Trends in Parasitology

CellPress
REVIEWS

Review

Genomic Advances in Avian Malaria Research

Elin Videvall [1,*]

Haemosporidian parasites causing malaria-like diseases in birds are globally distributed and have been associated with reduced host fitness and mortality in susceptible bird species. This group of parasites has not only enabled a greater understanding of host specificity, virulence, and parasite dispersal, but has also been crucial in restructuring the evolutionary history of apicomplexans. Despite their importance, genomic resources of avian haemosporidians have proved difficult to obtain, and they have, as a result, been lagging behind the congeneric *Plasmodium* species infecting mammals. In this review, I discuss recent genomic advances in the field of avian malaria research, and outline outstanding questions that will become possible to investigate with the continued successful efforts to generate avian haemosporidian genomic data.

Highlights

Avian malaria research has experienced multiple challenges of generating genomic data, primarily due to host contamination.

One *Haemoproteus* and two avian *Plasmodium* genome sequences have been assembled.

One *Haemoproteus* and four avian *Plasmodium* transcriptome sequences have been assembled, with one incorporating gene expression levels.

Targeted multilocus sequencing of specific genes has proved useful to generate phylogenetic haemosporidian data.

The transcriptional responses of hosts to avian malaria infection have been investigated in only one study so far.

Avian Malaria Parasites

The haemosporidian parasites infecting birds (Apicomplexa: Haemosporida), commonly referred to as avian malaria parasites, can be separated into three related genera: *Plasmodium*, *Haemoproteus*, and *Leucocytozoon*. Recent studies have, however, challenged this traditional classification, as *Plasmodium* seems to be polyphyletic, and the subgenus *Parahaemoproteus* is potentially highly divergent from *Haemoproteus* [1–3]. With an increasing number of taxa sampled and sequences generated, these taxonomic conflicts will hopefully be better resolved in the near future [4]. Until then, I will refer to the original genera *Plasmodium*, *Haemoproteus*, and *Leucocytozoon* in this study.

Avian malaria parasites differ in their biology, morphology, and ecology [5], but all depend on a bird and a dipteran vector to complete their life cycle. These blood parasites are widespread and have been the focus of an increasing number of studies [6], primarily in naturally infected wild birds. Yet avian *Plasmodium* parasites have also been successfully studied in experimentally infected birds and mosquitoes (see e.g., [7–13]). The acute phase of infection with avian haemosporidians has, in some instances, led to high mortality [14], particularly in naive bird species without previous parasite exposure [10,15,16]. The effects of milder chronic infections in birds have, in some studies, been associated with reduced reproductive fitness [17–19], and in other cases the results have been more ambiguous [20–22] (but see also [23–25]). Because of their large variation in host ranges, we are slowly but steadily getting a better understanding of the specificity of parasite lineages within vertebrate [26–30] and invertebrate hosts [31–33].

With increased sampling efforts for birds and vectors world-wide, the global pattern of avian haemosporidian distributions is beginning to unravel [34–36]. Almost all studies characterizing the ecology and biogeography of avian malaria parasites have identified the species using either morphological characters seen in blood smears or sequence information based on a short fragment of the mitochondrial cytochrome b (cyt b) gene. The database MalAvi congregates

[1]Center for Conservation Genomics, Smithsonian Conservation Biology Institute, Washington, DC, USA

*Correspondence:
elin.videvall@gmail.com (E. Videvall).

 https://doi.org/10.1016/j.pt.2018.12.005

cyt b sequences of these parasites and has, at the time of writing, amassed more than 3100 unique lineages [37], indicating a potentially massive diversity of cryptic species. We know extremely little, however, of how this cyt b diversity transpires into genome-wide diversity [38–40], and how the genomes of avian haemosporidian parasites compare to each other [41,42]. Furthermore, even though gene expression has been intensively studied from multiple angles in human malaria parasites [43–45], similar information for avian haemosporidians has been very scarce [46].

In this review, I discuss the current challenges of sequencing genomes and transcriptomes of avian haemosporidians, present recent genomic advances made in spite of these challenges, and highlight potentially useful methodologies that may assist in generating genome-wide haemosporidian sequence data.

Challenges of Sequencing Avian Haemosporidian Genomes

It has proved very difficult to generate genomic sequence data from avian haemosporidians for a number of reasons (Figure 1). First of all, birds have nucleated red blood cells, which means that any procedures to deplete white blood cells [47] to increase the concentration of erythrocytes (as is common practice for primate malaria research) will not be as advantageous for avian haemosporidians, and downright detrimental for leucocyte-infecting parasites like *Leucocytozoon*. Second, wild birds with chronic infections have normally very low parasitemia levels (especially true for *Plasmodium* infections); often ranging from 0.0001 to 0.1% of the red blood cells infected [24,48,49]. Third, while bird genomes typically encompass ~1200 Mb, avian haemosporidian genomes are 50 times smaller, only ~23 Mb in size [41,42]. Fourth, haemosporidian cells in blood carry only one set of chromosomes (haploid) while the host cells have two (diploid). Fifth, the amount of blood that is possible to collect from most passerines is small, and can therefore restrict the use of high-throughput sequencing techniques that require

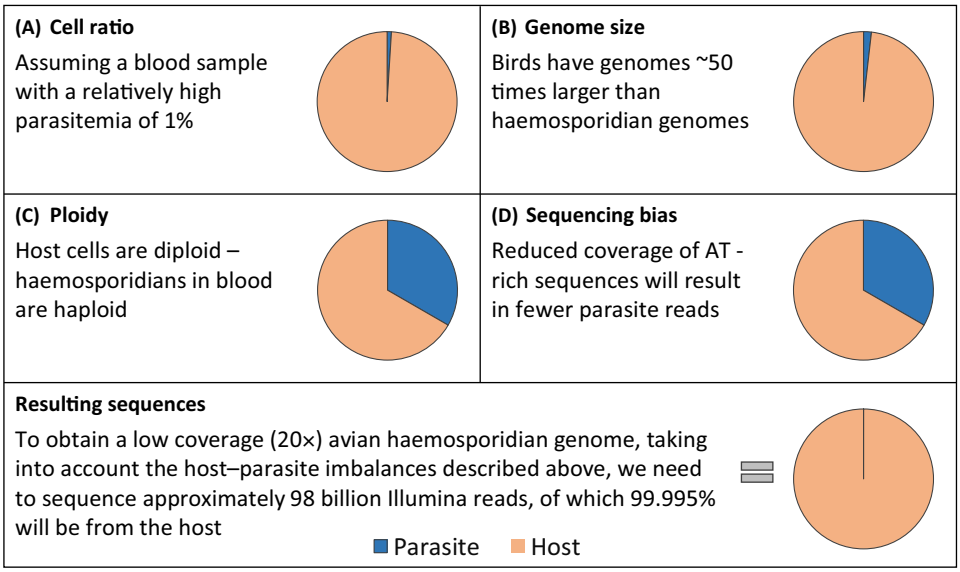

Figure 1. A Description of Challenges to Obtain Haemosporidian Genomic Sequence Data from Bird Blood. Taking into account different factors (A–D), the total number of sequences necessary for a low-coverage haemosporidian genome would come close to 100 billion. There are, however, various methods proposed to increase the parasite ratio in most of these aspects, thereby reducing the number of sequence reads needed (see section 'Promising Solutions to Facilitate Avian Haemosporidian Genome Sequencing').

large input quantities of high-quality parasite DNA [50,51]. Lastly, the genome sequences of haemosporidian parasites are in general extraordinarily AT-rich, with the avian *Plasmodium* species having an even more extreme genic AT-content compared to the already highly AT-rich mammalian *Plasmodium* parasites [52]. This base composition has led to complications during sequencing, largely due to the inherent difficulties working with nucleotides of high homogeneity [50,53–55], but it has also complicated genome assemblies [56] and phylogenetic analyses [2,57].

With these outlined challenges in mind, we can start to estimate from a hypothetical blood sample approximately how much we would need to sequence in order to obtain enough data for a low-coverage avian haemosporidian genome, broadly assuming a relatively high parasitemia level of 1%, one parasite cell per infected red blood cell, no white blood cells, and a rough 50% reduction in read coverage for the AT-rich parasite DNA [55]. The cell ratio, together with the genome size and ploidy differences, means that we will need to generate approximately 98 billion Illumina reads (100 bp) in order to obtain enough data for a low-coverage (20×) haemosporidian genome. This large number results from the fact that 99.99% of all sequences will be derived from host DNA (Figure 1). In reality, some infected cells are likely to contain multiple parasites; however, the sample is also likely to contain several white blood cells, making these details unnecessary to include in the calculations. Even if we disregard certain aspects of the assumption, such as the potential sequencing bias of standard next-generation sequencing methods, the cell ratio and genome size differences between host and parasite alone result in an insurmountable sequencing depth. As a consequence, any potential method that is successful at sorting parasite cells or enriching for parasite DNA will be extremely valuable to implement before sequencing avian haemosporidian genomes. Similarly, because the resulting sequence data will be derived from both host and parasite, bioinformatic filtering methods to remove host contamination after sequencing will be necessary to obtain clean parasite assemblies (Figure 2).

The difficulties in obtaining genomic DNA from avian malaria parasites were unequivocally demonstrated by Lutz *et al*. who attempted to sequence the genome of *Plasmodium relictum* SGS1 [58]. To circumvent the problems with host sequences, the authors used laser capture microdissection under a microscope to isolate parasite cells from individual host erythrocytes [59], which they subsequently sequenced. Despite the efforts to single out parasite cells with this technique, Lutz *et al*. managed to obtain parasite DNA in only 0.07% of the sequences, while 99.93% consisted of contamination from the host and other sources [58].

Avian Haemosporidian Genome Assemblies

In 2016, Bensch *et al*. published the very first genome sequence of an avian haemosporidian, *Haemoproteus tartakovskyi* [42]. This milestone was accomplished thanks to an innovative method by Palinauskas *et al*. to separate haemosporidian cells from bird blood cells [60], thereby greatly increasing the parasite-to-host sequence ratio. With this method, blood is drawn from the bird into sodium citrate solution, gently mixed, and then left exposed to air for 4 min. The brief air exposure induces exflagellation in the *Haemoproteus* gametocytes (Figure 3) [60], since the circumstances mimic blood ingestion by a vector. The blood sample is subsequently centrifuged and the plasma (which now contains numerous microgametes) is collected, and later used for sequences [42,60]. By using this exflagellation technique, Bensch *et al*. successfully obtained approximately 50% of the 454-sequencing reads from the parasite [42]. The finished *H. tartakovskyi* genome assembly was used to compare synteny and similarities with other genomes from apicomplexan parasites, and a large number of genes (~40%) was identified as being entirely unique to *H. tartakovskyi* [42]. Furthermore, the genome

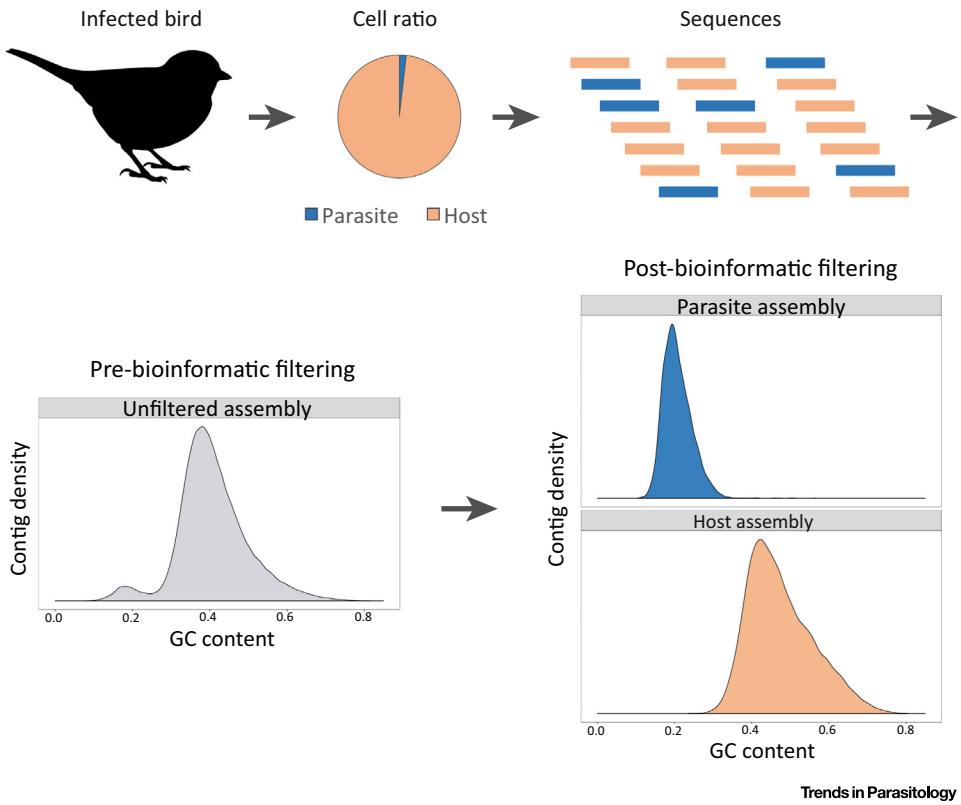

Figure 2. Schematic Example of a Possible Process to Build Genome/Transcriptome Sequences of Avian Haemosporidians. Starting with an infected bird, the skewed parasite/host cell ratio (together with the other biases highlighted in Figure 1) will affect the resulting ratio of parasite/host sequence reads. Subsequent bioinformatic filtering of an unfiltered meta-assembly is therefore needed to separate the host assembly from the parasite assembly. The density plots are adapted and reprinted from Videvall *et al*. [46].

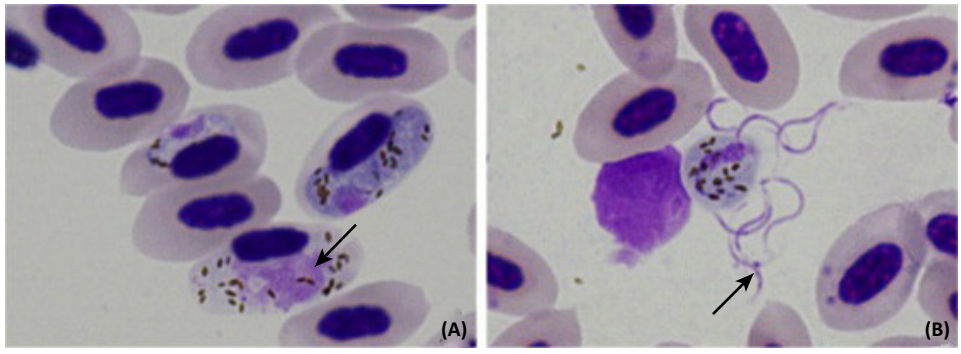

Figure 3. Example of a Method Designed to Separate Parasite Cells from Host Cells. Brief air exposure of bird blood infected with *Haemoproteus* parasites (A) induces exflagellation and leads to the production of extracellular microgametes (B). The blood sample is subsequently centrifuged to separate host blood cells from parasite microgametes. Figures depict Giemsa-stained thin blood films, and arrows point at parasite nuclei. Adapted from Palinauskas *et al*. [60].

of *H. tartakovskyi*, in combination with a transcriptome of *Plasmodium ashfordi* [46], allowed us to use genome-wide sequence resolution to construct phylogenies of Apicomplexa. These analyses showed that *Haemoproteus* is a suitable outgroup to *Plasmodium* and that the human parasite *Plasmodium falciparum* is indeed more closely related to the other mammalian *Plasmodium* species [42], and not the avian *Plasmodium* species.

The second and third avian haemosporidian (and the first avian *Plasmodium* species) to have their full genome sequences published were the chicken parasite *Plasmodium gallinaceum* 8A and the passerine parasite *Plasmodium relictum* DONANA05 [41] (a lineage closely related to the widespread SGS1). Böhme *et al*. successfully managed to generate enough sequencing material for these two parasite genome sequences by dissecting out 50 heavily infected mosquito midguts with >100 oocysts (*P. relictum*) [41], by enriching for nonmethylated DNA [61] (*P. gallinaceum*), and by using a whole-genome amplification technique [62] (*P. gallinaceum*). These new genome sequences were similarly used in phylogenomic trees to evaluate evolutionary relationships within the *Plasmodium* genus, and Böhme *et al*. also found the bird parasites to be positioned outside the mammalian-infecting *Plasmodium* clade. Furthermore, some genes that had previously been assumed to be specific to the primate-infecting *Laverania* clade of *Plasmodium* were found within the genomes of these two bird parasites (Figure 4) [41]. DNA segments entirely unique to avian *Plasmodium* genomes (as far as we know with available data) included 50 genes and multiple transposable elements [41], potential future targets in trying to pinpoint the underlying molecular mechanisms of mammal-specificity and bird-specificity within *Plasmodium*.

Avian Haemosporidian Transcriptome Assemblies

In 2014, prior to any genome assemblies, Lauron *et al*. generated the very first genome-wide sequences from an avian haemosporidian when they assembled a blood transcriptome of the chicken parasite *Plasmodium gallinaceum* [63,64] (Table 1). Despite difficulties in removing host transcripts from the final transcriptome assembly, the parasite transcripts provided valuable sequence information on genes involved in the purine salvage pathway, and on genes involved

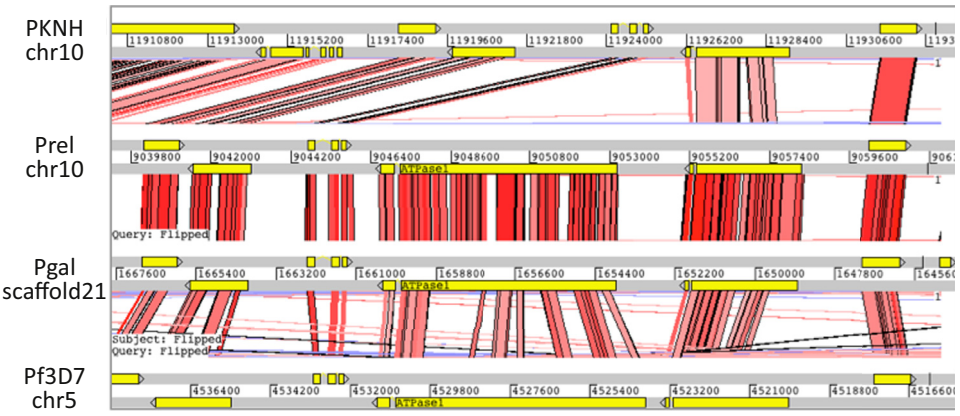

Figure 4. An Example of How Avian Haemosporidian Genomes Can Expand Our Current View of Primate *Plasmodium* Evolution. The gene ATPase1 was previously believed to be specific to the *Laverania* clade of *Plasmodium*. However, recent whole-genome sequencing of *P. relictum* and *P. gallinaceum* has demonstrated that ATPase1 is present in avian *Plasmodium* genomes as well. In the figure, gray bars represent forward and reverse DNA strands, and the yellow boxes represent genes. The red blocks show sequence similarity between species (tBLASTX). Abbreviations: PKNH, *P. knowlesi*; Prel, *P. relictum*; Pgal, *P. gallinaceum*; Pf3D7, *P. falciparum*. Adapted from Böhme *et al*. [41].

Table 1. Avian Haemosporidian Genome and Transcriptome Assemblies

Species (lineage)	Assembly	From host	Tissue	Experimentally infected	Host contigs removed	Total assembled bases excl. gaps	Assembly available at	Refs
Plasmodium relictum (DONANA05)	Genome	*Culex pipiens*	Mosquito midgut	Yes	Yes	22 579 648	NCBI, PlasmoDB	[41]
Plasmodium gallinaceum (8A)	Genome	*Gallus gallus*	Whole blood	Yes	Yes	23 877 187	NCBI, PlasmoDB	[41]
Plasmodium gallinaceum	Transcriptome	*Gallus gallus*	Whole blood	Yes	Partially	N/A	Supp. Info.	[63]
Plasmodium ashfordi (GRW2)	Transcriptome	*Carduelis spinus*	Whole blood	Yes	Yes	9 010 380	MalAvi, Supp. Info.	[46]
Plasmodium delichoni (COLL6)	Transcriptome	*Carduelis spinus*	Whole blood	Yes	Yes	5 680 962	Supp. Info.	[66]
Plasmodium homocircumflexum (COLL4)	Transcriptome	*Carduelis spinus*	Whole blood	Yes	Yes	17 175 763	Supp. Info.	[66]
Haemoproteus tartakovskyi (SISKIN1)	Genome	*Carduelis spinus*	Blood plasma	No	Yes	22 756 238	NCBI, MalAvi	[42]
Haemoproteus columbae	Transcriptome	*Columba livia*	Whole blood	No	Yes	13 214 764	NCBI	[67]
Leucocytozoon toddi	Transcriptome	*Buteo buteo*	Whole blood + pooled organs	No	No	N/A	N/A, only raw reads	[65]

in the invasion of host cells [63,64]. Pauli *et al*. similarly used RNA sequencing on whole blood and on pooled organ tissue from juvenile common buzzards naturally infected with a *Leucocytozoon* parasite [65]. The difficulty of filtering host transcripts from the assembly led the authors to report it as a dual transcriptome assembly, a collection of both bird and parasite contigs [65]. Following these two partial transcriptome assemblies, Videvall *et al*. took on the challenge of building a fully filtered parasite transcriptome assembly of the virulent *Plasmodium ashfordi* GRW2 in experimentally infected Eurasian siskins [46]. *De novo* assemblies of parasites are inherently more complicated to construct when reference genomes of both host and related parasites are nonexistent. However, by filtering the assembly in a series of bioinformatic steps, we successfully managed to remove the transcripts originating from the host and produced a clean parasite transcriptome (Figure 2) [46]. This assembly allowed us to evaluate parasite gene expression over time, showing that the clonal *Plasmodium* parasites exhibited differential gene expression depending on which host individual they had infected [46].

Encouragingly, more transcriptome assemblies of avian haemosporidians are now gradually being reported (Table 1), as RNA sequencing seems to provide a better host–parasite sequence ratio compared to DNA sequencing (personal observation). Recently, two *Plasmodium* transcriptome sequences were assembled by Weinberg *et al*. who also utilized experimentally infected Eurasian siskins as host species [66]. In this study, the authors evaluated transcripts recovered from two closely related species, *P. delichoni* and *P. homocircumflexum*, and could clearly show that their sequence similarity searches against other *Plasmodium* species became significantly improved with more avian haemosporidian sequence data included [66]. The first transcriptome assembly of a *Haemoproteus* parasite (*H. columbae*)

was also recently reported by Toscani Field *et al*. who sequenced the blood of a wild-caught rock pigeon [67]. By using 600 genes from the transcriptome assembly, combined with previously published haemosporidian genomic data [42], Toscani Field *et al*. built phylogenies showing that the genus *Haemoproteus* (including *Parahaemoproteus*) is monophyletic [67].

Targeted Sequence Approaches

Due to the aforementioned difficulties in sequencing whole genomes of avian haemosporidians, several studies have taken alternative routes to obtain multigene sequences from multiple taxa in order to construct better phylogenies.

The mitochondrial genome sequence (mtDNA) of haemosporidian parasites is very short (~6 kb) and it contains only three protein-coding genes (cyt b, cox1, and cox3). Prior to any nuclear genomes, complete mtDNA sequences had therefore already been successfully generated from multiple bird-infecting parasites: including *Plasmodium* [68–70], *Haemoproteus* [68,71–73], and *Leucocytozoon* lineages [71,74–76]. The number of complete avian haemosporidian mtDNA sequences has recently more than tripled, thanks to a study by Pacheco *et al*. [3]. The authors generated approximately 56 new avian haemosporidian mtDNA genomes, comprising 28 *Plasmodium*, 25 *Haemoproteus*, and three *Leucocytozoon* lineages, which they subsequently used in phylogenetic analyses [3]. The apicoplast genome sequence, on the other hand, is larger (~35 kb) and contains 30 protein-coding and several noncoding genes [77,78]. Near-complete apicoplast genomes from avian haemosporidians had, prior to the whole-genome assemblies [41], been sequenced only from parasites infecting chickens, specifically *P. gallinaceum* [78] and *Leucocytozoon caulleryi* [77]. The present *H. tartakovskyi* genome assembly does not contain an apicoplast genome sequence because it was generated from microgametes which lack apicoplasts [42]. As a result, avian haemosporidian apicoplast genomes have currently only been sequenced from *P. gallinaceum, P. relictum*, and *L. caulleryi*.

Borner *et al*. used a Sanger sequencing approach to obtain fragments of 21 nuclear genes in seven haemosporidians and used these sequences to build multiple phylogenetic trees [1]. This dataset of 21 genes was subsequently expanded with greater taxon sampling in a study by Galen *et al*. [2] (Figure 5). Both Borner *et al*. and Galen *et al*. could also show in their phylogenetic analyses that the genus *Plasmodium* is indeed polyphyletic, containing nested clades of the mammalian parasites *Nycteria* and *Hepatocystis* [1], and even the genera *Polychromophilus* and *Haemocystidium* when including *Plasmodium* parasites from ungulates and lizards [2]. Recently, Galen *et al*. also found an extensive cryptic species diversity within the *Leucocytozoon* genus by targeting and partially sequencing seven nuclear genes from 69 infected blood samples [79].

A targeted sequence capture approach based on the genome of *H. tartakovskyi* [42] was developed by Huang *et al*. in 2018 to sequence haemosporidian genes from naturally infected birds [80]. Of the 2565 exons tested, the authors obtained on average 82% of all nucleotides covered in four samples containing the same lineage as the reference genome, though this number jumped to 99% when excluding one poorly sequenced sample. Gradually fewer sequences were recovered with increasing cyt b phylogenetic distance from the reference genome: 19% nucleotide coverage in lineages that were 4.2% distant, to 0.9% coverage in the more genetically distant lineages (>7%) [80]. This sequence capture method was subsequently extended by Barrow *et al*. who evaluated a wide range of infected bird samples, some with very low parasitemia levels, and they successfully managed to obtain sequences from >70 haemosporidian loci in 15 of the 51 tested samples [81].

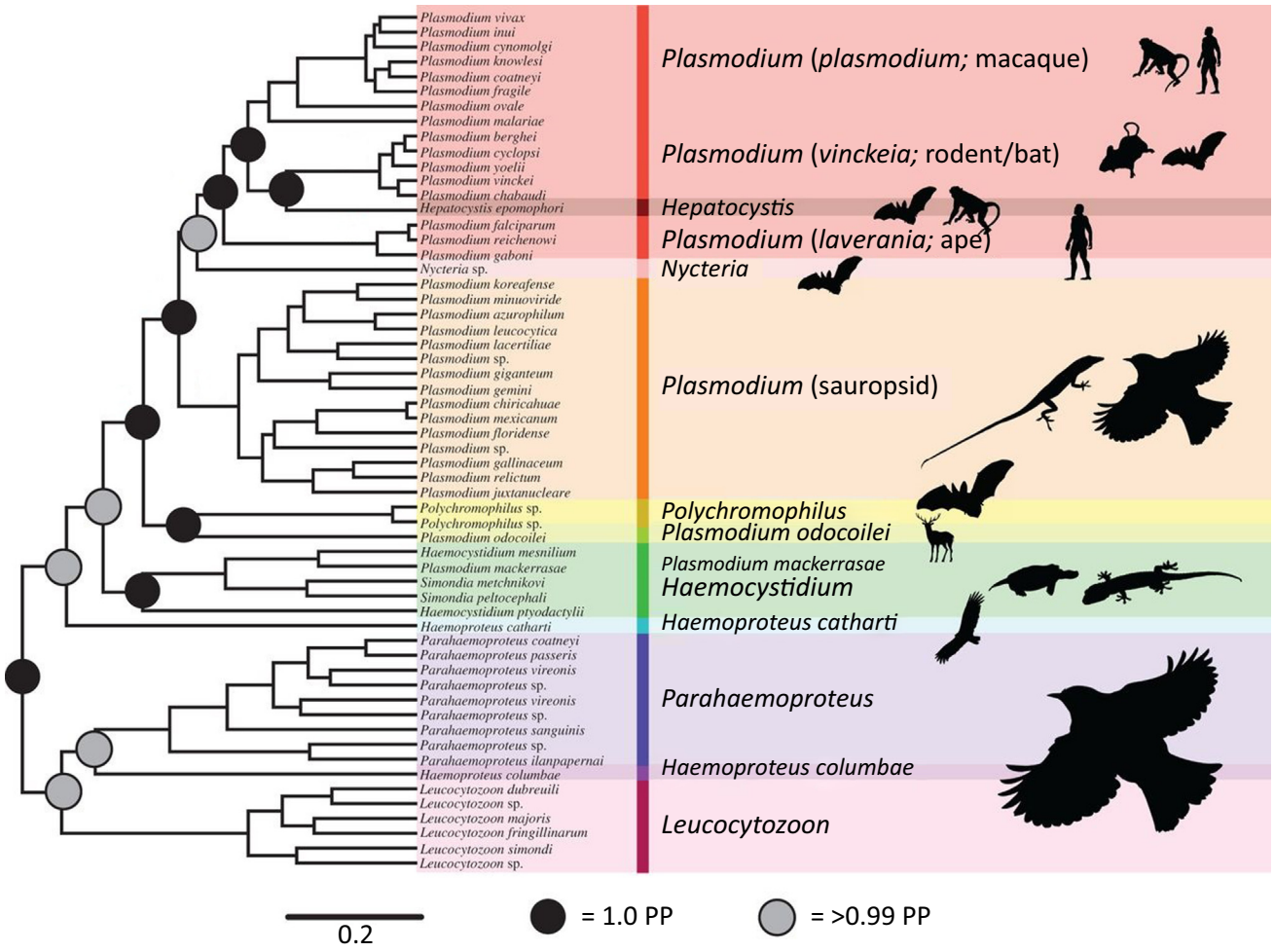

Figure 5. Phylogeny of Haemosporidians Constructed Using Multiple Nuclear Gene Sequences. This phylogeny was constructed from a fully partitioned amino acid dataset and lognormal relaxed molecular clock in BEAST. Silhouettes depict vertebrate host groups. Adapted from Galen *et al*. [2].

Promising Solutions to Facilitate Avian Haemosporidian Genome Sequencing

Despite the numerous challenges in obtaining genomic sequence data of avian haemosporidians (Figure 1), researchers have showed that this is not an impossible feat, as can be seen in Table 1. Knowledge of parasite biology, in combination with innovative laboratory solutions and bioinformatic filtering, has been very important in progressing the field forward. One of the more apparent ways to tackle the first outlined problem of skewed host–parasite cell ratio (Figure 1A) is to use samples from hosts with higher parasitemia levels. However, high parasitemia of avian haemosporidians can be difficult to obtain. Most birds with natural haemosporidian infection show low levels [5,22,24], and setting up an infection challenge experiment requires not only extensive knowledge, suitable facilities, and susceptible uninfected birds, but most importantly ethical permissions. Nevertheless, all current genome and transcriptome assemblies of avian *Plasmodium* spp. have been sequenced from experimentally challenged hosts (Table 1). The current *Haemoproteus* assemblies, on the other hand, have both been obtained from naturally

infected wild-caught birds [42,67]. After the sample has been collected, several laboratory procedures have been proposed to further reduce the number of host cells (Figure 1A). These methods include, for example, the aforementioned induced exflagellation of infected blood to obtain microgametes [60] (Figure 3), disruption of red blood cell membranes using EDTA-20 [82], and laser-capture microdissection of *Plasmodium*-infected erythrocytes [59]. Flow cytometry methods to sort parasite cells in blood have been utilized by researchers studying mammalian-infecting *Plasmodium* for decades [83–85]. In avian malaria research, flow cytometric cell sorting has been evaluated for *Leucocytozoon* [86,87].

To address the challenge of small amounts of haemosporidian DNA (Figure 1B,C), solutions tested primarily on mammalian *Plasmodium* often include methods to enrich samples directly for parasite DNA, such as the depletion of methylated DNA [41,61,88], whole-genome amplification [62], and selective whole-genome amplification [89,90]. Another promising solution to reduce the discrepancy in genome sizes is to use samples from invertebrate (rather than vertebrate) hosts. The dipteran vectors of haemosporidian parasites have much smaller genomes than birds (*Culex* ~45% smaller [91], *Culicoides* ~85% smaller [92]), and a single oocyst in the midgut of a mosquito can contain thousands of sporozoites [93]. Experimental infections of vectors followed by microdissection of oocysts could therefore prove highly valuable for avian malaria genomics. This technique has so far only been successfully implemented for genome-wide sequencing of one species, *P. relictum* [41]. Nonetheless, the large number of parasite sequences obtained with this method made the genome of *P. relictum* the best assembled avian haemosporidian genome as of date and is currently also the only genome with a chromosome-level assembly [41].

Since the specific problem of sequencing AT-biased genomes (Figure 1D) is largely shared with genomic studies of human malaria parasites, avian malaria researchers can greatly benefit from previous work on *P. falciparum* and other AT-biased organisms. PCR-free library preparation steps have been suggested to reduce AT-bias [50,53], although they require relatively large quantities of DNA [54]. Optimized library preparation procedures that include PCR have therefore been developed for *P. falciparum* [54,94], and these methods could prove useful for avian haemosporidians as well. High-throughput sequencing techniques differ in their costs, DNA requirements, data output, error rate, and nucleotide bias, but continue to rapidly improve. Researchers planning to sequence avian haemosporidian genomes are therefore advised to obtain the most up-to-date information on sequencing technologies, including their possibilities and limitations, to make informed decisions. Single-cell genome sequencing of haemosporidians is an intriguing opportunity, which has proved possible for *P. falciparum* [95,96]. However, this approach still suffers from low coverage [95,96], with mostly incomplete genome sequences as a result [97]. Chromatin conformation capture methods, for example Hi-C and 3C, map chromatin contacts to obtain structural information of genomes and organize sequenced contigs. This technique has been used, for example, to characterize the genomic architecture of *P. falciparum* [98] and improve the genome sequence of *P. knowlesi* [99]. However, these types of approaches have primarily used parasite cultures or strategies to first sort parasite cells from host cells, together with pre-existing parasite genome sequences, so they will likely become relevant for avian malaria parasites only after efficient cell sorting/DNA enrichment methods have been implemented.

Parasite Sequences Mined from Genomic Data of Hosts

There are minor opportunities of acquiring avian haemosporidian sequence data as a by-product from next-generation sequencing projects of hosts [100,101]. Borner and Burmester searched through 920 published animal genome/transcriptome assemblies and found that 51 contained sequences from apicomplexan parasites [100]. Similarly, Holmes and Davis Rabosky found

apicomplexan matches in 3–5% of ddRAD sequence data from host tissue samples [101], and Orosz found a species of *Sarcocystis* (Coccidia) in a published bird genome [102]. Obtaining parasite sequence data as a by-catch from sequencing hosts will, however, rarely amount to a large number of sequences. The efficacy of this method is therefore both unreliable and limited. In addition, most published genomic data sets of hosts have likely already been subjected to various filtering and screening checks to remove undesired sequences before being uploaded to sequence repositories. As a result, the possibility of finding parasite sequences from host genome and transcriptome projects is the greatest from raw unfiltered sequence data. Nevertheless, this kind of bioinformatic parasite screening has in some instances expanded the known host range of parasites [103], and with the ongoing massive genome sequencing of all living bird species [104], there is potential for finding additional avian haemosporidian sequences.

Host Transcriptome Responses during Malaria Infection

The possibility of experimentally infecting hosts and monitoring their gene expression over time in response to malaria can yield incredibly useful information of how different hosts vary in their molecular response to different parasite species. To date, no study has yet investigated the transcriptome of dipteran vectors (e.g., mosquitoes) in response to avian haemosporidian infection. The transcriptional response of birds to malaria parasites has been evaluated in only one study so far [105]. In 2015, Videvall *et al*. sequenced and quantified genome-wide gene expression of Eurasian siskins experimentally infected with *P. ashfordi* at three timepoints during the experiment: before infection, during peak parasitemia, and during the decreasing parasitemia stage. Multiple differences in host gene expression between infected and uninfected samples were detected, and genes responding to the parasites were found to be involved in various host functions, such as the immune system, stress response, metabolism, cell death, and miRNA gene regulation [105]. Additional studies in this area of research are desperately needed to achieve a better understanding of the molecular interactions taking place inside hosts during a haemosporidian infection. The resulting data will not only allow for comparative studies, network analyses, and meta-analyses, but also for follow-up investigations of specific genes of interest. Discovering which genes that are ubiquitously responding to malaria similarly across host species, and which genes that are responding in only some host species or to specific parasite lineages, will enable a much greater understanding of the mechanisms behind host susceptibility to blood parasites.

Concluding Remarks

Genomic sequence data of avian haemosporidians have proved highly difficult to generate. Small quantities of DNA, skewed host–parasite cell ratio, nucleated host erythrocytes, 50-fold smaller genome sizes, sequencing biases, and assembly complications due to extreme AT-content constitute some of the current challenges. Nevertheless, researchers have demonstrated that these difficulties are possible to overcome. Genome sequences of two avian *Plasmodium* and one *Haemoproteus*, together with blood transcriptomes of four avian *Plasmodium* and one *Haemoproteus*, have so far been made available (Table 1). One partial low-coverage *Leucocytozoon* transcriptome assembly exists but has not yet been filtered from the host contigs, and only one study so far has evaluated the transcriptome gene expression of an avian haemosporidian.

To gain a better understanding of the biology and evolution of avian malaria parasites, it is imperative that we continue to sequence the genomes or transcriptomes of a wide range of haemosporidians. Innovative methodological solutions will be required to surmount the multitude of challenges in generating genomic data of these organisms, and subsequent bioinformatic filtering procedures will be crucial to remove sequences from the host and other sources of contamination.

Outstanding Questions

How does *Plasmodium* differ genetically from *Haemoproteus* and *Leucocytozoon*? And how does that genetic variation relate to their biological differences?

What are the genetic mechanisms that make some *Plasmodium* species infect birds and other *Plasmodium* species adapted to mammals or lizards?

How do the genomes of specialist haemosporidian parasites compare to those of generalists? (What makes a specialist a specialist?)

How do the genomes of avian haemosporidians interact with the genomes of their hosts?

How strongly do the processes of evolution contribute to the genetic variation and effective population sizes of avian haemosporidians?

How does the gene expression of avian and dipteran hosts during avian malaria infection relate to the pathogenesis of the disease?

Future genomic data of a more diverse set of haemosporidians will allow us to answer outstanding questions on how these organisms evolve and function (see Outstanding Questions). Combining demographic data of avian haemosporidians with population genetics can bring important insights into the genetic mechanisms of host specificity and parasite dispersal. Finally, there are great opportunities to investigate host–parasite interactions using transcriptome expression data, and comparative studies of the genetic structure of different parasite lineages will yield tremendously valuable information on key features of parasite life-history traits.

Acknowledgments

This work was supported by the National Science Foundation (grant no: DEB-1717498). The author wishes to thank Michael Campana, Robert Fleischer, and the anonymous reviewers for kindly providing comments on the paper. Vaidas Palinauskas, Spencer Galen, and Ulrike Böhme generously gave their permission to republish Figures 3–5.

References

1. Borner, J. *et al.* (2016) Phylogeny of haemosporidian blood parasites revealed by a multi-gene approach. *Mol. Phylogenet. Evol.* 94, 221–231
2. Galen, S.C. *et al.* (2018) The polyphyly of *Plasmodium*: comprehensive phylogenetic analyses of the malaria parasites (order Haemosporida) reveal widespread taxonomic conflict. *R. Soc. Open Sci.* 5, 171780
3. Pacheco, M.A. *et al.* (2018) Mode and rate of evolution of haemosporidian mitochondrial genomes: timing the radiation of avian parasites. *Mol. Biol. Evol.* 35, 383–403
4. Outlaw, D.C. and Ricklefs, R.E. (2014) Species limits in avian malaria parasites (Haemosporida): How to move forward in the molecular era. *Parasitology* 141, 1223–1232
5. Valkiunas, G. (2004) *Avian Malaria Parasites and Other Haemosporidia,* CRC Press
6. Rivero, A. and Gandon, S. (2018) Evolutionary ecology of avian malaria: past to present. *Trends Parasitol.* 34, 712–726
7. Cornet, S. *et al.* (2013) Malaria infection increases bird attractiveness to uninfected mosquitoes. *Ecol. Lett.* 16, 323–329
8. Palinauskas, V. *et al.* (2008) *Plasmodium relictum* (lineage P-SGS1): Effects on experimentally infected passerine birds. *Exp. Parasitol.* 120, 372–380
9. Dimitrov, D. *et al.* (2015) *Plasmodium* spp.: An experimental study on vertebrate host susceptibility to avian malaria. *Exp. Parasitol.* 148, 1–16
10. Atkinson, C.T. *et al.* (2000) Pathogenicity of avian malaria in experimentally-infected Hawaii amakihi. *J. Wildl. Dis.* 36, 197–204
11. Cellier-Holzem, E. *et al.* (2010) Effect of repeated exposure to *Plasmodium relictum* (lineage SGS1) on infection dynamics in domestic canaries. *Int. J. Parasitol.* 40, 1447–1453
12. Žiegytė, R. *et al.* (2014) Complete sporogony of *Plasmodium relictum* (lineages pSGS1 and pGRW11) in mosquito *Culex pipiens pipiens* form *molestus*, with implications to avian malaria epidemiology. *J. Parasitol.* 100, 878–882
13. Permin, A. and Juhl, J. (2002) The development of *Plasmodium gallinaceum* infections in chickens following single infections with three different dose levels. *Vet. Parasitol.* 105, 1–10
14. Krams, I.A. *et al.* (2013) Acute infection of avian malaria impairs concentration of haemoglobin and survival in juvenile altricial birds. *J. Zool.* 291, 34–41
15. 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, 610–619
16. Atkinson, C.T. *et al.* (1995) Wildlife disease and conservation in Hawaii: Pathogenicity of avian malaria (*Plasmodium relictum*) in experimentally infected Iiwi (*Vestiaria coccinea*). *Parasitology* 111, S59–S69
17. Knowles, S.C.L. *et al.* (2010) Chronic malaria infections increase family inequalities and reduce parental fitness: Experimental evidence from a wild bird population. *J. Evol. Biol.* 23, 557–569
18. Marzal, A. *et al.* (2005) Malarial parasites decrease reproductive success: An experimental study in a passerine bird. *Oecologia* 142, 541–545
19. Merino, S. *et al.* (2000) Are avian blood parasites pathogenic in the wild? A medication experiment in blue tits (*Parus caeruleus*). *Proc. R. Soc. B Biol. Sci.* 267, 2507–2510
20. Zylberberg, M. *et al.* (2015) *Haemoproteus* infected birds have increased lifetime reproductive success. *Parasitology* 142, 1033–1043
21. Lachish, S. *et al.* (2011) Fitness effects of endemic malaria infections in a wild bird population: The importance of ecological structure. *J. Anim. Ecol.* 80, 1196–1206
22. Bensch, S. *et al.* (2007) Temporal dynamics and diversity of avian malaria parasites in a single host species. *J. Anim. Ecol.* 76, 112–122
23. Schoenle, L.A. *et al.* (2017) An experimental test of the physiological consequences of avian malaria infection. *J. Anim. Ecol.* 86, 1483–1496
24. Asghar, M. *et al.* (2011) Are chronic avian haemosporidian infections costly in wild birds? *J. Avian Biol.* 42, 530–537
25. Asghar, M. *et al.* (2015) Hidden costs of infection: Chronic malaria accelerates telomere degradation and senescence in wild birds. *Science* 347, 436–438
26. Clark, N.J. and Clegg, S.M. (2017) Integrating phylogenetic and ecological distances reveals new insights into parasite host specificity. *Mol. Ecol.* 26, 3074–3086
27. Ellis, V.A. and Bensch, S. (2018) Host specificity of avian haemosporidian parasites is unrelated among sister lineages but shows phylogenetic signal across larger clades. *Int. J. Parasitol.* 48, 897–902
28. Ellis, V.A. *et al.* (2015) Local host specialization, host-switching, and dispersal shape the regional distributions of avian haemosporidian parasites. *Proc. Natl. Acad. Sci. U. S. A.* 112, 11294–11299
29. Svensson-Coelho, M. *et al.* (2016) Resource predictability and specialization in avian malaria parasites. *Mol. Ecol.* 25, 4377–4391
30. Loiseau, C. *et al.* (2012) Host and habitat specialization of avian malaria in Africa. *Mol. Ecol.* 21, 431–441
31. Martínez-de la Puente, J. *et al.* (2011) On the specificity of avian blood parasites: Revealing specific and generalist relationships between haemosporidians and biting midges. *Mol. Ecol.* 20, 3275–3287
32. Njabo, K.Y. *et al.* (2011) Nonspecific patterns of vector, host and avian malaria parasite associations in a central African rainforest. *Mol. Ecol.* 20, 1049–1061
33. Kimura, M. *et al.* (2010) Avian malaria parasites share congeneric mosquito vectors. *J. Parasitol.* 96, 144–151
34. Clark, N.J. *et al.* (2014) A review of global diversity in avian haemosporidians (*Plasmodium* and *Haemoproteus*:

Haemosporida): New insights from molecular data. *Int. J. Parasitol.* 44, 329–338

35. Ellis, V.A. *et al.* (2018) The global biogeography of avian haemosporidian parasites is characterized by local diversification and intercontinental dispersal. *Parasitology* Published online July 16, 2018. http://dx.doi.org/10.1017/S0031182018001130
36. Beadell, J.S. *et al.* (2006) Global phylogeographic limits of Hawaii's avian malaria. *Proc. R. Soc. B Biol. Sci.* 273, 2935–2944
37. Bensch, S. *et al.* (2009) MalAvi: A public database of malaria parasites and related haemosporidians in avian hosts based on mitochondrial cytochrome b lineages. *Mol. Ecol. Resour.* 9, 1353–1358
38. Garcia-Longoria, L. *et al.* (2014) Molecular identification of the chitinase genes in *Plasmodium relictum*. *Malar. J.* 13, 1–5
39. Bensch, S. *et al.* (2004) Linkage between nuclear and mitochondrial DNA sequences in avian malaria parasites: multiple cases of cryptic speciation? *Evolution* 58, 1617–1621
40. Nilsson, E. *et al.* (2016) Multiple cryptic species of sympatric generalists within the avian blood parasite *Haemoproteus majoris*. *J. Evol. Biol.* 29, 1812–1826
41. Böhme, U. *et al.* (2018) Complete avian malaria parasite genomes reveal features associated with lineage-specific evolution in birds and mammals. *Genome Res.* 28, 547–560
42. Bensch, S. *et al.* (2016) The genome of *Haemoproteus tartakovskyi* and its relationship to human malaria parasites. *Genome Biol. Evol.* 8, 1361–1373
43. Siau, A. *et al.* (2008) Temperature shift and host cell contact up-regulate sporozoite expression of *Plasmodium falciparum* genes involved in hepatocyte infection. *PLoS Pathog.* 4, e1000121
44. Otto, T.D. *et al.* (2010) New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq. *Mol. Microbiol.* 76, 12–24
45. Bozdech, Z. *et al.* (2003) The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol.* 1, 85–100
46. Videvall, E. *et al.* (2017) The transcriptome of the avian malaria parasite *Plasmodium ashfordi* displays host-specific gene expression. *Mol. Ecol.* 26, 2939–2958
47. Auburn, S. *et al.* (2011) An effective method to purify *Plasmodium falciparum* DNA directly from clinical blood samples for whole genome high-throughput sequencing. *PLoS One* 6, e22213
48. Zehtindjiev, P. *et al.* (2008) Dynamics of parasitemia of malaria parasites in a naturally and experimentally infected migratory songbird, the great reed warbler *Acrocephalus arundinaceus*. *Exp. Parasitol.* 119, 99–110
49. Ishtiaq, F. *et al.* (2017) Estimating prevalence of avian haemosporidians in natural populations: A comparative study on screening protocols. *Parasites Vectors* 10, 1–10
50. Aird, D. *et al.* (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* 12, R18
51. Frey, K.G. *et al.* (2014) Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. *BMC Genomics* 15, 1–14
52. Videvall, E. (2018) *Plasmodium* parasites of birds have the most AT-rich genes of eukaryotes. *Microb. Genomics* 4, 1–9
53. Kozarewa, I. *et al.* (2009) Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat. Methods* 6, 291–295
54. Oyola, S.O. *et al.* (2012) Optimizing illumina next-generation sequencing library preparation for extremely at-biased genomes. *BMC Genomics* 13, 1
55. Ross, M.G. *et al.* (2013) Characterizing and measuring bias in sequence data. *Genome Biol.* 14, R51
56. Gardner, M.J. *et al.* (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419, 498–511
57. Dávalos, L.M. and Perkins, S.L. (2008) Saturation and base composition bias explain phylogenomic conflict in *Plasmodium*. *Genomics* 91, 433–442
58. Lutz, H.L. *et al.* (2016) Laser capture microdissection microscopy and genome sequencing of the avian malaria parasite, *Plasmodium relictum*. *Parasitol. Res.* 115, 4503–4510
59. Palinauskas, V. *et al.* (2010) Laser microdissection microscopy and single cell PCR of avian hemosporidians. *J. Parasitol.* 96, 420–424
60. Palinauskas, V. *et al.* (2013) A new method for isolation of purified genomic DNA from haemosporidian parasites inhabiting nucleated red blood cells. *Exp. Parasitol.* 133, 275–280
61. Feehery, G.R. *et al.* (2013) A method for selectively enriching microbial DNA from contaminating vertebrate host DNA. *PLoS One* 8, e76096
62. Oyola, S.O. *et al.* (2014) Optimized whole-genome amplification strategy for extremely AT-biased template. *DNA Res.* 21, 661–671
63. Lauron, E.J. *et al.* (2014) Transcriptome sequencing and analysis of *Plasmodium gallinaceum* reveals polymorphisms and selection on the apical membrane antigen-1. *Malar. J.* 13, 1–14
64. Lauron, E.J. *et al.* (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, 1–10
65. Pauli, M. *et al.* (2015) *De novo* assembly of the dual transcriptomes of a polymorphic raptor species and its malarial parasite. *BMC Genomics* 16, 1038
66. Weinberg, J. *et al.* (2018) *De novo* transcriptome assembly and preliminary analyses of two avian malaria parasites, *Plasmodium delichoni* and *Plasmodium homocircumflexum*. *Genomics* Published online December 13, 2018. http://dx.doi.org/10.1016/j.ygeno.2018.12.004
67. Toscani Field, J. *et al.* (2018) Delineation of the genera *Haemoproteus* and *Plasmodium* using RNA-Seq and multi-gene phylogenetics. *J. Mol. Evol.* 86, 646–654
68. Beadell, J.S. and Fleischer, R.C. (2005) A restriction enzyme–based assay to distinguish between avian hemosporidians. *J. Parasitol.* 91, 683–685
69. Omori, S. *et al.* (2007) Complete nucleotide sequences of the mitochondrial genomes of two avian malaria protozoa, *Plasmodium gallinaceum* and *Plasmodium juxtanucleare*. *Parasitol. Res.* 100, 661–664
70. Mantilla, J.S. *et al.* (2013) Identification of *Plasmodium* (*Haemamoeba*) *lutzi* (Lucena, 1939) from *Turdus fuscater* (Great Thrush) in Colombia. *J. Parasitol.* 99, 662–668
71. Perkins, S.L. (2008) Molecular systematics of the three mitochondrial protein-coding genes of malaria parasites: Corroborative and new evidence for the origins of human malaria. *Mitochondrial DNA* 19, 471–478
72. Matta, N.E. *et al.* (2014) Description and molecular characterization of *Haemoproteus macrovacuolatus* n. sp. (Haemosporida, Haemoproteidae), a morphologically unique blood parasite of black-bellied whistling duck (*Dendrocygna autumnalis*) from South America. *Parasitol. Res.* 113, 2991–3000
73. Mantilla, J.S. *et al.* (2016) *Haemoproteus erythrogravidus* n. sp. (Haemosporida, Haemoproteidae): Description and molecular characterization of a widespread blood parasite of birds in South America. *Acta Trop.* 159, 83–94
74. Omori, S. *et al.* (2008) Two extra chromosomal genomes of *Leucocytozoon caulleryi*; Complete nucleotide sequences of the mitochondrial genome and existence of the apicoplast genome. *Parasitol. Res.* 103, 953–957
75. Lotta, I.A. *et al.* (2015) *Leucocytozoon pterotenuis* sp. nov. (Haemosporida, Leucocytozoidae): description of the morphologically unique species from the Grallariidae birds, with remarks on the distribution of *Leucocytozoon* parasites in the Neotropics. *Parasitol. Res.* 114, 1031–1044
76. Lotta, I.A. *et al.* (2016) *Leucocytozoon* diversity and possible vectors in the neotropical highlands of Colombia. *Protist* 167, 185–204
77. Imura, T. *et al.* (2014) The apicoplast genome of *Leucocytozoon caulleryi*, a pathogenic apicomplexan parasite of the chicken. *Parasitol. Res.* 113, 823–828

78. Arisue, N. *et al.* (2012) The *Plasmodium* apicoplast genome: conserved structure and close relationship of *P. ovale* to rodent malaria parasites. *Mol. Biol. Evol.* 29, 2095–2099
79. Galen, S.C. *et al.* (2018) Integrating coalescent species delimitation with analysis of host specificity reveals extensive cryptic diversity despite minimal mitochondrial divergence in the malaria parasite genus *Leucocytozoon*. *BMC Evol. Biol.* 18, 128
80. Huang, X. *et al.* (2018) The success of sequence capture in relation to phylogenetic distance from a reference genome: a case study of avian haemosporidian parasites. *Int. J. Parasitol.* 48, 947–954
81. Barrow, L.N. *et al.* (2018) Genomic sequence capture of haemosporidian parasites: Methods and prospects for enhanced study of host-parasite evolution. *Mol. Ecol. Resour.* Published online December 16, 2018. http://dx.doi.org/10.1111/1755-0998.12977
82. Graczyk, T.K. *et al.* (1994) Extraction of *Haemoproteus columbae* (Haemosporina: Haemoproteidae) antigen from rock dove pigeons (*Columba livia*) and its use in an antibody ELISA. *J. Parasitol.* 80, 713–718
83. Boissière, A. *et al.* (2012) Isolation of *Plasmodium falciparum* by flow-cytometry: implications for single-trophozoite genotyping and parasite DNA purification for whole-genome high-throughput sequencing of archival samples. *Malar. J.* 11, 163
84. Shapiro, H.M. and Mandy, F. (2007) Cytometry in malaria: moving beyond Giemsa. *Cytometry A* 71, 643–645
85. Howard, R.J. *et al.* (1979) *Plasmodium*-infected blood cells analyzed and sorted by flow fluorimetry with the deoxyribonucleic acid binding dye 33258 Hoechst. *J. Histochem. Cytochem.* 27, 803–813
86. Omori, S. *et al.* (2010) Use of flow cytometry to separate *Leucocytozoon caulleryi* gametocytes from avian blood. *Parasitology* 137, 1899–1903
87. Chakarov, N. *et al.* (2012) Label-free enrichment of avian *Leucocytozoon* using flow cytometric sorting. *Parasitology* 139, 1547–1552
88. Oyola, S.O. *et al.* (2013) Efficient depletion of host DNA contamination in malaria clinical sequencing. *J. Clin. Microbiol.* 51, 745–751
89. Oyola, S.O. *et al.* (2016) Whole genome sequencing of *Plasmodium falciparum* from dried blood spots using selective whole genome amplification. *Malar. J.* 15, 1–12
90. Leichty, A.R. and Brisson, D. (2014) Selective whole genome amplification for resequencing target microbial species from complex natural samples. *Genetics* 198, 473–481
91. Arensburger, P. *et al.* (2010) Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. *Science* 330, 86–88
92. Morales-Hojas, R. *et al.* (2018) The genome of the biting midge *Culicoides sonorensis* and gene expression analyses of vector competence for bluetongue virus. *BMC Genomics* 19, 624
93. Rosenberg, R. and Rungsiwongse, J. (1991) The number of sporozoites produced by individual malaria oocysts. *Am. J. Trop. Med. Hyg.* 45, 574–577
94. Quail, M.A. *et al.* (2012) Optimal enzymes for amplifying sequencing libraries. *Nat. Methods* 9, 10–11
95. Nair, S. *et al.* (2014) Single-cell genomics for dissection of complex malaria infections. *Genome Res.* 24, 1028–1038
96. Trevino, S.G. *et al.* (2017) High-resolution single-cell sequencing of malaria parasites. *Genome Biol. Evol.* 9, 3373–3383
97. Parks, D.H. *et al.* (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055
98. Lemieux, J.E. *et al.* (2013) Genome-wide profiling of chromosome interactions in *Plasmodium falciparum* characterizes nuclear architecture and reconfigurations associated with antigenic variation. *Mol. Microbiol.* 90, 519–537
99. Lapp, S.A. *et al.* (2018) PacBio assembly of a *Plasmodium knowlesi* genome sequence with Hi-C correction and manual annotation of the SICAvar gene family. *Parasitology* 145, 71–84
100. Borner, J. and Burmester, T. (2017) Parasite infection of public databases: A data mining approach to identify apicomplexan contaminations in animal genome and transcriptome assemblies. *BMC Genomics* 18, 1–12
101. Holmes, I. and Davis Rabosky, A.R. (2018) Natural history bycatch: a pipeline for identifying metagenomic sequences in RADseq data. *PeerJ* 6, e4662
102. Orosz, F. (2015) Two recently sequenced vertebrate genomes are contaminated with apicomplexan species of the Sarcocystidae family. *Int. J. Parasitol.* 45, 871–878
103. Pascar, J. and Chandler, C.H. (2018) A bioinformatics approach to identifying *Wolbachia* infections in arthropods. *PeerJ* 6, e5486
104. Zhang, G. (2015) Bird sequencing project takes off. *Nature* 522, 34
105. Videvall, E. *et al.* (2015) The avian transcriptome response to malaria infection. *Mol. Biol. Evol.* 32, 1255–1267