across 11 bird orders. Although pathology in many established host species tends to be low (Palinauskas et al., 2008), it has allegedly contributed heavily to the extinction of newly acquired hosts in the honeycreeper species in Hawaii (Atkinson and LaPointe, 2009) and zoo birds such as penguins (Fix et al., 1988). Important to note is that within the morphologically defined species of *P. relictum* there exist several differently defined mitochondrial lineages that differ both in virulence, host responses (Videvall et al., 2020) and in geographical distribution. The most prevalent, geographically ubiquitous lineage and best studied to date are the lineages SGS1, GRW4 and GRW11. Each of these lineages has dissimilar host preferences, parasitemia intensity depending on host, and reproduction period patterns within their preferred host (Drovetski et al., 2014). SGS1 may cause a long term chronic infection in one species, but cause mortality in another species (Valkiūnas et al., 2018). Why such closely related lineages appear to have such distinctive host-related characteristics requires further examination. If the inter-lineage genomic content is so similar, perhaps the differences in gene expression can explain the host-specific parasite responses.

Current experimental models used to study gene expression of human malarial parasites are performed *in vitro*, which require specific media and perfusions to represent each stage in the development (Cowman and Crabb, 2006; Yang et al., 2017; Roth et al., 2018). *In vitro* studies are useful since the environment is easily controlled, so reproducibility is improved. In order to identify stage specific genes, such an experimental set up is ideal. However, when trying to understand how the parasites behave in natural host populations that are genetically heterogenous, other study systems are needed. In the wild a parasite does not act in isolation and its expression and responses are a direct effect of the environment it inhabits i.e., the host or vector. Specifically, the host has biological functions and responses that are not present in tissue culture, such as the immune system, adaptive thermal regulation, and a wide variation in genetic make-up of the hosts (the hosts in the wild are not clonal, and will therefore show variation in their responses). Thus, if we want to understand the mechanisms involved in the host-pathogen arms-race and how the parasite reacts to the host immune system, stress levels or manipulates the host for its own purposes, we also need information from more natural study systems as it is the variation in expression and response that holds the key to understanding the differences in sickness severity, tolerance, and resistance. Therefore, despite the difficulties, this approach is needed to identify the genes that are interacting in the host and the parasite, which determines the fate of the infected individual. As a result, experimental set-ups have been developed where it is possible to study natural malaria infections in birds under controlled experimental settings (Palinauskas et al., 2008; Dimitrov et al., 2015). In short, uninfected juveniles are infected with blood from an infected donor bird. This method ensures the use of the same malaria isolate and the same starting parasite concentration, which is not possible using mosquitoes as vectors (Churcher et al., 2017).

In a previous study using synchronised infections of birds with *Plasmodium ashfordi* (lineage GRW2), the gene expression was compared between early and late time points within the same hosts, no significantly differentially expressed genes were found between time points (Videvall et al., 2017). However, the overall patterns of expression revealed greater similarity within individual hosts than within time points. In this study, we follow up that investigation by investigating the responses of avian malaria by using a larger experimental design and take advantage of a recently published and annotated *P. relictum* parasite genome (Böhme et al., 2018) which was not available in previous studies in order to further investigate avian malaria parasite gene expression, and track how the expression changes over time.

2. Methods

2.1. Infections

Juvenile Eurasian siskins (*Spinus spinus*) were caught using mist nets in mid-June to limit the risk of the birds already being infected, and housed in mosquito-free aviaries. Samples were checked for parasites by PCR screening and blood smear to ensure that the birds did not have any ongoing infections. A total of 9 birds were injected with blood containing the erythrocytic stages of the *P. relictum* lineage SGS1. This specific isolate had previously been isolated from *Loxia curvirostra* (common crossbill) and multiplied further in common crossbills and siskins in order to obtain enough inoculum for all experimentally infected birds. The donor blood was mixed before inoculation to ensure all birds received the same number of parasites. Every four days, 70 μl of blood was taken for investigations of parasitemia levels using microscopic examination. At days 8, 20, and 36 post inoculation, 40 μl of blood was collected for RNA extraction. Blood samples from day 8 and 20 were selected for RNA-sequencing as they represent the periods of parasitemia just prior and post peak infection respectively. The blood was directly put into liquid nitrogen and kept at −80C until extraction. For a detailed description of the infection procedure, see Videvall et al (2020) and Dimitrov et al (2015).

2.2. RNA extraction and sequencing

Total RNA was extracted from the blood samples by first adding 1000 μl TRIzol® to each sample, and after vortex homogenisation and incubation at room temperature for 5 min, 200 μl of chloroform was added. The sample was then incubated for 3 min at room temperature, and centrifuged at 11 000 rpm for 15 min at 4 °C. The supernatant was then processed following the RNeasy Mini Kit (Qiagen, GmbH, Hilden, Germany) manufacturer's instructions, starting from point 4 in the instruction manual. The final lysate was vacuum dried and shipped on dry ice to the sequencing facility at Beijing Genomics Institute (BGI), China, where RNA quality control, DNAse treatment, rRNA reduction, and amplification using the SMARTer® Ultra™ Low kit (Clontech Laboratories, Inc.) was done. BGI performed library preparation, cDNA synthesis, and paired end sequencing using Illumina HiSeq 2000. For a detailed description of the RNA extraction, see Videvall et al (2020).

Obtained reads were filtered using the Illumina Chastity filter, trimmed from 100 to 75 base pairs and subsequently mapped to the *P. relictum*-like genome (PlasmoDB) using the splice-aware mapper STAR (ver 2.7.0). To assess the quality of the transcriptome and ensure that host reads were excluded, a reciprocal BLAST (blastn 2.9.0) was performed: all reads mapping to the parasite genome, using STAR, were blasted (E-value of 1e-5 cut-off) against the zebra finch (*Taeniogypia guttata*) cDNA library retrieved from NCBI. The zebra finch hits, reads with an E-value of at least 1e-5, were blasted against the *P. relictum* gene blast database.

2.3. Data analysis

The BAM files, obtained from mapping the reads to the reference genome, were indexed using SAMtools (ver 1.9) and the reads were assigned to genes using their gff coordinates on the *P. relictum* -like genome and counted using HTSEQ Count (ver 0.11.2).

The raw gene count data was analysed with DESeq2 (ver 1.22.2) (Love et al, 2014) using un-normalised data. In order to characterise the similarity in gene expression between samples, the counts were transformed using a regular log transformation (rlog), which makes the data homoscedastic, i.e., the variance is independent of the mean. This variance stabilisation accounts for the differences in overall read depth which are a result of differences in parasitemia. A rlog transformation was used because its robustness when there are large differences in size factors (Love, Huber and Anders, 2014), which is the case in this dataset.

In order to visualise the similarities and dissimilarities between samples their Principal Component Analysis (PCA) values, i.e., PC1-PC3, were plotted in two dimensions. The PCA values were calculated based on their pairwise Euclidean distance as implemented in DESeq2. We further constructed a cluster dendrogram, also based on Euclidean distances using a Ward criterion as implemented in DESeq2. Both the PCA and the cluster dendrogram cluster samples together based on similarities in their gene expression (Ringnér, 2008).

2.4. Differential expression

To detect whether different genes showed significantly different gene expression patterns between day 8 and day 20 (data from day 36 was excluded as the parasitemia levels where too low in order to generate enough coverage to perform any analysis), the normalised counts were used in a test for differential gene expression as implemented in DESeq2. Genes that showed significantly different gene expression were further analysed for functional enrichment using the Gene Ontology (GO) analysis tool on the PlasmoDB site. By analysing GO-terms that were significantly overrepresented among genes that showed significantly different gene expression, specific biological functions or molecular pathways that are important during the different time points can be identified. Testing for enrichment for any particular GO-term involves finding the likelihood that the term would be enriched by chance, and results in a p-value. We used the typical 0.05p-value, the 5% chance that our positive result is false.

After testing for differences in expression based on time point 8 and 20 days post infection (dpi), the link between parasitemia levels and differential gene expression was investigated for samples originating from day 8. The 20 dpi samples were not included because they, with one exception, all had very low parasitemia. The low parasitemia resulted in very low expression and therefore many genes were not sequences/mapped, which caused zero counts. These missing genes would excessively restrict the total set of genes for analysis. Samples with a parasitemia of 15% and lower have low virulence, and above 15% have high virulence. This divided the samples in the 8 dpi group in approximately half (4 and 5 samples with low and high parasitemia respectively). These groups were subjected to the same differential expression analysis as described between timepoints above. This categorical division may reduce the statistical power of the test, but it makes it easier to compare the time-point test and parasitemia test.

2.4.1. Correlating gene expression and parasitemia

In parallel, we investigated the relationship of overall gene expression in individual birds and the parasitemia. A Spearman correlation test was performed on the first two principal component values and the parasitemia of each sample. In the case that there was a significant correlation, we selected the 200 genes each that had the highest positive or negative loading factors for that PC-component for a GO enrichment analysis performed on Plasmo-DB.

2.5. Comparing overall variation in gene expression between and within time points

We hypothesised that genes interacting with the host at the individual level would also be genes that show large variation in expression between samples at a given time point, whereas genes involved in housekeeping and general metabolic processes do not respond significantly to the surrounding environment of the parasite. In contrast, large variation in expression would be expected for genes that are interacting with the hosts when the hosts are not genetically identical. Hence, to characterise the plasticity, defined as the variation across individuals, of the parasite's gene expression, we analysed the genes that varied the most between individual birds at same time point. These highly varied genes would reveal the way the parasite responds to the differences between hosts. This was accomplished by calculating the co-efficient of variation (CoV) of the normalized gene expression of each gene at each of the time points. As there were many genes that did not have any expression in at least two samples, the variation was inflated. To avoid this inflation in variance, genes with two or more null counts were removed from the analysis. The remaining 3565 genes (out of a total of 5305 total genes) were used to construct violin plots, one for each day where the width represents the number of genes with the CoV on the vertical axis. Then the difference in gene CoV between 20 dpi and 8 dpi samples was calculated. The genes in the top 5% of the distribution of differences were selected for GO term and metabolic pathway enrichment analysis using the PlasmoDB analysis tool (Aurrecoechea et al., 2009).

3. Results

3.1. Data analysis

3.1.1. Host vs parasite reads

Approximately 13% of the total sequence reads originated from a *Plasmodium* parasite, but ranged between 0.3% and 43% depending largely on the parasitemia level of the sample. The reciprocal BLAST for any bird RNA that potentially could contaminate the analysis which was done using the reads that first had been mapped against the *P. relictum* genome revealed only five significant hits against the bird genome. The extremely low number of reads which also mapped to the bird genome in the filtered dataset gives us confidence that we received a "clean" transcriptome, more or less free of contaminating host reads. The genes that were found among the reciprocal BLAST hits were the 28S ribosomal gene, eukaryotic translation initiation factor 3 subunit C, heat shock protein 70, dynactin subunit 2, and one gene with an unknown function. The first four are highly conserved genes in eukaryotes, therefore these may simply be malarial homologs of the avian genes and were left in the dataset.

Two of the 20 dpi samples, from bird 4 and 5 (App. A Table 1) were removed from the analysis due to very low parasitemia (0.04% and 0.03%, respectively) and thus too low read coverage (Fig. 1A) which produced a very large number of zero count genes, i.e., genes to which no reads were mapped against (over 60%).

When samples from birds 4 and 5 were included in the differential expression analysis, we received more significant genes (19) but nearly all of these additional genes were uncharacterised and un-annotated.

There was a clear link between parasitemia and mapped reads per sample (Fig. 1B), both when analysing all the samples as well as for the individual days. This resulted in both a lower total number of reads and thus more genes with null counts at 20 dpi where all the samples had a lower level of parasitemia compared to samples from 8 dpi (App. A Fig. 1).

To visualise how similar or dissimilar the different samples were in gene expression, we constructed a cluster dendrogram (Fig. 2), which group samples based the Euclidean distance between samples. This variation was further visualized through Principal Component analysis (Fig. 3).

When visualizing the similarity or dissimilarity between gene expression between samples we observed groupings by time point (8 dpi samples tended to cluster together and 20dpi samples formed one clade or cluster), and not by biological replicate. This was found both in the cluster dendrogram (Fig. 2), which group samples based the Euclidean distance between samples as well as when mapping the PC values for each sample (Fig. 3). Only two exceptions were observed, birds 1 and 3, which grouped by host.

3.2. Differential expression

Before testing for differential expression, individual gene expression was made comparable through gene-by-gene normalisation to adjust for technical and biological effects. The scaling factors used to normalise the

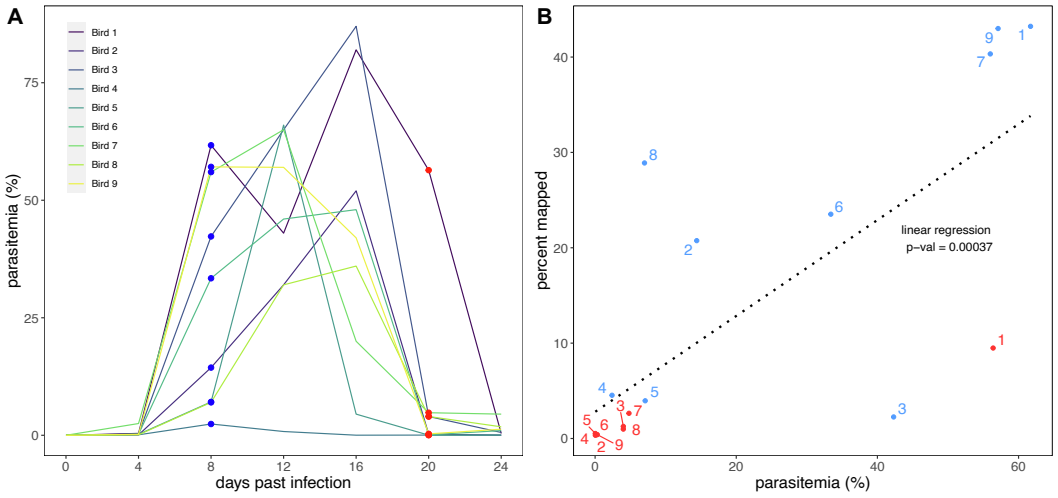

Fig. 1. **A** Parasitemia in the samples used in differential expression analysis of nine birds experimentally infected with *Plasmodium relictum*. RNA from blood samples taken at 8 and 20 days post infection (dpi) were sequenced. Samples from birds 4 and 5 were removed from the entire time-point based differential expression analysis due to low read coverage. **B**: Scatter plot showing a correlation between the parasitemia of each sample and the number of reads that mapped to the parasite genome for that sample. Red dots are the samples taken 20 dpi, and blue dots 8 dpi. The numbers are bird IDs. The linear regression has a p-value of 0.00037. The correlation was significant even when day of sampling was included as a factor. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

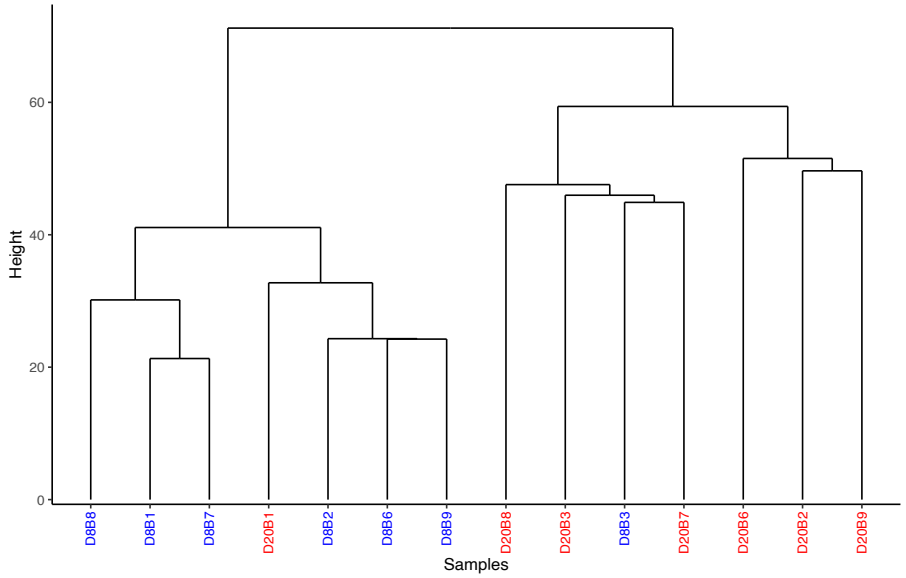

Fig. 2. A cluster dendrogram based on the Euclidean distance of each sample's count data. The samples group by the day the blood was taken, with the exception of day 20-bird 1 and day 8-bird 3.

count data ranged from 12.78 to 0.04 (App. A, Table 2). After the normalisation the mean and variation around the $\log_2$ count were more equal between the samples allowing for further testing of differential expression (App. A Fig. 2 boxplots). For example, sample day 8-bird 3 got in line with the other samples from day 8 and should not be a source of bias in the analysis. The distribution was similar for all samples, but the samples from birds 2 and 6 in day 20 are exceptions due to the large number of zero counts.

To reduce the number of genes we examined (and therefore number of unnecessary tests), DESeq2 performs independent filtering, with the assumption that genes with a very low count will have a very high dispersion (where the variance is greater than the mean) and will therefore not have significant differences in expression. Of the 5305 genes, 1322 genes were discarded in this step. Although missing out of some of the genes, this increases the power of the overall test, as fewer tests must be corrected for, lowering the number of false negatives.

We fit the data to a negative binomial distribution and estimate the coefficients of the model counts ~ day. These are interpreted as $\log_2$ fold change and are tested to get p-values. After Benjamini-Hochberg correction for multiple testing, we get the adjusted p-values (Table 1). There were 12 genes significantly upregulated in day 20 samples (Table 1). Most of the GO terms associated with these genes in day 20 were related to cell signalling and metabolism (Table 2 and App. Table 3). Nine genes were found to be upregulated in 8 dpi samples, and their GO-terms were related to cell movement (Table 3 and App. Table 4).

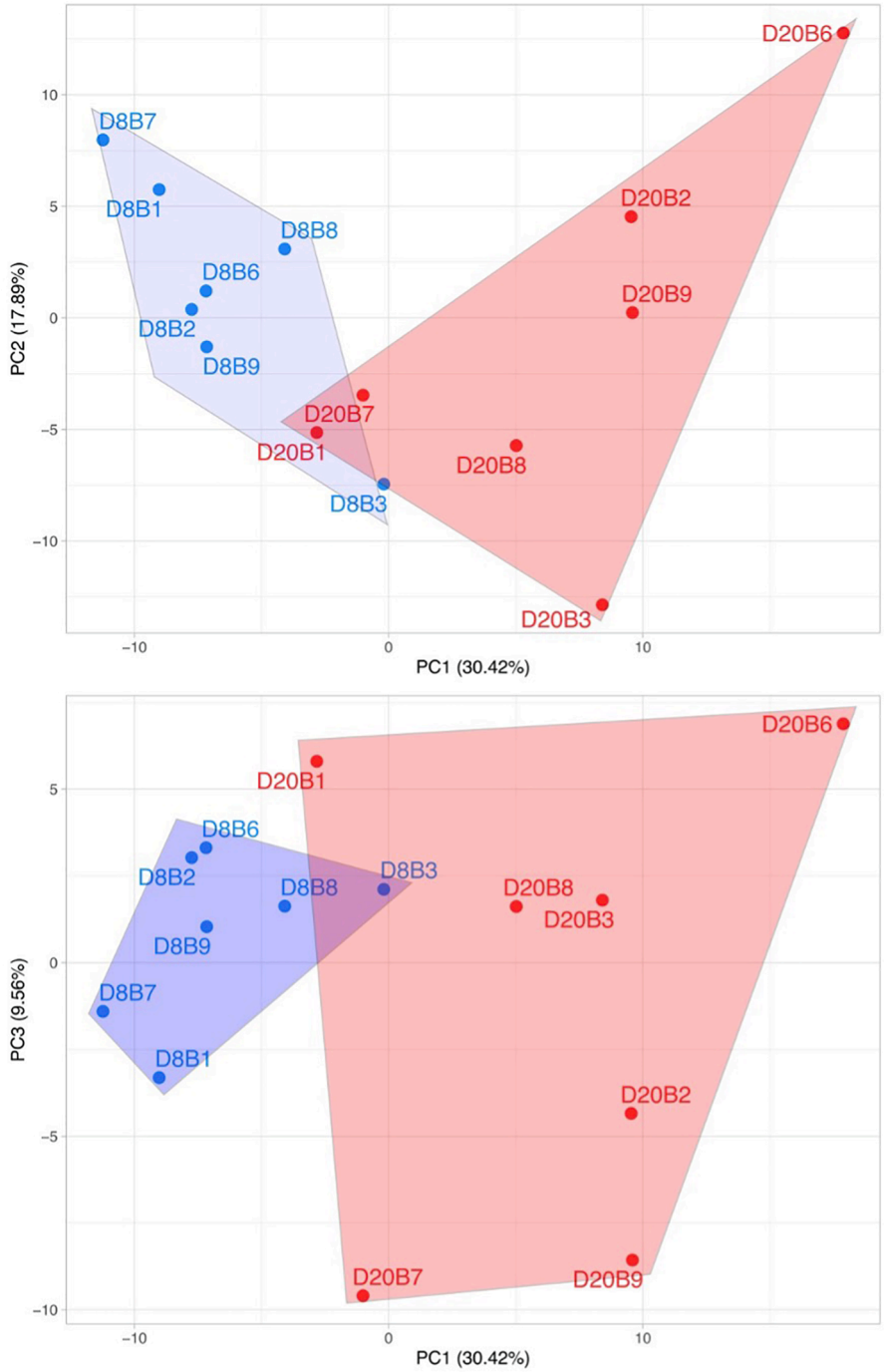

Fig. 3. The first 3 principal components of variation comparing RNA-seq samples taken 8 and 20 days post infection. Day 20 samples (red) have more variation compared to day 8 samples (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.3. Parasitemia linked differential expression in day 8 samples

Because the strong link between parasitemia, mapping success, and the day the samples were taken (Fig. 1), we investigated if the significantly differentially expressed genes we found between the time points (see above) were an effect of high parasitemia. Therefore, we wanted to test if parasitemia levels can predict gene expression. If this was the case, then the time point differential expression results would be, if not invalid, then unreliable.

We selected only the 8 dpi samples (the 20 dpi samples were all too low in virulence) and divided them into high and low virulence, based on a 15% parasitemia cut off which resulted in roughly equally sized groups – five samples in the high parasitemia group and four in the low parasitemia group. We included all 9 birds in this analysis. There was only one significant differentially expressed gene, which was the large subunit ribosomal gene (adjusted p-value 0.046, fold change of 2.06). This result is confirmed in the PCA, which showed no clustering by parasitemia for the first 3 principal components (App. Fig. 3, only first 2 components; third component not shown).

3.3.1. Correlating gene expression and parasitemia

In parallel, we looked for correlation of individual gene expression and parasitemia, using the Spearman correlation test. After correcting for multiple testing, we did not find any significant correlations. The

Table 1
Genes with a significantly different expression in 20 dpi samples compared to 8 dpi samples. A positive fold change indicates higher expression in 20 dpi samples and genes with a negative fold change were expressed significantly more in 8 dpi samples.

Name/description	log2FoldChange	Padj	ID
28S ribosomal RNA	5,655	3,01E−10	PRELSG_0000120
RNA-binding protein, putative	3,339	0,0003	PRELSG_1425000
conserved *Plasmodium* protein, unknown function	4,17	0,0003	PRELSG_1431700
OST-HTH associated domain protein, putative	3,717	0,0005	PRELSG_1242200
liver specific protein 1, putative	2,274	0,0005	PRELSG_1333000
nucleus export protein BRR6, putative	2,906	0,0011	PRELSG_1438800
phosphodiesterase gamma, putative	4,629	0,0042	PRELSG_1419400
reticulocyte binding protein, putative	1,254	0,0101	PRELSG_0008000
conserved *Plasmodium* protein, unknown function	1,471	0,0284	PRELSG_1202300
Plasmodium exported protein, unknown function	2,388	0,0349	PRELSG_0003110
conserved *Plasmodium* membrane protein, unknown function	2,748	0,0496	PRELSG_0401900
zinc finger protein, putative	−1,105	0,0004	PRELSG_1112000
reticulocyte binding protein, putative	−1,537	0,0103	PRELSG_0021100
kinesin-19, putative	−0,898	0,0103	PRELSG_0313500
conserved *Plasmodium* protein, unknown function	−0,726	0,0252	PRELSG_0722200
fam-e protein	−2,551	0,0265	PRELSG_0003700
conserved *Plasmodium* protein, unknown function	−0,736	0,0265	PRELSG_0917500
conserved *Plasmodium* protein, unknown function	−0,933	0,0349	PRELSG_0938400
conserved *Plasmodium* protein, unknown function	−1,278	0,0449	PRELSG_1418800
CPW-WPC family protein	−1,58	0,0489	PRELSG_1265500

number of tests was large, 2765 genes, so we looked for a testing regimen that would be less specific about the genes, but still reveal significant trends. Therefor we looked for correlations between parasitemia and the principal components of the PCA. Using the Spearman correlation test, we found a significant ($p = 0.031$) positive correlation ($rho = 0.73$) between the parasitemia level and the second principal component (PC2) (Fig. 4). The second principal component explains 17% of the variation between samples. Following on that analysis, we identified the genes most influential on this PC axes, and selected the 200 most positive (suppl. Table 1) and negative (suppl. Table 2) genes for GO enrichment analysis. Some of the most positive genes correlated with high parasitemia in descending order are inner membrane complex protein 1b, putative; major facilitator superfamily domain-containing protein, putative; large subunit ribosomal RNA; LCCL domain-containing protein; conserved protein, unknown function; upregulated in late gametocytes ULG8, putative. The GO terms significantly enriched in the larger gene set are related to cell movement and DNA repair/replication (suppl. Table 3). The genes most correlated with low parasitemia in descending order are OST-HTH associated domain protein, putative; 28S ribosomal RNA; 60S ribosomal protein L28, putative; 60S ribosomal protein L44, putative; 60S ribosomal protein L35, putative; 60S ribosomal protein L39. This gene set was significantly enriched for GO terms related to translation and biosynthesis (suppl. Table 4).

3.4. Comparing variation in gene expression between and within time points

The amount of variation in gene expression between individuals (the same gene in different hosts) was visualised using violin plots where the samples were divided up by days post infection (Fig. 5A). Here you can see the distribution of the CoV in gene expression showed very different patterns between 8 dpi and 20 dpi. At 8 dpi most of the genes showed low amount of variation between the different infected individuals with a few genes with large variation. However, 20 dpi the pattern was very different, showing almost a bimodal distribution in CoV, with an almost equal number of genes with low variability between individuals as there were genes showing large variation in expression between individuals, i.e., genes found in the upper distribution of the violin plot. That there was an increase in gene variation at 20 dpi can also be seen in the PCA plots where the PCA area that the samples contribute to is much larger on 20 dpi compared to day 8 (Fig. 3). Both results indicate an increase in variation in gene expression between individuals at 20 dpi compared to 8 dpi. To identify the genes that may be responding to their particular host, the change in CoV for each gene between 20 dpi and 8 dpi was calculated. The CoV 8 dpi was subtracted from 20 dpi for each gene, and the results were plotted as a histogram (Fig. 5B). The genes with the greatest difference from the top 5% of the distribution were selected (Supplemental Data: 77_Change_in_CoV.csv; genes_greatestChangeCoV.txt) for GO term enrichment analysis. These genes had a change in CoV of at least 77. The GO terms are presented in Table 3 (complete table in Supplemental Data: GOterms_greatestChangeCoV.txt) and relate to transcription and metabolism, all low-level processes in the parasite. Metabolic pathway enrichment (KEGG and MetaCyc) revealed pathways for sucrose and carbon metabolism (Supplemental Data: pathways_greatestChangeCoV.csv). However, after adjustment for multiple testing, the p-values for the GO and pathway terms no longer met the 5% false-positive rate threshold.

4. Discussion

In this study, we examined the gene expression of the malaria parasite *P. relictum* during different time points within its avian hosts with the use of RNA-sequencing. Gene expression of the parasite was measured at two different time points when the parasite multiplies in the bloodstream of its host. In order to separate the gene expression of the parasite from that of its host, multiple steps of *in silico* filtering was done prior to analysis to ensure that a cleaned data set only including sequencing reads from the parasite was investigated.

In a prior transcriptome study of avian malaria (Videvall et al., 2017), there were indications that parasites tuned their gene expression to fit individual hosts and that this effect had a higher impact on shaping the overall gene-expression profiles compared to days of sampling. This was found as gene expression tended to group on hosts and not time points during the infection. Here we continued this investigation using a larger experimental set-up and working with a parasite for which we had an annotated reference genome to map gene expression against. The aim was to understand how the parasite regulates their gene expression in relation to the individual hosts it infects. Are the parasites responses static irrespective of the different environments in the hosts where their gene expression profiles will be similar, although the expression can change over the course of infection as the parasite transit into different life stages (Bozdech et al., 2008), or do they show a plastic gene expression response meaning that they will tune their gene expression to fit the difference in the environment each individual host consists of?

4.1. Differentiation in expression between different time points

The main biological difference between the timepoints of sampling is that during 8 dpi the population of parasites are increasing, i.e., they have not yet reached peak of parasitemia in the host (Fig. 1A) and is therefore not under control of the hosts immune system. Whereas at 20 dpi the peak has been passed and the parasitemia is on its way down to a low, potentially chronical level. During 20 dpi, the parasites are supressed by an active immune response and might to a larger extent be

Table 2
The GO-term enrichment analysis of the genes overexpressed in 20 dpi (shaded in red) samples revealed functions related to cell signalling and metabolism. The genes that are overexpressed in 8 dpi samples (shaded in blue) are related to cell movement. The * indicates the GO term is still significant after Benjamini-Hochberg adjustment for multiple testing.

GO-term	Uncorrected P-value
lipid homeostasis	0.0012*
cGMP metabolic process	0.0046
chemical homeostasis	0.0081
cyclic nucleotide metabolic process	0.0093
biological regulation	0.0104
signal transduction	0.0231
Signalling	0.0231
cell communication	0.0242
purine ribonucleotide metabolic process	0.0401
ribonucleotide metabolic process	0.0412
purine nucleotide metabolic process	0.0424
homeostatic process	0.0435
ribose phosphate metabolic process	0.0458
purine-containing compound metabolic process	0.0458
microtubule-based movement	0.0139*
movement of cell or subcellular component	0.0208*
microtubule-based process	0.0262*

Table 3
The genes which depicted the highest difference in CoV between days post infection (Fig. 5B) were subjected to GO-term enrichment. The results, while not significant after correction for multiple testing, are presented anyway.

Name	Uncorrected P-value
vesicle-mediated transport	8.61E−03
phosphorylation	9.97E−03
protein phosphorylation	1.19E−02
cellular protein localization	2.54E−02
cellular macromolecule localization	2.54E−02
cellular glucose homeostasis	3.08E−02
carbohydrate homeostasis	3.08E−02
glucose homeostasis	3.08E−02
proton-transporting ATP synthase complex assembly	3.08E−02
proton-transporting two-sector ATPase complex assembly	3.08E−02
protein localization	4.75E−02

fighting for survival and are living in a host whose gene regulation is responding to the stress induced by the parasite during peak infection (Videvall et al., 2020).

When comparing the gene expression between the different time points in the experiment 20 genes were found to differ in their expression between 8 dpi and 20 dpi. When using these genes in a GO-term enrichment analysis 3 GO-terms were found to be significantly enriched for 8 dpi and 14 for 20 dpi. The GO terms at 8 dpi (Table 2) were cell movement related which might be due to that during RBC invasion there is cell membrane manipulation (Keeley and Soldati, 2004; Cowman and Crabb, 2006) present that requires cytoskeleton control. At 20 dpi, the genes found to be significantly up regulated did not appear to have a common denominator (Table 2), however an issue that arises is that a third of the genes are to-date unannotated and can therefore not be placed in a GO-term analysis.

One factor that might influence the time point analysis by introducing variation in the analysis is if the parasite's gene expression is linked to the parasitemia levels as there is large variation in parasitemia both within 8 dpi and between the two time points. This would happen either if high parasitemia occurs as an effect of gene specific expression patterns or that parasites adjust their gene expression as a result of their parasitemia i.e., their population size. One such reason could be that once the parasite reaches a certain level of parasitemia it starts to produce a higher number of gametocytes in order to be able to transmit to a new host (Bruce et al., 1990; Liu et al., 2011; Schneider et al., 2018).

4.2. *Different gene expression linked to parasitemia levels*

When testing for differences in expression in relation to parasitemia in 8 dpi samples, only the large ribosomal gene was found to be differentially expressed between birds with high and low parasitemia. The main challenge with analysis gene per gene in relation to parasitemia levels is the large number of correlation tests that must be performed and controlled for, potentially creating a huge number of false negatives. One way to simplify the analysis was to use the PCA values against the parasitemia and if there is a correlation between those values continue with investigating the genes that have the highest loading coefficient to the specific principal component value. Although such an analysis reduced the strength of the link between the parasitemia and specific genes, it makes it possible to test if there are specific biological processes, through GO-term enrichment analysis, that are significantly overrepresented in the genes that "builds up" the significant principal component value. In our case we found an aggregation of genes related

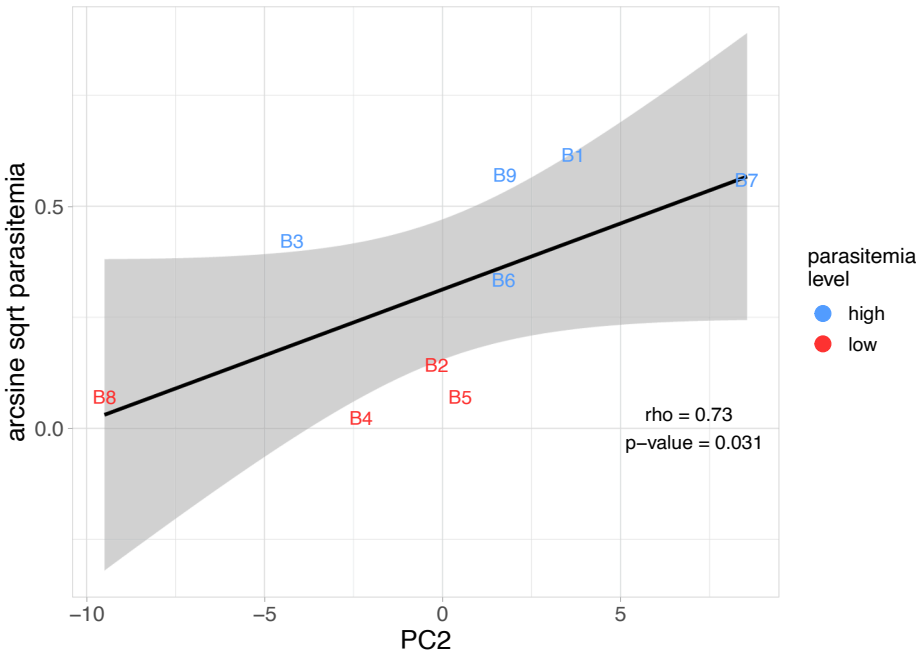

Fig. 4. The correlation between the 2nd principal component in App. Fig. 5 and the parasitemia intensity positively correlated with an R value of 0.73 and p-value of 0.031.

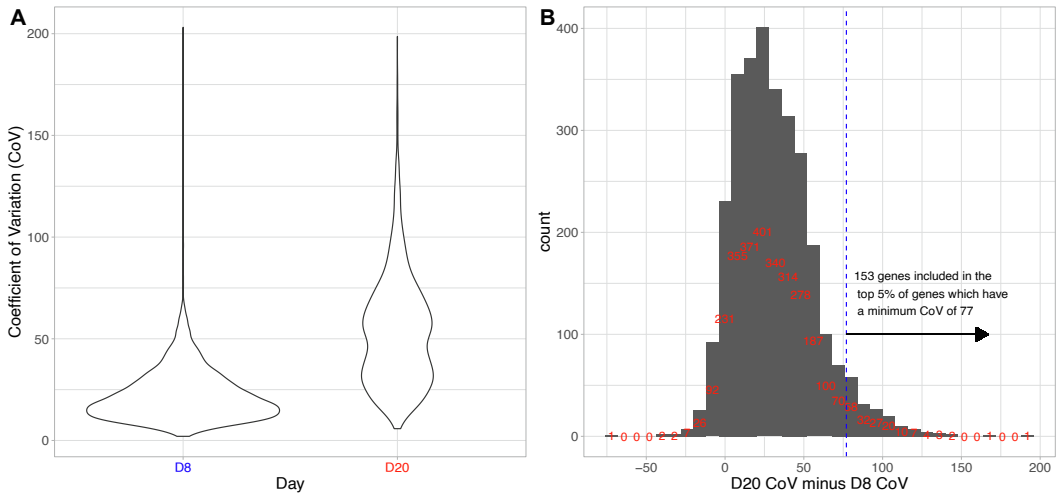

Fig. 5. A: The coefficient of variation (CoV) for each normalised gene count was calculated and plotted to compare expression in 8 dpi (blue) and 20dpi (red) samples. The broader the violin, the greater the number of genes with that CoV. 8 dpi samples tended to have lower CoV, whereas 20 dpi samples have a higher CoV. **B:** The difference in CoV between 20 dpi and 8 dpi samples was plotted as a histogram. Then the genes with the largest difference (at 5% of all genes) were selected for GO-term enrichment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

to cell replication and cell movement among the genes with a loading factor associated with high parasitemia. This might be explained by the significant exploitation of RBC's by parasites in high parasitemia samples, where high parasitemia requires, at the very least, cellular replication and invasion into new host RBC's. In contrast, the GO terms linked to low parasitemia were involved in cellular metabolism.

When doing the parasitemia correlation analysis, one unexpected set of genes, i.e., genes linked to meiosis, were found to be significantly upregulated. The presence of meiotic gene expression (Tat binding protein 1(TBP-1)-interacting protein, putative, and meiotic recombination protein DMC1, putative) in the dataset is unexpected, given meiosis occurs exclusively in the mosquito. The annotations of the genes are derived from their homology with genes found in the *P. falciparum* genome, which diverged on the order of 10 million years ago (Böhme et al., 2018). These genes may have evolved other functions related to but independent of meiosis, such as homologous DNA repair or the regulation thereof. The meiotic recombination protein DMC is part of the Ortholog Group OG5_126834 (https://plasmodb.org/plasmo/app/record/gene/PRELSG_0507200#category:evolutionary-biology) which is annotated either as a recombination repair or meiotic recombination protein within the *Plasmodium* genus. TBP-1 interacting protein is associated with meiosis because it has a winged helix-like DNA-binding domain which is involved in homologous pairing (https://plasmodb.org/plasmo/app/record/gene/PF3D7_1206500#category:functionanalysis, http://www.ebi.ac.uk/interpro/entry/InterPro/IPR010776/).

4.3. Increased variation in gene expression over the course of infection

In our data there was no initial indication that gene expression was individual-specific as there was no apparent grouping of samples based on individuals in the PCA analysis as had been found in other studies (Videvall et al., 2017). However, although the gene expression doesn't stay similar in the same host individual over time doesn't mean that the parasite doesn't fine tune its gene expression to individual hosts. We hypothesised that genes that are interacting with individual hosts (i.e., exhibit gene-gene interactions) would also be the genes that show largest variation between the infected hosts at a given time point. In fact, genes that are not linked to gene-gene interactions between the host and the parasite, e.g., such as house-keeping genes, would keep a similar gene expression independent on the host it infects.

When visualising the CoV (Fig. 5A) the number of genes that showed large variation was very different between 8 dpi and 20 dpi, where the 20 dpi samples showed a much larger number of genes with a high CoV value. This pattern was mirrored in the PCA plots (Fig. 3) where the area the samples occupied increased at 20 dpi as a result of a more diverged overall gene expression profile. Why this pattern emerges one can only speculate, but we hypothesise that it is due to one, or a combination, of two factors; A) as the infection progresses the parasites that comprises the population within the infection becomes unsynced in their life stages where different proportions of the population being at different life stages in the different individual populations thus creating an increase in gene expression variation. B) that the parasites in fact tune their gene expression profiles to fit their individual hosts. C) It also might be that secondary *exo*-erythrocytic stages already appeared at 20 dpi (they are induced by blood stages, but it takes some time for their development, and they should be absent at 8 dpi). As a result, transcriptomes at 20 dpi might include new genes, which i) came from tissue stages or ii) blood stages, which are initiated by merozoites from *exo*-erythrocytic stages (Ilgunas et al., 2016). Early during the infection this might not be as pronounced as later when the immune system of the host is suppressing (or not suppressing) the parasites. Since, in this experiment, the hosts are genetically heterogenous, the parasites will respond differently depending on the individual host thus increasing the CoV as the infection progresses. One way to address this question in the future is to link the specific genotype or gene expression of the host to that of the parasite. By doing so it would be possible to test if parasites that infect genetically similar hosts also exhibit more similar gene expression profiles.

4.4. Comparing P. relictum and P. ashfordi

Both *P. ashfordi* and *P. relictum* are considered to be host generalist as they have been found to infect multiple host species (Bensch, Hellgren and Pérez-Tris, 2009). However, *P. relictum* have temperate and subtropical transmission whereas to date no indications have been found that *P. ashfordi* have transmission in temperate regions. These species were selected for study because they were some of the few isolates we had at the time.

When this paper compares the high-level transcriptome profiles of *P. relictum* and *P. ashfordi,* the differences in speed of reproduction should be considered. The sampling days chosen for *P. ashfordi* were 21 and 31, because they represented peak and post-peak parasitemia, and days 8 and 20 in this study were chosen for the same reason.

It is therefore possible that the host responses might be different between the two studies where the adaptive immune system of the host would have had longer time to become activated in both sampling points in Videvall et al (2017) study hosts compared to this study. Therefore, in the future when trying to study differences and similarities at the genetic level in this system it would be important to use parasite species that show similar temporal dynamics in the host to reduce any potential differential effects the host immune system may have on the parasite.

The study here and the study of Videvall et al (2017) represent some of the first studies that investigate the genetic responses during the vertebrate phase of the infection cycle. Although that the two studies differed in their findings where Videvall et al found that the parasites tended to adjust their gene expression to fit individual host whereas this study could not find similar patterns one should bear in mind that none of the studies can be used to draw any general conclusion of the system but merely represent responses in separate species of parasites. *Plasmodium relictum* and *P. ashfordi* exhibit a sequence divergence of 7.7% of the cyt b gene. Even when implementing the fastest suggested rate of evolution of 1.3% per million year would mean that they have evolved independent of each other for more than 5 million years (Bensch et al., 2013). It is therefore not unrealistic to assume that these two different species of malaria parasites also have evolved large differences in how they interact with their hosts and how the host in turn is responding to the parasites. Therefor we belief that as intriguing it would have been to find general patterns that span across greater than 5 millions of years of independent evolution it is equally fascinating if we can establish different patterns of host-parasite interactions at the genetic level in this group of parasites. As this system exhibits huge diversity in terms of hosts and parasite species, together with frequent host shifts, we can, by studying both closely related and unrelated parasites, unlock the potential to study whether similarity in hosts species or the effect of a common ancestry of the parasites drives the observed differences.

5. Conclusion

In this study we have found that the avian malaria parasite *P. relictum* show, although not extensively, less variation in gene expression prior to peak infection compared to past peak infection and that some gene functions were linked to the level of parasitemia. We further show that the overall gene expression wasn't tuned to single individuals as indicated in prior studies. However, over the course of an infection, the gene expression diverged between individual infected hosts thus creating an increase in the variation of expression within the experimental group compared to the earlier time point in the experiment. We speculate that this either is an effect of the parasite populations within the infections which are becoming more "un-synched" later in the infections or that the parasites in fact adapt their gene expression to the single individual hosts over the course of infection.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank the staff of the Biological Station "Rybachy" of the Zoological Institute of Russian Academy of Science. Funding for VP from European Social Fund (project No 09.3.3-LMT-K-712-01-0016) under grant agreement with the Research Council of Lithuania (LMTLT).

Funding for OH, the Swedish Research Council (grant 2016-03419 and 0000-0002-4062-7276) and DA Swedish Research Council (grant 0000-0003-4713-0032).

Data Availability

All sequence data have been deposited in the European Nucleotide Archive at EMBL-EBI (PRJEB42477).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2021.145723.

References

Atkinson, C.T., LaPointe, D.A., 2009. Introduced avian diseases, climate change, and the future of Hawaiian honeycreepers. J. Avian Med. Surg. 23 (1), 53–63. https://doi.org/10.1647/2008-059.1.

Aurrecoechea, C., et al., 2009. PlasmoDB: A functional genomic database for malaria parasites. Nucleic Acids Res. https://doi.org/10.1093/nar/gkn814.

Baker, D.A., 2010. Malaria gametocytogenesis. Mol. Biochem. Parasitol. https://doi.org/10.1016/j.molbiopara.2010.03.019.

Bensch, S., et al., 2004. Linkage between nuclear and mitochondrial DNA sequences in avian malaria parasites: Multiple cases of cryptic speciation? Evolution 58 (7), 1617–1621. https://doi.org/10.1111/j.0014-3820.2004.tb01742.x.

Bensch, S., et al., 2013. 'How can we determine the molecular clock of malaria parasites? Trends Parasitol. 363–369. https://doi.org/10.1016/j.pt.2013.03.011.

Bensch, S., Hellgren, O., Pérez-Tris, J., 2009. MalAvi: a public database of malaria parasites and related haemosporidians in avian hosts based on mitochondrial cytochrome b lineages. Mol. Ecol. Resour. 9 (5), 1353–1358. https://doi.org/10.1111/j.1755-0998.2009.02692.x.

Böhme, U., et al., 2018. Complete avian malaria parasite genomes reveal features associated with lineage-specific evolution in birds and mammals. Genome Res. Cold Spring Harbor Lab. Press 28 (4), 547–560. https://doi.org/10.1101/gr.218123.116.

Bozdech, Z., et al., 2008. The transcriptome of Plasmodium vivax reveals divergence and diversity of transcriptional regulation in malaria parasites. PNAS 105 (42), 16290–16295. https://doi.org/10.1073/pnas.0807404105.

Bruce, M.C., et al., 1990. Commitment of the malaria parasite plasmodium falciparum to sexual and asexual development. Parasitology Cambridge Univ. Press 100 (2), 191–200. https://doi.org/10.1017/S0031182000061199.

Butlin, R., Thompson, J.N., 1995. The Coevolutionary process. J. Anim. Ecol. 121–133. https://doi.org/10.2307/5910.

Churcher, T.S., et al., 2017. Probability of transmission of malaria from mosquito to human is regulated by mosquito parasite density in naïve and vaccinated hosts. PLOS Pathogens 13 (1), e1006108. https://doi.org/10.1371/journal.ppat.1006108. Edited by E. Wenger. Public Library of Science.

Cowman, A.F., Crabb, B.S., 2006. Invasion of red blood cells by malaria parasites. Cell 124 (4), 755–766. https://doi.org/10.1016/j.cell.2006.02.006.

Dimitrov, D., et al., 2015. Plasmodium spp.: An experimental study on vertebrate host susceptibility to avian malaria. Exp. Parasitol. 148, 1–16. https://doi.org/10.1016/j.exppara.2014.11.005.

Drovetski, S.V., et al., 2014. Does the niche breadth or trade-off hypothesis explain the abundance-occupancy relationship in avian Haemosporidia? Mol. Ecol. Blackwell Publishing Ltd 23 (13), 3322–3329. https://doi.org/10.1111/mec.12744.

Fix, A.S., et al., 1988. Plasmodium relictum as a cause of avian malaria in wild-caught magellanic penguins (Spheniscus magellanicus). J. Wildl. Dis. 24 (4), 610–619. https://doi.org/10.7589/0090-3558-24.4.610.

Hellgren, O., et al., 2015. Global phylogeography of the avian malaria pathogen Plasmodium relictum based on MSP1 allelic diversity. Ecography 38 (8), 842–850. https://doi.org/10.1111/ecog.01158.

Ilgunas, M., et al., 2016. Mortality and pathology in birds due to Plasmodium (Giovannolaia) homocircumflexum infection, with emphasis on the exoerythrocytic development of avian malaria parasites. Malar. J. BioMed Central 15 (1), 256. https://doi.org/10.1186/s12936-016-1310-x.

Keeley, A., Soldati, D., 2004. The glideosome: a molecular machine powering motility and host-cell invasion by Apicomplexa. Trends Cell Biol. 14 (10), 528–532. https://doi.org/10.1016/j.tcb.2004.08.002.

Liu, Z., Miao, J., Cui, L., 2011. Gametocytogenesis in malaria parasite: Commitment, development and regulation. Future Microbiol. 1351–1369. https://doi.org/10.2217/fmb.11.108.

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15 (12), 1–21. https://doi.org/10.1186/s13059-014-0550-8.

Palinauskas, V., et al., 2008. Plasmodium relictum (lineage P-SGS1): Effects on experimentally infected passerine birds. Exp. Parasitol. 120 (4), 372–380. https://doi.org/10.1016/j.exppara.2008.09.001.

Penczykowski, R.M., Laine, A.-L., Koskella, B., 2016. Understanding the ecology and evolution of host-parasite interactions across scales. Evol. Appl. 9 (1), 37–52. https://doi.org/10.1111/eva.12294.

Ringnér, M. (2008) What is principal component analysis?, Nat. Biotechnol. Available at: http://www.nature.com/naturebiotechnology (Accessed: 28 August 2020).

Roth, A., et al., 2018. Unraveling the Plasmodium vivax sporozoite transcriptional journey from mosquito vector to human host. Sci. Rep. Springer, US 8 (1), 1–20. https://doi.org/10.1038/s41598-018-30713-1.

Schneider, P., et al., 2018. 'Adaptive plasticity in the gametocyte conversion rate of malaria parasites. PLOS Pathogens 14 (11), e1007371. https://doi.org/10.1371/journal.ppat.1007371. Edited by R Tewari. Public Library of Science.

Thomas, F., et al., 2002. Understanding parasite strategies: a state-dependent approach? Trends Parasitol. England 18 (9), 387–390. https://doi.org/10.1016/s1471-4922(02)02339-5.

Valkiūnas, G., et al., 2018. Characterization of Plasmodium relictum, a cosmopolitan agent of avian malaria. Malar. J. BioMed Central Ltd. 17 (1), 184. https://doi.org/10.1186/s12936-018-2325-2.

Valkiunas, G., Iezhova, T.A., 2018. Keys to the avian malaria parasites. Malar. J. BioMed Central 17 (1), 17–19. https://doi.org/10.1186/s12936-018-2359-5.

Videvall, E., et al., 2017. The transcriptome of the avian malaria parasite Plasmodium ashfordi displays host-specific gene expression. Mol. Ecol. Blackwell Publishing Ltd 26 (11), 2939–2958. https://doi.org/10.1111/mec.14085.

Videvall, E., et al., 2020. Host transcriptional responses to high- and low-virulent avian malaria parasites. Am. Nat. 195 (6), 1070–1084. https://doi.org/10.1086/708530.

Yang, A.S.P., et al., 2017. Cell traversal activity is important for plasmodium falciparum liver infection in humanized mice. Cell Rep. ElsevierCompany 18 (13), 3105–3116. https://doi.org/10.1016/j.celrep.2017.03.017.