

Chapter 4

The Use of Molecular Methods in Studies of Avian Haemosporidians



Staffan Bensch and Olof Hellgren

Abstract During the last two decades, molecular methods to study mitochondrial DNA sequence variation have become an important part in the studies of avian haemosporidians. Up until recently, these methods have primarily been used for identification of the parasites and for tentative phylogenetic reconstructions, allowing researchers not trained in traditional parasitology to compare data across the globe. However, with the introduction of genome and transcriptome sequencing, studies are emerging that go deeper into the genetics and molecular biology of the parasites. In this chapter, we describe and summarize the common methods used for genetic barcoding of the parasites and give an introduction of what to take into account when designing a molecular study of avian haemosporidians. This chapter further discusses why nuclear genetic data are needed in order to answer several important ecological and evolutionary questions and which methods to use in order to overcome the obstacles of obtaining nuclear data of the parasites. Finally, this chapter highlights the challenges and opportunities that come with the use of molecular methods, such as how to study and interpret prevalence, the challenge of aborted developments, and how to obtain data for more robust phylogenies and population structure studies of the parasites.

Keywords Barcoding · Genomics · Haemosporida · Molecular parasitology · Molecular phylogenies · Transcriptomics

4.1 Introduction

The prevailing molecular methods for identification of haemosporidian parasites of birds are based on DNA sequence variation within the mitochondrial genome. One region in particular, a 479 bp fragment of the cytochrome b gene, has been the

S. Bensch (✉) · O. Hellgren
Department of Biology, Ecology Building, Lund University, Lund, Sweden
e-mail: staffan.bensch@biol.lu.se

target in several hundred publications and, thus by sheer number, has become the “barcoding” region for avian haemosporidians (Bensch et al. 2009). Unique haplotypes of this region are called “lineages”. Published records of haemosporidian lineages are stored and available for comparison via BLAST (Altschul et al. 1990) or for download in the database “MalAvi”, along with associated information such as host species and geographic locality (Fig. 4.1). The MalAvi database presently (June 2020) contains information on 1577 lineages of *Haemoproteus*, 1287 of *Plasmodium*, and 1138 of *Leucocytozoon*. This barcoding region of the *cytb* gene, for simplicity called the MalAvi region, can be amplified by a number of primer pairs (Table 4.1) with the only requirement for identification of lineages that the sequenced amplicons cover most (~90%) of this domain. The coordinated efforts of barcoding have made a huge impact in the field of avian haemosporidian research (Rivero and Gandon 2018). However, to correctly interpret results from *cytb* studies, that is, whether the lineages represent good species or intraspecies variation, there is a need for more genetic information, especially from genes not linked to the mitochondrial genome.



Fig. 4.1 The MalAvi database. Unique *cytb* haplotypes are called “lineages” and are named, typically by a 5–6-letter acronym corresponding to the scientific name of the first recorded host species combined with a two-digit number. For example, the lineage ACAED02 was the second lineage to be found in *Acrocephalus aedon*. To check whether a lineage is already present in the MalAvi database, one can use the BLAST function on the website, or download all of the lineages as a FASTA file for examination locally on the computer (click FASTA => All sequences). All the report tables can be checked online or downloaded in Excel format. For example, previous records of a lineage (host species or locations) can be found in the “Hosts And Sites Table.” The database is curated by Staffan Bensch and updated online every 2–3 months. To submit data to the MalAvi database, one can use the Excel form provided at the main web page (<http://mbio-serv2.mbioekol.lu.se/Malavi/>) and email this (along with questions) to staffan.bensch@biol.lu.se

Table 4.1 A selective list of primers that amplify haemosporidian mitochondrial DNA from the genera *Haemoproteus* (H), *Plasmodium* (P), and *Leucocytozoon* (L)

Primer combinations ^a	mtDNA Region	Main use	Length ^b (bp)	H	P	L	MalAvi Overlap (%)	Reference
(1) HAEMF × HAEMR2	<i>cytb</i>	Identification	479 ^b	-	-		100	Bensch et al. (2000)
(1) HAEMNF × HAEMNR2 (2) HAEMF × HAEMR2	<i>cytb</i>	Identification / screening	479 ^b	-	-		100	Waldenström et al. (2004)
(1) HAEMNFI × HAEMNR3 (2) HAEMF × HAEMR2	<i>cytb</i>	Identification / screening	479 ^b	-	-		100	Hellgren et al. (2004)
(1) HAEMNFI × HAEMNR3 (2) HAEMFL × HAEMR2L	<i>cytb</i>	Identification / screening	479 ^b			- ^c	100	Hellgren et al. (2004)
(1) DW2 × DW4 (2) DW1 × DW6	<i>cytb</i>	Phylogeny	1138 ^b	-	-	-	100	Perkins and Schall (2002)
(1) 621F × 983R	<i>cytb</i>	Identification / screening	229 ^b	-	-		15	Fallon et al. (2003a)
(1) 3760F × 4292Rw2	<i>cytb</i>	Identification / screening	533 ^b	-	-		100	Beadell et al. (2004)
(1) DW2 × DW4 (2) LeucoF × LeucoR	<i>cytb</i>	Identification / screening	818 ^b			-	100	Perkins and Schall (2002), Sehgal et al. (2006)
(1) HML × HMR	<i>cytb</i>	Screening	367	-			68	Martinez et al. (2009)
(1) Plas-F × 4292Rw	<i>cytb</i>	Screening	422		-		70	Martinez et al. (2009)
(1) Palu-F × Palu-R	<i>cytb</i>	Screening	391 ^d	-	-		64	Martinez et al. (2009)
(1) CytF1 × CytR1 (2) CytFN × CytRN	<i>cytb</i>	Phylogeny	1181 ^b	-	-	-	100	Schmid et al. (2017)
(1) F × R (2) Finternal × Rinternal	Whole mtDNA	Phylogeny	5451	-	-	-	100	Pacheco et al. (2018b)
(1) 213F × 372R	n-c	Screening	160 ^e	-	-		0	Beadell and Fleischer (2005)
(1) 343F × 496R	n-c	Screening ^f	188	-	-	-	0	Fallon et al. (2003b)
(1) AE064 × AE066			1109	-	-	-	100	Pacheco et al. (2018a)
(2) AE980 × AE982	<i>cytb</i>	Screening	346	-			0	
(2) AE983 × AE985			580		-		47	

(continued)

Table 4.1 (continued)

Primer combinations ^a	mtDNA Region	Main use	Length ^b (bp)	H	P	L	MalAvi Overlap (%)	Reference
(1) PMF × PMR	n-c		377	-			0	
(1) HMF × HMR	n-c	Screening	525		-		0	
(1) LMF × LMR	co1		212			- ^c	0	

The main use of the protocols is indicated as screening (presence/absence on gels), lineage identification, or phylogenetics (whole *cytb* or whole mtDNA genome). The sequences of the primers can be found in the original publications

^a(1) Denotes primers in first reaction, (2) the primers in a second nested reaction

^bExcluding the length of primers (i.e., the length of novel sequences obtained)

^cThe primers appear to fail in amplifying many lineages of *Leucocytosoon* having fusiform gametocytes (Lotta et al. 2019)

^dTreatment of the amplicon with the endonuclease *Hpy* CH4III cuts *Haemoproteus* sequences in two fragments (363 bp and 27 bp) and *Plasmodium* sequences in three fragments (327 bp, 36 bp, and 27 bp)

^eTreatment of the amplicon with the endonuclease *Xba*I cuts *Haemoproteus* sequences (121 bp and 39 bp), whereas *Plasmodium* sequences remain intact (160 bp)

^fWorks also for estimating total parasitemia by qPCR (Ishtiaq et al. 2017)

4.2 Barcoding, Species Limits, and Population Structures

A genetic region suitable for barcoding must first of all exist in all of the species attempted to be barcoded; second, it should show sufficient variation so that species can be distinguished. At the beginning of this century when the first MalAvi sequences were generated (Bensch et al. 2000), there was so little knowledge in the field that these assumptions could not be critically tested. A few parasites have failed amplification even when tested by multiple primers (Valkiūnas et al. 2016), in one case possibly indicating that the mitochondrial genome had been deleted (Zehindjiev et al. 2012) like in *Cryptosporidium*, a distant relative within Apicomplexa parasites (Kuo et al. 2008). With a few possible exceptions, it appears that the MalAvi region is present in the genomes of all avian haemosporidians investigated to date. Encouraging for species discrimination, it was observed that the sequence diversity seemed to be many fold higher than estimates of species diversity based on morphological analyses alone (i.e., the observed genetic variation showed potential for investigating the presence of cryptic species). It also seemed that the MalAvi region captured most of the lineage diversity when comparing the same samples sequenced for the full mtDNA *cytb* coding region; Hellgren et al. (2007a) tested 8 lineages (2–13 samples / lineage) without finding additional lineages with the full *cytb* coding region data set. Not surprisingly, later studies have found cases of the same MalAvi lineage breaking up into more lineages when sequencing the full *cytb* gene (Musa et al. 2018) or the whole mtDNA genome (Pacheco et al. 2018a, b; Huang et al. 2018). However, these cases are comparably few. Hence, for barcoding avian haemosporidians and trying to minimize sequence

length while maximizing lineage discrimination, it seems that the MalAvi region meets the purpose quite well. That more diversity can be uncovered with longer mtDNA sequences is not in itself a problem for barcoding. Note, for example, that adding sequences from nuclear genes up to the complete nuclear genome will always increase the resolution. The purpose of barcoding is to have one sufficiently informative and easily studied region that a research community agrees to sequence across studies, enabling unambiguous identification of parasites for direct comparisons of parasite diversity among host species and geographic regions.

The mitochondrial genome of haemosporidian parasites is minimal in size (6 kb) with a linear instead of a circular organization, which is otherwise the common arrangement in eukaryotes (Hikosaka et al. 2013). In mammal *Plasmodium* parasites, each mitochondrion contains multiple copies of the 6 kb genome, typically arranged in tandem and repeated dozens to up to 150 times (Wilson and Williamson 1997). Direct studies of avian haemosporidian mitochondrial DNA structure have not been carried out, but it is plausible that the situation is similar. An important consequence of this is that DNA samples will always contain many more copies of parasite mtDNA (10–100×) than any nuclear gene of the parasite. Hence, protocols targeting mtDNA fragments will be correspondingly more sensitive than protocols amplifying nuclear genes. Attempts have been done to use primers for 18S rRNA for avian malaria infections screening (Feldman et al. 1995), a commonly used marker for identification of a wide range of organisms, particularly protozoans. This protocol has not been widely used, as it seems to be specific for *Plasmodium relictum*. In contrast to animals that often have hundreds of tandem repeated rRNA copies in the nuclear genome, in haemosporidians, they are single copy genes arranged on different chromosomes and are often highly divergent from each other (Gunderson et al. 1987; Rooney 2004). Hence, it seems unlikely that we will find sensitive and universal primers for amplifying rRNA of haemosporidians.

Within the mitochondrial genome, any, or at least many other regions would probably work equally well for barcoding. One main reason for keeping focus on the MalAvi region is the wealth of information already available for comparisons. It is important to note that screening for prevalence using primers that generate long amplicons will more likely result in false negatives. This is because such reactions are more sensitive to variation in template quality, quantity, and PCR conditions. Hence, a more efficient strategy is to first use protocols for shorter amplicons, such as the primers for the MalAvi region, to identify the infected samples. The barcoding sequence of these samples can then form the basis when selecting samples for generating more complete mtDNA sequences, preferably to be combined with sequences from multiple nuclear genes (Borner et al. 2016) for robust phylogenetic reconstructions.

A lingering question has been whether similar mitochondrial lineages (i.e., sequences that differ by 1–5 bp) represent variation within species or indeed unique species. To date, there is no consensus regarding the rate of molecular change for the mitochondrial genes in the haemosporidian parasites. There is however good evidence that parasite mitochondrial DNA evolves at a slower rate than the mitochondria of hosts, and the parasite's nuclear genes evolve 6–10 times faster than

their mitochondrial DNA (Nilsson et al. 2016). Rate estimates for the *cytb* gene range between 0.1% and 1.3% divergence per million year (Ricklefs and Outlaw 2010; Bensch et al. 2013; Pacheco et al. 2018b). In practice, this would translate to one mutation within the MalAvi fragment (equal to 0.2% sequence divergence) every 154,000 to 2 million years. This has two important implications. First, cases where lineages only differ by a single mutation will include both very newly arisen mutations where the lineages are part of the same recombining population and cases where these have diverged for up to four million years (assuming a divergence rate of 0.1% and that the lineages are analyzed just before the second mutation has occurred). Second, when parasites are sharing the same *cytb* lineage, these might consist of populations that have been isolated for up to 154,000–2 million years. With a higher rate of change in nuclear genes, such populations can be substantially divergent in their nuclear genomes.

It was early observed that lineages that differ by only 1–2 bp sometimes had drastically different host species distributions (e.g., HIICT1 and HIPOL1) (Reullier et al. 2006) or gametocyte morphology (*Haemoproteus minutus* and *Haemoproteus pallidus*; Hellgren et al. 2007a). Moreover, the few studies that have investigated nuclear genes of closely related *cytb* lineages have found that these frequently are associated with distinctly different nuclear alleles: in *Haemoproteus* (Nilsson et al. 2016), *Plasmodium* (Beadell et al. 2006), and *Leucocytozoon* (Galen et al. 2018b). This suggests an absence of successful reproduction between these closely related lineages. Hence, such strict associations between mitochondrial and nuclear variants support that they should represent independently evolving nonrecombining units (e.g., Bensch et al. 2004), a hallmark for good biological species. With that said, it is important to point out that similar lineages also can represent within species variation; for example, the *Plasmodium relictum* lineages SGS1 and GRW11 share MSP1 alleles (Hellgren et al. 2013b), suggesting that they are recombining. Unfortunately, there does not seem to be a general cut-off, or so-called barcoding gap, which can separate within-species variation from between-species variation.

To what extent similar lineages represent one or several species needs therefore to be addressed case by case. Compared to animals, it is surprising that mtDNA distances as low as 0.2% often appears to represent unique species in haemosporidian parasites. In birds, within species divergence is generally <2%, whereas the between-species divergence is >4% (Ward 2009). The lack of a clear barcoding gap for haemosporidians emphasizes the importance of recording the findings of all unique lineages in the primary publications (i.e., not lumping lineages that do differ into the same entity), as their species status may be recognized in the future.

4.3 The Design of Molecular Studies

Before the introduction of molecular techniques, most of the species or community surveys of haemosporidians reported data on the overall prevalence of the three genera *Haemoproteus*, *Plasmodium*, and *Leucocytozoon*, without further attempts

to separate the species within these genera (Pierce 1981; see Chap. 1 for a synthesis of avian haemosporidian research in tropical regions during the twentieth century). This is understandable given that microscopic identification requires high infection levels and high-quality blood smears, which is rarely the case in data sets of hundreds of samples from wild captured birds. Molecular methods now make it possible also for the nonspecialist to identify haemosporidian parasites to a level that enables unambiguous identification of the parasites, not only to genus but also down to lineages (see Chap. 2 for avian haemosporidian life cycles and study methods). However, even if molecular methods have a very strong identification power, they can be implemented very differently and also come with a number of caveats that will be discussed below (see also Chap. 2).

4.3.1 Selecting Buffer for Storing Blood Samples

There are many buffers equally good for obtaining high-quality DNA from avian blood. These include inexpensive alternatives such as ethanol and more expensive commercial buffers. It has been reported that buffers containing sodium dodecyl sulfate (SDS) may reduce the success of amplifying haemosporidian DNA (Freed and Cann 2006) and in general, one should use buffers and sampling procedures that minimize polymerase inhibitors, such as heparin, EDTA, and Giemsa staining (Palinauskas et al. 2010; Owen 2011). Because each buffer will require its specific extraction technique, the choice of buffer for storing the blood should be decided with the plan of how the DNA will be extracted. A good starting point is to find out what buffers and extraction techniques are normally used in the laboratory where the work is to be carried out. When working with a technique that is already well established in a laboratory, troubleshooting will be easier as there will be experienced colleagues who can give advice on how to deal with low-yield or low-purity DNA.

4.3.2 Selecting Primers

There are many published PCR primers that work equally well if correctly optimized (Table 4.1). Importantly, the optimal condition for a specific set of primers may differ between laboratories due to different brands of DNA polymerase, PCR instruments, and other reagents. Just because one set of primers works better than another pair in a particular laboratory, it does not predict global success. Most of the general PCR and sequencing protocols have been developed based on parasites in passerine birds. Hence, there are reasons to believe that lineages from nonpasserines may be undetected by these primers. High lineage richness can result in failure to amplify some parasites, as some of the ~10,000 lineages just by chance may have mutations at crucial primer-binding sites. The task for any project is to develop an

optimized assay for the study: an assay that generates repeatable results and that does not miss infections identified by alternative methods (i.e., microscopy of blood smears). However, many data sets consist of just blood samples (no blood smear for verification). Starting screening such data without verifying that the selected protocol is working well in the laboratory is risky. A recommended strategy is to select about 30–50 samples for testing with two different screening protocols that produce absence/presence data (Fallon et al. 2003b; Beadell and Fleischer 2005; Ishtiaq et al. 2017; Pacheco et al. 2018a; Ciloglu et al. 2019), in parallel with developing the protocol for generating sequence data.

4.3.3 Coinfections

Wild birds are frequently infected with two or more different parasites (Valkiūnas et al. 2006). If both parasites have similar infection intensities and the primers bind equally well to the DNAs of these parasites, they will be amplified together and sequencing electropherograms will show double base-calling at the positions where the parasites have different nucleotides (Fig. 4.2). If the common parasites of the host species at the specific study site are known, one might be able to resolve the coinfections, by finding two lineages that can explain all the positions of the double base-calling. If the mix is from more than two parasites, this approach does not work. The standard way to resolve coinfections in unknown or complex situations is by cloning the PCR product and sequencing multiple clones (Perez-Tris and Bensch 2005). Note that sequences of cloned PCR products contain polymerase errors that normally go undetected when employing direct sequencing of PCR products, as the sequence is the average of millions of sequence reads. To verify the correct sequence from sequenced clones, multiple clones need to be analyzed.

In many instances, the PCR will favor only one of the parasites, either the one with the highest parasitemia or the one with the best match to the primers,

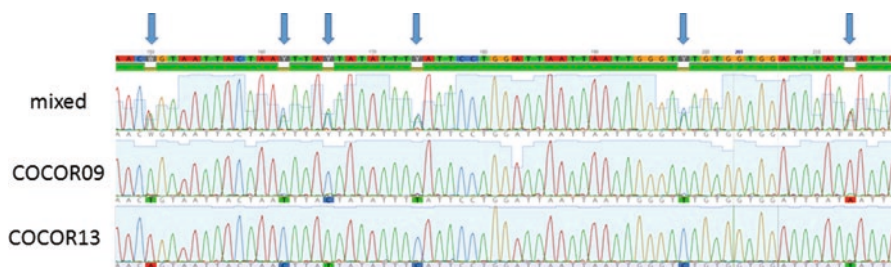


Fig. 4.2 At the top is a sequence of coinfections (six positions of double base-calling: blue arrows) that can be explained by a combination of the *Leucocytozoon* lineages COCOR09 and COCOR13. The samples are from Jackdaws (*Corvus monedula*)

erroneously showing the result of single infections (Bernotienė et al. 2016). The presence of infections that include parasites of different genera can be efficiently identified by examination of blood smears or by protocols with multiplex primers (Ciloglu et al. 2019) or by combining PCR with restriction enzymes (Beadell and Fleischer 2005; Martinez et al. 2009). However, revealing the hidden sequence by sequencing may require design of primers that exclude the masking sequence. Coinfections of lineages from the same genus that go undetected by standard PCR and sequencing protocols will generally be missed unless blood smears show presence of morphologically distinct species or that lineage-specific primers are used.

The consequences of failing to detect and determine the sequences in coinfections depend on the specific questions asked. If the aim is to describe the lineage diversity of a host species, missing some coinfections is probably not so important as all lineages of the species at some point will occur in single infections (assuming the infections are fairly independent). In bird species with generally high prevalence (> 70%), rare lineages will often be in coinfections and thus easy to miss. Hence, the higher the prevalence, the more samples need to be screened in order to find most of the lineages infecting a species. The failure to identify coinfections should generally be more problematic in studies aimed at precisely estimating lineage prevalence, particularly in bird species with high lineage diversity and high prevalence where the lineages will mask each other occurrence. In such cases, lineage-specific qPCR may be a solution (Asghar et al. 2011). However, each lineage needs its own PCR design and optimization, which can be very laborious and comes with fairly high running costs. It is important to remember that even if all efforts are taken, the observed frequency of coinfections will most likely be an underestimate of the true frequency (Valkiūnas et al. 2006).

4.3.4 Contamination

A successful PCR may generate many millions of copies of the targeted DNA fragment, each copy a potential source of contamination in future experiments. This is the reason why it is important to add negative template controls (NTC) to check for the presence of contaminants in the preparation of reagents. However, adding one or a few NTC on a 96 plate may not be sufficient to identify low-level background contamination. If low levels of contaminating PCR products are present in the laboratory, it may show up in only some of the NTCs. For example, in our laboratory in Lund, we ran a project screening 382 samples from birds from the Azores that had overall low rates of infections (Hellgren et al. 2011). We used 1 NTC per row of 8 samples, that is, 12 NTC per plate. All of these were negative as were most of the wild samples from 10 species of birds with the exception of blackbirds (prevalence 57%). Sequencing showed that one European robin (*Erithacus rubecula*), one blackcap (*Sylvia atricapilla*), one Eurasian blackbird (*Turdus merula*), and two house sparrows (*Passer domesticus*) were each infected by the lineage

PARUS1, a common parasite of tits in Europe, and not previously recorded in these species. We examined blood smears from these specimens but could not find gametocytes and analyses of subsequent DNA extractions failed to amplify. So, how did we find the lineage PARUS1 in these samples? The most likely reason was due to spill-over from a previous project in the lab, completed just a few months earlier that included amplifications from 53 blue tits (*Cyanistes caeruleus*) infected with PARUS1 (Stjernman et al. 2008). In this case, we could detect the contamination because the records stood out as surprising (wrong host species) and we had blood smears to check with microscopy, but in many other situations, such contamination may go undetected (including in our lab in Lund). This calls for a general awareness of contamination, not only between samples within a study but also from previous projects carried out in the particular laboratory. To confirm that a surprising finding indeed comes from a parasite infecting the bird and not from a PCR-contaminated product, one can check the sample with primers targeting other parts of the mitochondrial genome and also under the microscope to search for gametocytes.

4.3.5 *Mixing Up the Samples*

This sounds like a problem so obvious that it is not even worth bringing up. However, when working with hundreds of samples, it can happen easily, and if it goes undetected, it will lead to erroneous host records of parasites. Samples can be mislabeled in the field, during DNA extraction, or when preparing dilutions for the PCR. In studies of single species, this is not fatal beyond the study itself, although mixing up the tubes from two individuals, one of which is infected and one is not, will give incorrect parasite status to two individuals in the data set. This will reduce the power of the statistical analyses aimed at finding phenotypic predictors (e.g., sex, age) of parasite infection status. The consequences of mixing up samples will however be more severe in studies including several species. For example, if the mixed-up samples belong to two different species, each infected with a host-specific lineage of parasites, it would lead to two erroneous host records in the databases that will inflate future calculations of host specificity. Mixed-up samples of host species can easily be detected by amplifying and sequencing a smaller fragment of the bird's mtDNA (e.g., *coxI* or *cytb*) using universal primers (Kocher et al. 1989), but due to the extra labor and costs for sequencing, this is rarely done on complete data sets. However, checking the host species identity of samples that have resulted in “surprising findings” is a good practice to reduce errors of parasite host records. For example, a surprising finding would be a typical host-specific parasite in very different host species. Particularly vulnerable to this kind of problem are studies including relatively few samples from many previously not investigated species (i.e., the common situation for community studies in the tropics).

4.4 How Can We Obtain Data from Nuclear Genes?

Many evolutionary and ecological questions require more genetic information than can be retrieved from mtDNA alone. However, to obtain data on nuclear genes has been far from easy for several reasons. The main reason is the fact that the avian hosts have nucleated erythrocytes and that there is a big difference in genome size between the host and the parasite, as well as a difference in ploidy during the vertebrate stage. Even if all the red blood cells were infected, such an infected sample would mainly contain bird DNA (< 1% would be from the parasite). In nature, the level of parasitemia is typically <1%, resulting in a parasite proportion of <0.01% of the total DNA. Although sequencing costs are decreasing and sequencing platforms steadily become more efficient, it would require a tremendous sequencing effort to obtain a good coverage parasite genome from a naturally infected bird. In fact, there would be a need to sequence around 10 trillion base-pairs in order to obtain a 20× genome of the parasite (Videvall 2019). This bias is reflected in the low yield of parasite reads found when looking for avian haemosporidian sequences in databases of genome-sequenced hosts (Borner and Burmester 2017). Due to these problems, until recently, the only way to amplify nuclear genes was by using primers designed based on genomes from either mammalian malaria parasites and an unpublished fragmented genome of *Plasmodium gallinaceum*. Despite much effort from many laboratories, the success was restricted because designing working primers using alignments of these very divergent genomes was problematic. Between 2000 and 2013, publications included one apicoplast gene (*clpc*) and a handful of nuclear genes (*DHRF-TS*, *TRAP*, *ASL*, *MAELB*), of which many only could be amplified from a limited number of lineages (Jarvi et al. 2003; Bensch et al. 2004; Martinsen et al. 2008; Santiago-Alarcon et al. 2010; Farias et al. 2012; Martinez et al. 2013). Simultaneously, a lot of unsuccessful effort was put into sequencing the first genome of avian haemosporidians. Since then, several different methods have been developed in order to circumvent the problem of skewed DNA ratios when obtaining nuclear data of the parasites. At present, there are three major strategies of obtaining large number of sequences from nuclear genes: 1) RNA/transcriptome sequencing, 2) Parasite enrichment followed by genome sequencing, and 3) Multigene sequencing using sequenced genomes as the backbone.

4.4.1 RNA/Transcriptome Sequencing

When the parasite has infected a host, the ongoing asexual replication will require the molecular machinery responsible for the basic metabolic processes along with those for the biological processes during host invasion. This means that DNA transcription in the parasites will be fully switched on to produce mRNA, followed by translation into the required proteins. The reason for using RNA sequencing instead of DNA sequencing to obtain nuclear data rests on the notion that the parasite is producing abundant mRNA in the peripheral blood, resulting in a more favorable ratio of parasite-host RNA for sequencing. The first avian haemosporidian RNA

data was from a Crossbill (*Loxia* sp.) experimentally infected with the *P. relictum* lineage SGS1 (Bensch et al. 2014). This first transcriptome was of low coverage and included relatively few and incompletely sequenced genes. Nevertheless, this partial transcriptome was sufficient for developing primers for investigating several nuclear genes in the parasite *P. relictum* (Hellgren et al. 2013a, 2015). As RNA isolation and sequencing have been improved, there are now several well-annotated transcriptomes available, primarily from different *Plasmodium* species (Lauron et al. 2014; Videvall et al. 2017; Weinberg et al. 2018), but also from the genus *Leucocytozoon* (Pauli et al. 2015). Strong advantages of using RNA sequencing are that there is no need to enrich for the parasite in the samples, it does not require any prior information of the genome when constructing the protocol, and offers the possibility of investigating the expression levels of different genes of the parasite. To date, transcriptomes have primarily been obtained from birds that have been experimentally infected as this allows for collecting blood samples when the parasitemia is at its peak (see Fig. 2.2 from Chap. 2). It therefore remains to be tested whether RNA sequencing will work on samples from wild caught birds that typically have low parasitemia. Samples for RNA sequencing must however either be stored directly at -80°C (freezer or liquid nitrogen) or in suitable RNA buffers that prevent the RNA from degrading.

When obtaining the RNA sequences from an infected sample, the first challenge is to filter out the reads originating from the host. This can be done either by (1) deleting reads that map to a reference genome of the host, (2) filtering reads or contigs based on the composition of G and C where the GC% is much lower in the parasite compared to the host, (3) selecting reads or contigs that map to a related parasite genome or through homology searches using BLAST algorithms or, (4) a combination of these approaches. To date, sequencing transcriptomes seems to be one of the easiest ways to obtain nuclear data of the parasites, as it requires no manipulation of the parasite or prior knowledge of the parasites genome.

4.4.2 Parasite Enrichment Followed by Genome Sequencing

In theory, there are many ways to enrich samples of blood or tissue for parasites, to increase the yield of parasite DNA when aiming for whole-genome sequencing. The first *Haemoproteus* (*H. tartakovskiyi*) genome was sequenced from samples of harvested microgametes (Bensch et al. 2016). For this to work, it requires access to live birds having infections that contain a high number of gametocytes. When the blood is withdrawn, the gametocytes respond as if they were in the gut of a vector (i.e., exflagellation is induced), probably triggered by the direct exposure to oxygen and the drop in temperature (see Chap. 2). Due to the size difference between microgametes and host cells, the microgametes can be enriched by simple centrifugation of the samples (Palinauskas et al. 2013). Some host DNA might still be left in the sample due to ruptured red blood cells; however, the proportion of parasite to host DNA can be sufficiently enriched for conducting whole-genome extraction and sequencing. Though this method works for *Haemoproteus* species, it seems more

difficult to induce exflagellation for species of *Plasmodium* and *Leucocytozoon* (Arai et al. 2001; Valkiūnas 2005; Valkiūnas et al. 2013, 2015). Another approach is to use laser capture microdissection microscopy (LCMM) on blood smears to directly separate parasites cells from host cells before sequencing (Lutz et al. 2016). This is a promising method for isolating parasite DNA; however, it is time consuming to collect a sufficient number of parasites needed for sequencing, and requires that the blood is prepared on specific membrane slides. A further challenge is that the microscopy must be done on unstained blood smears as Giemsa staining is a potent PCR inhibitor (Palinauskas et al. 2010).

Another way to harvest parasites was used by Böhme et al. (2018) before sequencing the genome of *P. relictum* (Böhme et al. 2018). In captivity, mosquitoes were fed blood from birds infected with *P. relictum*. After 7 days, oocyst-infested midguts were dissected out from the mosquitoes, DNA was extracted and used as a template for genome sequencing. This method is specifically suitable for *Plasmodium* species as they often produce a large number of oocysts, where each may contain thousands of developing sporozoites (see Chaps. 2 and 6). For enrichment of *Leucocytozoon* spp., one can take advantage of the large difference in shape of infected blood cells compared to uninfected cells. This morphological difference was used to separate infected from uninfected cells by flow-cytometry (Chakarov et al. 2012). As this method requires a size or shape difference of infected and uninfected cells, it is less likely to work for *Plasmodium* and *Haemoproteus*; however, it might be worth testing. In the future, there will most likely be other ways to manipulate or to take advantage of the biology of the parasites life cycle in order to separate the parasites from the host cells.

The following two enrichment methods have not yet been applied successfully for avian haemosporidians but might be worth considering. It has been suggested that host and parasite DNAs can be separated based on the observation that CpG methylated sites are much more abundant in the nuclear genomes of the vertebrate hosts than in the genomes of the host mitochondria and the parasites. Feehery et al. (2013) developed this method for *Plasmodium falciparum* and observed a tenfold increase in the number of parasite reads (Feehery et al. 2013). An alternative approach is called selective whole-genome amplification (SWGAs). By selecting short primers and amplifying sequence motifs that are common in the genome of the parasite but rare in the host, the sequence coverage of the parasite genome can be enriched a tenfold (Leichty and Brisson 2014). However, for natural infection levels of avian haemosporidians that are typically <1%, a tenfold enrichment is probably not sufficient to reduce the excessive amount of host DNA.

4.4.3 Nuclear Gene Sequencing Using Available Genomes as Backbone

In the beginning of the molecular era of avian malaria research, much effort was put into using sequenced human or rodent malaria genomes as backbones when designing primers for avian malaria parasites. In some cases, these efforts have been

successful (Bensch et al. 2004; Borner et al. 2016), especially for genes that are evolutionarily conserved. However, many of the ecological and evolutionary questions that we want to answer involve gaining knowledge of genes that, due to their function (i.e., genes involved in host invasion or immune system evasion), evolve at a faster rate, or because such fast-evolving genes can provide us with a higher resolution about population structures or species boundaries. In those cases, primers or protocols need to be developed using genomes/transcriptomes from closely related species or using the combined information from multiple genomes of more distant parasites.

Several studies have used traditional Sanger sequencing when developing sequencing protocols either for single- (Bensch et al. 2004; Hellgren et al. 2013a; Garcia-Longoria et al. 2014; Nilsson et al. 2016) or multiple-gene approaches (Borner et al. 2016; Galen et al. 2018a). For studies requiring data from multiple genes, PCR protocols need optimization for each single gene under investigation, which are, of course, time, money, and template consuming. An alternative method for sequencing multiple genes is by sequence capture, which recently has been tested in haemosporidians with satisfying results (Huang et al. 2018; Barrow et al. 2019). When developing a sequence capture protocol, a researcher utilizes one or multiple genomes as a template for designing probes for the targeted genes. The protocols developed by Barrow et al. (2019) and Huang et al. (2018) targeted 498 and 1000 preselected genes, respectively. In brief, the method exploits the fact that target sequences are bound to biotin-labeled probes that can be captured by streptavidin-coated magnetic beads. Following a step of amplification with sample-specific index primers, many samples can be sequenced in parallel on a next-generation sequencing platform. Since samples can be indexed before capture (Barrow et al. 2019), the cost can be reduced substantially. One advantage with this method is that multiple probes, designed from different species, can be used in parallel, which should increase the success of recovering sequences from species and lineages that are evolutionarily distant from the available reference genomes. In the studies by Barrow et al. (2019) and Huang et al. (2018), approximately half of targeted loci were retrieved from lineages with an mtDNA difference of 5% from the reference genomes. It should however be noted that the success rate of loci recovered decreases drastically when the parasitemia drops below 0.1% (Barrow et al. 2019).

The number of available genomes and transcriptomes will progressively make it easier to develop specific protocols that include a larger phylogenetic range of the parasites. This will in turn enable larger phylogenetic comparative studies (see Chap. 3 for an introduction to systematic and phylogenetic concepts) of parasites with different life-history traits, varying, for example, in host specificity or virulence. This will vastly increase our knowledge of the epidemiology of the parasites, and what might be limiting transmission success and host range. However, in order to take the next step in this line of research, there is a need to expand the genomic research, which to date has primarily been done for parasites transmitted in temperate regions, to also include more data on avian parasites that are transmitted in the tropics.

4.5 Challenges and Opportunities with Molecular Methods

4.5.1 *Molecular Methods to Estimate Prevalence*

We now know that the three genera *Haemoproteus*, *Plasmodium*, and *Leucocytozoon* are very species rich. To set up a molecular study with the goal of just determining the prevalence of genera seems therefore a waste of resources, given the relative ease by which lineage identification can be obtained. No ornithologist would be satisfied with a study that describes bird communities by the proportion of individuals belonging to different orders, Passeriformes, Galliformes, etc. Although the age of the genera *Haemoproteus*, *Plasmodium*, and *Leucocytozoon* is a contentious issue (Bensch et al. 2013), it seems likely that they diverged before the radiation of extant birds (Pacheco et al. 2018b) and thus represent very divergent organisms that deserve as much attention as the species they infect.

The ideal goal of any screening project is to identify all infections in the collected samples (i.e., to find the true parasite prevalence and diversity). In practice, this is impossible because low-level infections will always be missed. Consider a small songbird of 10 grams that has a total blood volume of ~1.0 ml, or about 5 billion erythrocytes (assuming five million erythrocytes/ μ l). For PCR, we typically add 25 ng of total DNA as a template, which corresponds roughly to 20,000 bird genomes. In order to confirm an infected sample to be positive by PCR, it then follows that it must have at least 250,000 infected erythrocytes! Admittedly, these calculations are rough and since the parasites have several mtDNA genomes per cell, the sensitivity of the PCR might be a lot higher. Dilution experiments have suggested that the detection limit is about 1 parasite / 100,000 erythrocytes for the most commonly used nested PCR protocol (Waldenström et al. 2004). The take-home message is that the bird must be infected by tens of thousands of parasites before we can detect infection by PCR. From these calculations, one would be tempted to add more template DNA to the PCR (e.g., 100 ng rather than 25 ng that theoretically would increase the sensitivity fourfold). We are not aware of any test of this possibility, however, because too much DNA is in itself a PCR inhibitor, it may in the end not increase the sensitivity.

A common experience when retesting samples is that, for some samples, the infection status is inconsistent (e.g., out of a total of five PCRs, only three are positive). Based on the calculations above, this is precisely what we are to expect if samples are close to the detection limit. When taking 25 ng of template repeatedly from the extract, sometimes it contains enough parasites for the PCR to amplify them, but sometimes there are no parasites in the template volume.

Whatever efforts we take, our estimates of prevalence will always be an underestimate of the true prevalence. However, the knowledge of the distribution of natural parasitemia can inform us on the reliability of our estimates. Each protocol has a detection curve relative to the infection intensity of a sample, with highly infected samples being less likely missed than low infection samples (Fig. 4.3). If the distribution of the parasitemia is high for the given sample set, there will be few false

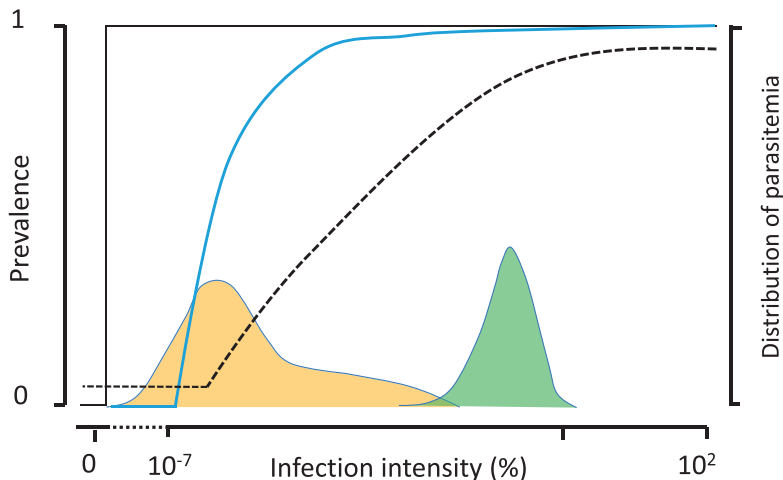


Fig. 4.3 A schematic illustration of the relationship between the successes of identifying true positive infections relative to infection intensity (i.e., parasitemia). The blue line shows the detection curve for a well-optimized protocol, the stippled line for a protocol that misses many infections already at median infection intensities and also results in false positives. The natural distributions of infection intensities of two parasites (orange and green) are illustrated at the bottom of the graph. The optimized protocol will correctly recover the infections of the green parasite, but for the orange parasite, it will result in many false negatives

negatives. However, if the distribution of the parasitemia overlaps with the decline of the detection curve, we can be sure that we have missed many infected samples. Thus, variation in parasitemia may lead to wrong conclusions of variation in prevalence. It is therefore important to consider the temporal aspect of parasitemia when estimating prevalence. Consider a parasite that has a seasonal variation in parasitemia, for example, higher in the wet season than in the dry season. By PCR, we will miss more infections in the dry season samples than in the wet season samples and conclude that prevalence is higher in the wet than in dry season, when in fact it is the parasitemia, but not the prevalence, that varies between the seasons. For any study that aims for carefully estimating prevalence, it is important to collect and examine blood smears or if these are not available, use qPCR to estimate parasitemia.

4.5.2 *Aborted Development*

To be a competent host, the parasite must be able to reach the blood and develop into mature gametocytes that are able to be transmitted to the vector and subsequently to new vertebrate host individuals. Similarly, to being a competent vector, the parasite must be able to develop into sporozoites that reach the insect's salivary glands (see

Chaps. 2 and 6). However, molecular screening of haemosporidian infections is based on determining the presence or absence of parasite DNA, without differentiating where this DNA came from. Molecular methods, particularly those that involve amplification steps, are very sensitive and may therefore detect DNA from other life stages of the parasites (e.g., sporozoites recently injected by a vector or parasite DNA released into the blood from exoerythrocytic replication points). Sporozoites circulating in the blood are probably only rarely amplified by PCR although these can sometimes be seen on blood smears (Valkiūnas et al. 2009). Based on the calculations above, we can assume that 25 ng of total DNA corresponds to 0.004 μl blood. To be regularly picked up by PCR, sporozoites hence need to reach a density of 250 μl^{-1} (to have one sporozoite genome/25 ng of total DNA). For a ten-gram bird, this would correspond to a total of 250,000 sporozoites in the blood volume and would require bites from hundreds or thousands of infected vectors, which is probably a rare situation in nature. In contrast, it has become increasingly clear that primers for haemosporidians can amplify parasite DNA released into the blood from replication points in other organs (Olias et al. 2011). Importantly, molecular detections of parasites cannot distinguish whether infections are within competent host species or cases of aborted development of the parasites. A striking example comes from the parasite *Haemoproteus witti*, which uses hummingbirds as competent hosts, but is frequently amplified from various species of passerines in which gametocytes do not seem to develop (Moens et al. 2017). It has been shown that abortive development infections can severely harm both hosts (Olias et al. 2011) and vectors (Valkiūnas et al. 2014), thus sometimes presenting a higher selection pressure than competent parasites. Although infections resulting in abortive development might be dead ends for both the parasite and sometimes also for the host, such infections are of both ecological and evolutionary interest as they might represent the first signs of host range expansion; the parasite obviously has reached the capacity to establish and replicate to some degree in the novel host(s) (one of the first stages of the aggressive symbiosis hypothesis; Ryan 2009).

The traditional way to verify that a host or a vector is competent is by confirming presence of gametocytes in analyses of blood smears or sporozoites in the salivary glands of the vector. For human and rodent malaria (*P. falciparum* and *P. chabaudi*), there are already molecular methods that utilize genes that are specifically expressed in the gametocytes that can be used to detect if transmissible stages exist in the vertebrate host (Wargo et al. 2006; Babiker and Schneider 2008). It would be possible to develop similar protocols for detection of sporozoites due to expression differences of genes involved in different development stages in the vector (Roth et al. 2018). For avian haemosporidians, there are already several RNA-sequencing studies that can be used to look for life-stage-specific genes (Videvall et al. 2017; Weinberg et al. 2018) and soon, there will be transcriptome data for life stages in the vectors. However, a particular challenge for identifying life-stage-specific genes in avian haemosporidians is that the data comes from samples with un-synchronized life cycles (i.e., the samples include parasites that are in different life stages). To overcome this problem, there is a need to be able to culture the parasites in vitro, as is done in human malaria research. For the vectors, this problem might be less

important as it is possible to dissect out the parasites belonging to the different life-cycle stages before sequencing, (i.e., separating the blood meal with zygote formation, oocysts, and sporozoites into different sequencing batches). Once we have this information, it might be possible to develop protocols that identify stage-specific RNA profiles of the parasite, thus allowing us to identify whether the parasite has completed its life cycle in that specific host.

4.5.3 Population Structure

Today, we are only at the beginning of understanding the genetic structure of parasites that have identical *cytb* lineages. This question needs to be addressed with a large number of nuclear genes across the genome in order to evaluate levels of linkage disequilibrium and gene flow between potential populations. Many *cytb* lineages have been recorded across vast geographical areas as well as in taxonomically divergent hosts (Bensch et al. 2009). Do these cases represent large genetically unstructured parasite populations or do these consist of a multiple isolated population with no or very little gene flow occurring between them? The lineage SGS1, found in 126 species in 39 countries (June 2019), appears to have active transmission both in temperate regions in Eurasia as well as in tropical parts of Africa (Beadell et al. 2006; Hellgren et al. 2007b, 2015; Marzal et al. 2011). However, when investigating data from a fast-evolving nuclear gene (MSP1), the lineage SGS1 was found to consist of several populations, each with its own MSP1 haplotypes with different variants transmitted in tropical and temperate regions (Garcia-Longoria et al. 2015; Hellgren et al. 2015). Another example is the lineage SISKIN1 (*Haemoproteus tartakovskyi*) that is a common parasite of siskins (*Spinus spinus*) and crossbills in northern Eurasia, but also of house finches (*Haemorhous mexicanus*) in Mexico. Huang et al. (2019) analyzed ~1000 genes of SISKIN1 isolates from Europe and Mexico using the sequence capture technique. The complete mtDNA genomes were found to be very similar although not identical (6 differences corresponding to 0.1% divergence), whereas the average nuclear gene divergence was 20-fold higher (2.85%) between the European and Mexican isolates. This shows that minimal divergences in mtDNA in some cases are associated with highly divergent genomes. In order for future avian haemosporidian researchers to understand the epidemiology/epizootiology of the parasites, host range evolution, and processes of speciation, it will be important to identify the genetic structure of populations across transmission areas as well as across phylogenetically divergent hosts.

4.5.4 Molecular Phylogenies

To be able to study the direction of evolution (e.g., do generalists evolve from specialists or vice versa?), how selection is acting on certain genes or to be able to account for phylogenetic constraints when studying any life-history trait of the

parasite, we need robust phylogenies (see Chap. 3 for a thorough introduction to systematic and phylogenetic concepts and methods). For phylogenetic analyses, the MalAvi fragment is too short for obtaining well-supported topologies. Ellis and Bensch (2018) constructed ML trees for all the lineages that presently were in MalAvi ($n = 2451$) and determined that only 20% of the nodes had bootstrap support values $>70\%$. In order to obtain better-supported phylogenies, it would clearly be advantageous to amplify and sequence the whole *cytb* or even the whole mtDNA genome (Pacheco et al. 2018b), or – even better – to use multiple gene-sets across the whole genome of the parasites.

We will approach the ultimate goal of a robust phylogeny of all avian haemosporidians much faster if future studies will direct efforts toward generating data for the same set of nuclear genes. Borner et al. (2016) developed primers for 21 nuclear genes that can also be found in the published genomes of avian haemosporidians (Bensch et al. 2016; Böhme et al. 2018). These consist of a promising set of genes for future studies. As mentioned above, it is hard work to obtain data for nuclear genes by traditional PCR and sequencing. Such genes will also be biased toward those that evolve slowly, which on the one hand will be helpful for resolving deeper nodes in the phylogeny, but may not provide enough variation for resolving phylogenies of closely related species at the tips of the tree. As more genomes and transcriptomes become available for avian haemosporidians, they will facilitate researchers in developing specific protocols and molecular markers for “their” group of parasites, whether these efforts will be done by PCR and sequencing, sequence capture, or other methods based on reference genomes.

Acknowledgment We are indebted to Gediminas Valkiūnas, Diego Santiago-Alarcon, and Susan L. Perkins for reading earlier versions of this manuscript.

References

- Altschul SF, Gish W, Miller W et al (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Arai M, Billker O, Morris HR et al (2001) Both mosquito-derived xanthurenic acid and a host blood-derived factor regulate gametogenesis of *Plasmodium* in the midgut of the mosquito. *Mol Biochem Parasitol* 116:17–24
- Asghar M, Hasselquist D, Bensch S (2011) Are chronic avian haemosporidian infections costly in wild birds? *J Avian Biol* 42:530–537
- Babiker HA, Schneider P (2008) Application of molecular methods for monitoring transmission stages of malaria parasites. *Biomed Mater* 3:034007
- Barrow LN, Allen JM, Huang X et al (2019) Genomic sequence capture of haemosporidian parasites: methods and prospects for enhanced study of host-parasite evolution. *Mol Ecol Resour* 19:400–410
- Beadell JS, Fleischer RC (2005) A restriction enzyme-based assay to distinguish between avian hemosporidians. *J Parasitol* 91:683–685
- Beadell JS, Gering E, Austin J et al (2004) Prevalence and differential host-specificity of two avian blood parasite genera in the Australo-Papuan region. *Mol Ecol* 13:3829–3844

- Beadell JS, Ishtiaq F, Covas R et al (2006) Global phylogeographic limits of Hawaii's avian malaria. *Proc R Soc Lond B* 273:2935–2944
- Bensch S, Canbäck B, DeBarry JD et al (2016) The genome of *Haemoproteus tartakovskyi* and its relationship to human malaria parasites. *Genome Biol Evol* 8:1361–1373
- Bensch S, Coltman DW, Davis CS et al (2014) Genomic resources notes accepted 1 June 2013–31 July 2013. *Mol Ecol Resour* 14:218–218
- Bensch S, Hellgren O, Krizanuskiene A et al (2013) How can we determine the molecular clock of malaria parasites? *Trends Parasitol* 29:363–369
- 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* 9:1353–1358
- Bensch S, Perez-Tris J, Waldenstrom J et al (2004) Linkage between nuclear and mitochondrial DNA sequences in avian malaria parasites: multiple cases of cryptic speciation? *Evolution* 58:1617–1621
- Bensch S, Stjernman M, Hasselquist D et al (2000) Host specificity in avian blood parasites: a study of *Plasmodium* and *Haemoproteus* mitochondrial DNA amplified from birds. *Proc R Soc Lond B* 267:1583–1589
- Bernotienė R, Palinauskas V, Iezhova T et al (2016) Avian haemosporidian parasites (Haemosporida): a comparative analysis of different polymerase chain reaction assays in detection of mixed infections. *Exp Parasitol* 163:31–37
- Borner J, Burmester T (2017) Parasite infection of public databases: a data mining approach to identify apicomplexan contaminations in animal genome and transcriptome assemblies. *BMC Genomics* 18:100.
- Borner J, Pick C, Thiede J et al (2016) Phylogeny of haemosporidian blood parasites revealed by a multi-gene approach. *Mol Phylogenet Evol* 94:221–231
- Böhme U, Otto TD, Cotton JA et al (2018) Complete avian malaria parasite genomes reveal features associated with lineage-specific evolution in birds and mammals. *Genome Res* 28:547–560
- Chakarov N, Greiner JFW, Hauser S et al (2012) Label-free enrichment of avian *Leucocytozoon* using flow cytometric sorting. *Parasitology* 139:1547–1552
- Ciloglu A, Ellis VA, Bernotienė R et al (2019) A new one-step multiplex PCR assay for simultaneous detection and identification of avian haemosporidian parasites. *Parasitol Res* 118:191–201
- Ellis VA, 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
- Fallon SM, Bermingham E, Ricklefs RE (2003a) Island and taxon effects in parasitism revisited: avian malaria in the Lesser Antilles. *Evolution* 57:606–615
- Fallon SM, Ricklefs RE, Swanson BL et al (2003b) Detecting avian malaria: an improved polymerase chain reaction diagnostic. *J Parasitol* 89:1044–1047
- Farias MEM, Atkinson CT, LaPointe DA et al (2012) Analysis of the trap gene provides evidence for the role of elevation and vector abundance in the genetic diversity of *Plasmodium relictum* in Hawaii. *Malar J* 11:305
- Feehery GR, Yigit E, Oyola SO et al (2013) A method for selectively enriching microbial DNA from contaminating vertebrate host DNA. *PLoS One* 8(10):e76096
- Feldman RA, Freed LA, Cann RL (1995) A PCR test for avian malaria in Hawaiian birds. *Mol Ecol* 4:663–673
- Freed LA, Cann RL (2006) DNA quality and accuracy of avian malaria PCR diagnostics: a review. *Condor* 108:459–473
- Galen SC, Borner J, Martinsen ES et al (2018a) The polyphyly of *Plasmodium*: comprehensive phylogenetic analyses of the malaria parasites (order Haemosporida) reveal widespread taxonomic conflict. *R Soc Open Sci*. <https://doi.org/10.1098/rsos.171780>
- Galen SC, Nunes R, Sweet PR et al (2018b) 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.

- Garcia-Longoria L, Hellgren O, Bensch S (2014) Molecular identification of the chitinase genes in *Plasmodium relictum*. *Malar J* 13:239.
- Garcia-Longoria L, Hellgren O, Bensch S et al (2015) Detecting transmission areas of malaria parasites in a migratory bird species. *Parasitology* 142:1215–1220
- Gunderson JH, Sogin ML, Wollett G et al (1987) Structurally distinct, stage-specific ribosomes occur in *Plasmodium*. *Science* 238:933–937
- Hellgren O, Atkinson CT, Bensch S et al (2015) Global phylogeography of the avian malaria pathogen *Plasmodium relictum* based on MSP1 allelic diversity. *Ecography* 38:842–850
- Hellgren O, Krizanauskiene A, Hasselquist D et al (2011) Low haemosporidian diversity and one key-host species in a bird malaria community on a mid-Atlantic island (São Miguel, Azores). *J Wildl Dis* 47:849–859
- Hellgren O, Krizanauskiene A, Valkiūnas G et al (2007a) Diversity and phylogeny of mitochondrial cytochrome b lineages from six morphospecies of avian *Haemoproteus* (Haemosporida: Haemoproteidae). *J Parasitol* 93:889–896
- Hellgren O, Kutzer M, Bensch S et al (2013a) Identification and characterization of the merozoite surface protein 1 (*mSP1*) gene in a host-generalist avian malaria parasite, *Plasmodium relictum* (lineages SGS1 and GRW4) with the use of blood transcriptome. *Malar J* 12:381
- Hellgren O, Waldenström J, Bensch S (2004) A new PCR assay for simultaneous studies of *Leucocytozoon*, *Plasmodium*, and *Haemoproteus* from avian blood. *J Parasitol* 90:797–802
- Hellgren O, Waldenström J, Pérez-Tris J et al (2007b) Detecting shifts of transmission areas in avian blood parasites: a phylogenetic approach. *Mol Ecol* 16:1281–1290
- Hellgren O, Wood MJ, Waldenström J et al (2013b) Circannual variation in blood parasitism in a sub-Saharan migrant passerine bird, the garden warbler. *J Evol Biol* 26:1047–1059
- Hikosaka K, Kita K, Tanabe K (2013) Diversity of mitochondrial genome structure in the phylum Apicomplexa. *Mol Biochem Parasitol* 188:26–33
- Huang X, Hansson R, Palinauskas V 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
- Huang X, Rapševičius P, Chapa-Vargas L et al (2019) Within-lineage divergence of avian haemosporidians: a case study to reveal the origin of a widespread haemoproteus parasite. *J Parasitol* 105:414–422
- Ishtiaq F, Rao M, Huang X et al (2017) Estimating prevalence of avian haemosporidians in natural populations: a comparative study on screening protocols. *Parasit Vector* 10:127
- Jarvi SI, Farias MEM, Baker H et al (2003) Detection of avian malaria (*Plasmodium* spp.) in native land birds of American Samoa. *Conserv Genet* 4:629–637
- Kocher TD, Thomas WK, Meyer A et al (1989) Dynamics of mitochondrial-DNA evolution in animals – amplification and sequencing with conserved primers. *Proc Natl Acad Sci U S A* 86:6196–6200
- Kuo C-H, Wares JP, Kissinger JC (2008) The Apicomplexan whole-genome phylogeny: an analysis of incongruence among gene trees. *Mol Biol Evol* 25:2689–2698
- Lauron EJ, Oakgrove KS, Tell LA et al (2014) Transcriptome sequencing and analysis of *Plasmodium gallinaceum* reveals polymorphisms and selection on the apical membrane antigen-1. *Malar J* 13:382
- Leichty AR, Brisson D (2014) Selective whole genome amplification for resequencing target microbial species from complex natural samples. *Genetics* 198:473–481
- Lotta IA, Valkiūnas G, Pachecoc MA et al (2019) Disentangling *Leucocytozoon* parasite diversity in the neotropics: descriptions of two new species and shortcomings of molecular diagnostics for leucocytozooids. *Int J Parasitol Parasites Wildl* 9:159–173
- Lutz HL, Marra NJ, Grewe F et al (2016) Laser capture microdissection microscopy and genome sequencing of the avian malaria parasite, *Plasmodium relictum*. *Parasitol Res* 115:4503–4510
- Martinez C, Marzec T, Smith CD et al (2013) Identification and expression of *maebl*, an erythrocyte-binding gene, in *Plasmodium gallinaceum*. *Parasitol Res* 112:945–954

- Martinez J, Martinez-De La Puente J, Herrero J et al (2009) A restriction site to differentiate *Plasmodium* and *Haemoproteus* infections in birds: on the inefficiency of general primers for detection of mixed infections. *Parasitology* 136:713–722
- Martins ES, Perkins SL, Schall JJ (2008) A three-genome phylogeny of malaria parasites (*Plasmodium* and closely related genera): evolution of life-history traits and host switches. *Mol Phylogenet Evol* 47:261–273
- Marzal A, Ricklefs RE, Valkiūnas G et al (2011) Diversity, loss, and gain of malaria parasites in a globally invasive bird. *PLoS One* 6:e21905
- Moens MAJ, Valkiūnas G, Paca A et al (2017) Parasite specialization in a unique habitat: hummingbirds as reservoirs of generalist blood parasites of Andean birds (vol 85, pg 1234, 2016). *J Anim Ecol* 86:171–171
- Musa S, Mackenstedt U, Woog F et al (2018) Avian malaria on Madagascar: prevalence, biodiversity and specialization of haemosporidian parasites. *Int J Parasitol.* <https://doi.org/10.1016/j.ijpara.2018.11.001>
- Nilsson E, Taubert H, Hellgren O et al (2016) Multiple cryptic species of sympatric generalists within the avian blood parasite *Haemoproteus majoris*. *J Evolution Biol* 29:1812–1826
- Olias P, Wegelin M, Zenker W et al (2011) Avian malaria deaths in parrots, Europe. *Emerg Infect Dis* 17:950–952
- Owen JC (2011) Collecting, processing, and storing avian blood: a review. *J Field Ornithol* 82:339–354
- Pacheco MA, Cepeda AS, Bernotienė R et al (2018a) Primers targeting mitochondrial genes of avian haemosporidians: PCR detection and differential DNA amplification of parasites belonging to different genera. *Int J Parasitol* 48:657–670
- Pacheco MA, Matta NE, Valkiūnas G et al (2018b) Mode and rate of evolution of haemosporidian mitochondrial genomes: timing the radiation of avian parasites. *Mol Biol Evol* 35:383–403
- Palinauskas V, Dolnik OV, Valkiūnas G et al (2010) Laser microdissection microscopy and single cell PCR of avian haemosporidians. *J Parasitol* 96:420–424
- Palinauskas V, Krizanauskiene A, Iezhova TA 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
- Pauli M, Chakarov N, Rupp O et al (2015) De novo assembly of the dual transcriptomes of a polymorphic raptor species and its malarial parasite. *BMC Genomics* 16:1038
- Perez-Tris J, Bensch S (2005) Diagnosing genetically diverse avian malarial infections using mixed-sequence analysis and TA-cloning. *Parasitology* 131:15–23
- Perkins SL, Schall JJ (2002) A molecular phylogeny of malarial parasites recovered from cytochrome b gene sequences. *J Parasitol* 88:972–978
- Pierce MA (1981) Distribution and host-parasite check-list of haematozoa of birds in Western Europe. *J Nat Hist* 15:419–458
- Reullier J, Perez-Tris J, Bensch S et al (2006) Diversity, distribution and exchange of blood parasites meeting at an avian moving contact zone. *Mol Ecol* 15:753–763
- Ricklefs RE, Outlaw DC (2010) A molecular clock for malaria parasites. *Science* 329:226–229
- Rivero A, Gandon S (2018) Evolutionary ecology of avian malaria: past to present. *Trends Parasitol* 34:712–726
- Rooney AP (2004) Mechanisms underlying the evolution and maintenance of functionally heterogeneous 18S rRNA genes in Apicomplexans. *Mol Biol Evol* 21:1704–1711
- Roth A, Adapa SR, Zhang M et al (2018) Unraveling the *Plasmodium vivax* sporozoite transcriptional journey from mosquito vector to human host. *Sci Rep* 8:12183
- Ryan F (2009) *Virolution*. Harper Collins, London
- Santiago-Alarcon D, Outlaw DC, Ricklefs RE et al (2010) Phylogenetic relationships of haemosporidian parasites in New World Columbiformes, with emphasis on the endemic Galapagos dove. *Int J Parasitol* 40:463–470
- Schmid S, Facht K, Dinkel A et al (2017) Carrion crows (*Corvus corone*) of southwest Germany: important hosts for haemosporidian parasites. *Malar J* 16:369

- Sehgal RNM, Hull AC, Anderson NL et al (2006) Evidence for cryptic speciation of *Leucocytozoon* spp. (Haemosporida, Leucocytozoidae) in diurnal raptors. *J Parasitol* 92:375–379
- Stjernman M, Raberg L, Nilsson J-A (2008) Maximum host survival at intermediate parasite infection intensities. *PLoS One* 3(6):e2463
- Valkiūnas G (2005) Avian malaria parasites and other haemosporidia. CRC, Boca Raton
- Valkiūnas G, Bensch S, Iezhova TA et al (2006) Nested cytochrome b polymerase chain reaction diagnostics underestimate mixed infections of avian blood haemosporidian parasites: microscopy is still essential. *J Parasitol* 92:418–422
- Valkiūnas G, Iezhova TA, Loiseau C et al (2009) Nested cytochrome b polymerase chain reaction diagnostics detect sporozoites of hemosporidian parasites in peripheral blood of naturally infected birds. *J Parasitol* 95:1512–1515
- Valkiūnas G, Iezhova TA, Palinauskas V et al (2015) The evidence for rapid gametocyte viability changes in the course of parasitemia in *Haemoproteus* parasites. *Parasitol Res* 114:2903–2909
- Valkiūnas G, Ilgunas M, Bukauskaitė D et al (2016) Description of *Haemoproteus ciconiae* sp. nov. (Haemoproteidae, Haemosporida) from the white stork *Ciconia ciconia*, with remarks on insensitivity of established polymerase chain reaction assays to detect this infection. *Parasitol Res* 115:2609–2616
- Valkiūnas G, Kazlauskienė R, Bernotienė R et al (2014) *Haemoproteus* infections (Haemosporida, Haemoproteidae) kill bird-biting mosquitoes. *Parasitol Res* 113:1011–1018
- Valkiūnas G, Palinauskas V, Krizanauskienė A et al (2013) Further observations on in vitro hybridization of hemosporidian parasites: patterns of ookinete development in *Haemoproteus* spp. *J Parasitol* 99:124–136
- Videvall E (2019) Genomic advances in avian malaria research. *Trends Parasitol.* <https://doi.org/10.1016/j.pt.2018.12.005>
- Videvall E, Cornwallis CK, Ahren D et al (2017) The transcriptome of the avian malaria parasite *Plasmodium ashfordi* displays host-specific gene expression. *Mol Ecol* 26:2939–2958
- Waldenström J, Bensch S, Hasselquist D et al (2004) A new nested polymerase chain reaction method very efficient in detecting *Plasmodium* and *Haemoproteus* infections from avian blood. *J Parasitol* 90:191–194
- Ward RD (2009) DNA barcode divergence among species and genera of birds and fishes. *Mol Ecol Resour* 9:1077–1085
- Wargo AR, Randle N, Chan BHK et al (2006) *Plasmodium chabaudi*: reverse transcription PCR for the detection and quantification of transmission stage malaria parasites. *Exp Parasitol* 112:13–20
- Weinberg J, Field JT, Ilgunas M et al (2018) De novo transcriptome assembly and preliminary analyses of two avian malaria parasites, *Plasmodium delichoni* and *Plasmodium homocircumflexum*. *Genomics.* <https://doi.org/10.1016/j.ygeno.2018.12.004>
- Wilson R, Williamson DH (1997) Extrachromosomal DNA in the apicomplexa. *Microbiol Mol Biol Rev* 61:1–16
- Zehtindjiev P, Krizanauskiene A, Bensch S et al (2012) A new morphologically distinct avian malaria parasite that fails detection by established polymerase chain reaction-based protocols for amplification of the cytochrome gene. *J Parasitol* 98:657–665