Genomic sequence capture of haemosporidian parasites: Methods and prospects for enhanced study of host-parasite evolution

Lisa N. Barrow[1*], Julie M. Allen[2], Xi Huang[3], Staffan Bensch[3], Christopher C. Witt[1]

[1]Museum of Southwestern Biology and Department of Biology, MSC03 2020, 1 University of New Mexico, Albuquerque, New Mexico, 87131-0001, USA

[2]Department of Biology, University of Nevada, Reno, Nevada, 89557, USA

[3]Department of Biology, Molecular Ecology and Evolution Laboratory, Lund University, SE-223 62, Lund, Sweden

*Corresponding Author: Lisa N. Barrow, E-mail: lnbarrow@unm.edu

Abstract

Avian malaria and related haemosporidians (*Plasmodium*, *[Para]Haemoproteus*, and *Leucocytoozoon*) represent an exciting multi-host, multi-parasite system in ecology and evolution. Global research in this field accelerated after 1) the publication in 2000 of PCR protocols to sequence a haemosporidian mitochondrial (mtDNA) barcode, and 2) the development in 2009 of an open-access database to document the geographic and host ranges of parasite mtDNA haplotypes. Isolating haemosporidian nuclear DNA from bird hosts, however, has been technically challenging, slowing the transition to genomic-scale sequencing techniques. We extend a recently-developed sequence capture method to obtain hundreds of haemosporidian nuclear loci from wild bird samples, which typically have low levels of infection, or parasitemia. We tested 51 infected birds from Peru and New Mexico and evaluated locus recovery in light of variation in parasitemia, divergence from reference sequences, and pooling strategies. Our method was successful for samples with parasitemia as low as ~0.03% (3 of 10,000 blood cells infected) and mtDNA divergence as high as 15.9% (one *Leucocytozoon* sample), and using the most cost-effective pooling strategy tested. Phylogenetic relationships estimated with >300 nuclear loci were well resolved, providing substantial improvement over the mtDNA barcode. We provide protocols for sample preparation and sequence capture including custom probe kit sequences, and describe our bioinformatics pipeline using aTRAM 2.0, PHYLUCE, and custom Perl and Python scripts. This approach can be applied to the tens of thousands of avian samples that have already been screened for haemosporidians, and greatly improve our understanding of parasite speciation, biogeography, and evolutionary dynamics.

Keywords

Haemoproteus, avian malaria, hybrid enrichment, *Leucocytozoon*, host-parasite relationships, Apicomplexa

Introduction

Multi-host, multi-parasite systems provide extensive opportunities to advance research in ecology and evolution. Haemosporidians (malaria and relatives, Order Haemosporida), the intracellular, protozoan parasites that infect vertebrates, are one great example, with studies ranging in scope from regional and temporal patterns of community turnover (e.g., Fallon *et al.* 2004, 2005; Olsson-Pons *et al.* 2015; Fecchio *et al.* 2017), to host-switching and diversification across long evolutionary timescales (e.g., Martinsen *et al.* 2008; Ricklefs *et al.* 2014; Galen *et al.* 2018a; Pacheco *et al.* 2018). Avian haemosporidians in particular (genera *Plasmodium*, [*Para*]*Haemoproteus*, and *Leucocytoozoon*) have attracted a large research community seeking to describe global patterns of diversity, abundance, and host range, and uncover mechanisms underlying parasite diversification, host-switching, and host susceptibility (e.g., Scheuerlein & Ricklefs 2004; Bensch *et al.* 2009; Clark *et al.* 2014; Lutz *et al.* 2015). The latter goal has particular importance for avian conservation, as exemplified by the Hawaiian honeycreepers, which have been severely impacted by the introduction of avian malaria (van Riper *et al.* 1986; Atkinson & LaPointe 2009).

The detection and description of avian blood parasites have accelerated with the application of molecular methods. While microscopy of thin blood smears remains essential for morphological verification and detailed species descriptions (Valkiunas 2005; Valkiŭnas *et al.* 2008), the field benefited substantially from the development of PCR primers for avian haemosporidians (Bensch *et al.* 2000; Fig. 1). Subsequent nested PCR protocols based on these primers (Hellgren *et al.* 2004; Waldenström *et al.* 2004) enable researchers to amplify and sequence a mitochondrial (mtDNA) barcode fragment, 478 base pairs of cytochrome *b* (*cytb*), from avian blood or tissue samples, even when infection levels (i.e., parasitemia) are too low to

detect by microscopy. Parasite barcode sequences can then be compared with and uploaded to the avian haemosporidian database, MalAvi (Bensch *et al.* 2009). The growth of this database over the last decade has allowed for global analyses of parasite distributions and community assembly (Clark *et al.* 2014, 2017; Clark 2018; Ellis & Bensch 2018). It has become clear, however, that incorporating multiple nuclear loci will be necessary to further advance haemosporidian research.

Relatively few studies thus far have included multi-locus nuclear data of haemosporidian parasites (Fig. 1). Studies demonstrate that the *cytb* barcode provides limited resolution; a single *cytb* haplotype can include multiple cryptic species (Falk *et al.* 2015; Galen *et al.* 2018b), and phylogenies estimated from multiple nuclear loci substantially improve inferences of evolutionary relationships (Borner *et al.* 2016; Galen *et al.* 2018a). Several challenges, however, have previously prevented any large-scale efforts to obtain genomic data from avian haemosporidians. In contrast to mammalian red blood cells, avian red blood cells are nucleated, and the ratio of host to parasite DNA can be as high as a million to one (Perkins 2014). High-throughput sequencing of genomic DNA from avian samples is thus inefficient. Isolation of parasite gametocytes is possible through laser microdissection microscopy (Palinauskas *et al.* 2010), though this process is time-consuming and requires sufficient material. Parasite and host DNA can also be separated by inducing in vitro exflagellation of gametocytes, followed by centrifugation (Palinauskas *et al.* 2013), but donor birds with fairly high levels of infection are needed. Furthermore, avian haemosporidians are highly divergent from one another, with mtDNA barcode divergences between genera of ~10–20%, making the task of designing general primers for multiple nuclear loci somewhat intractable. An ideal method would: 1) work on any sample that contained preserved DNA, including the tens of thousands of frozen blood and tissue samples that have been screened for haemosporidians, 2) be broadly applicable across a diverse group of

haemosporidian parasites, and 3) enable cost-effective sequencing of hundreds of haemosporidian nuclear loci.

Sequence-capture methods used in conjunction with high-throughput sequencing are rapidly resolving the evolutionary Tree of Life for a variety of vertebrate, invertebrate, and plant taxa (Faircloth *et al.* 2012; Lemmon *et al.* 2012; Buddenhagen *et al.* 2016; Hamilton *et al.* 2016; Faircloth 2017; Quattrini *et al.* 2018). These techniques allow for the enrichment of genomic regions of interest by hybridizing oligonucleotide probes to genomic samples and removing non-target regions prior to sequencing (Albert *et al.* 2007; Gnirke *et al.* 2009). Probe sets can be designed from any existing genomic resources and are often useful across divergent taxa, although locus recovery tends to decline with increasing levels of divergence (Lemmon *et al.* 2012; Huang *et al.* 2018). The first genome for an avian *Haemoproteus* parasite was published in 2016 (Bensch *et al.* 2016), and no genomes for *Leucocytozoon* are available thus far. Given the vast differences between bird and haemosporidian genomes, the prospects are promising for targeted sequence capture of parasite genes from infected bird samples.

Huang *et al.* (2018) applied the first sequence-capture assay to haemosporidians, including eight *Haemoproteus* and one *Plasmodium* lineage, primarily from Europe and Asia. They successfully sequenced >100 nuclear exons from samples with up to ~6% mtDNA divergence from the reference, *H. tartakovskyi*. It is not yet known, however, whether this approach will be useful for sequencing low-level infections that are most commonly observed in naturally infected birds. The minimum parasitemia tested in Huang *et al.* (2018) was 0.25%, while most naturally infected birds exhibit parasitemia less than 0.1% (Atkinson *et al.* 2001; Zehtindjiev *et al.* 2008; Ishtiaq *et al.* 2017). It is possible that when parasitemia is too low, parasite DNA will be overwhelmed by host bird DNA.

Our primary goal was to design a cost-effective sequence capture assay to work broadly across the genus *Haemoproteus* because of its global abundance, diversity, and variation in host specificity. Secondarily, we include promising results from a single *Leucocytozoon* sample, and generate nuclear sequences for this genus that can be incorporated into subsequent probe designs. Our specific objectives were to: 1) describe the relationship between parasitemia levels and sequence-capture success, 2) test how sequence-capture success is affected by percent divergence from the reference sequences used for probe design, and 3) compare strategies for pooling samples before capture to increase cost-effectiveness. To facilitate use of this method by scientists studying avian haemosporidians, we provide detailed laboratory protocols, including probe-kit sequences. We also describe our bioinformatics pipeline using aTRAM 2.0 (Allen *et al.* 2015, 2018) for locus assembly, and PHYLUCE (Faircloth 2015) and custom PERL and Python scripts for downstream processing.

Materials and Methods

Locus selection and probe design

The recently sequenced genome of *H. tartakovskyi* was used for initial probe design and sequence capture, targeting 1,000 genes (Bensch *et al.* 2016; Huang *et al.* 2018). We used the parasite-enriched sequences from three *Haemoproteus* species produced with this initial probe kit as references to design the probe kit used in this study. These three species provide representative variation across a large portion of the *Haemoproteus* phylogeny, with up to 6.5% sequence divergence in the mtDNA *cytb* barcode between them.

Paired reads obtained from *H. tartakovskyi* (lineage SISKIN1, sample ID 126/11c), *H. majoris* (PARUS1, 1ES86798), and *H. nucleocondensus* (GRW01, 512022) captures were

trimmed, set as paired reads, and mapped to the initial 1,000 *H. tartakovskyi* genes in Geneious 8.1.9 (Biomatters Ltd). We selected 498 loci that were successfully captured and sequenced for at least one of the non-*tartakovskyi* samples, using a threshold of 3X coverage with no mismatches and alignment lengths of at least 200 base pairs (bp) as cutoffs for success. Most loci (471 out of 498) included three reference species, and the average alignment lengths were 1,251 bp (range: 232–8,319; total targeted bp: 622,788; Supplementary Table S1). Locus alignments of the three species were submitted to MYcroarray (now Arbor Biosciences, Ann Arbor, MI) for design and synthesis of a custom MYbaits kit with 19,973 biotinylated RNA probes and 2X tiling.

Sample selection and quantification

We selected 51 bird samples for sequence capture; 50 with putative single infections of known *Haemoproteus* lineages and one with a mixed infection of *Leucocytozoon* and *Haemoproteus*. All samples consisted of pectoral muscle previously collected from wild birds in Peru or New Mexico, USA in accordance with approved animal care guidelines and permits. Samples were stored at -80°C in the Museum of Southwestern Biology Division of Genomic Resources at the University of New Mexico. Genomic DNA was extracted using an Omega Bio-tek EZNA Tissue DNA Kit following manufacturer protocols. For initial assessment of infection, three nested PCR protocols were used to maximize detection of all three parasite genera (Hellgren *et al.* 2004; Waldenström *et al.* 2004). Positive infections were identified by visualizing PCR products on an agarose gel, and haplotypes were assigned by sequencing the 478-base pair haemosporidian mtDNA barcode (*cytb*), as described in Marroquin-Flores *et al.* (2017). We chose samples infected with 14 *Haemoproteus* lineages (9 Peru, 5 New Mexico) and one *Leucocytozoon* lineage from Peru for subsequent quantification, capture, and sequencing (Table 1; Table S2).

Overall sample DNA concentrations were quantified using a Qubit 3.0 Fluorometer. Relative parasite DNA concentrations were assessed using quantitative PCR with primers targeting a 154-base pair portion of haemosporidian ribosomal RNA (Fallon *et al.* 2003). We used 2X iTaq Universal SYBR Green Supermix (Bio-Rad), 0.5 µM of each primer (343F and 496R), and 30 ng sample DNA in a total reaction volume of 20 µL. Reactions were run on a Bio-Rad CFX96 Real-Time PCR System with the following temperature profile: 95 °C for 3 min, 40 cycles of 95 °C for 15 sec and 57°C for 1 min, followed by a melt curve analysis (47 °C to 95 °C at 0.5 °C and 5 sec per cycle). Each plate included three no-template controls. To generate a standard curve, we made a six-step 1:10 serial dilution (30–0.0003 ng DNA per well) of one sample with high parasitemia as estimated by microscopy (NK168012; 1.86% cells infected). Each sample was run in triplicate and cycle threshold (CT) values were averaged across the three replicates.

Library preparation, capture, and sequencing

We prepared libraries for each sample using the KAPA Hyper Prep Kit (Kapa Biosystems) and dual-indexing with the iTru system (Faircloth & Glenn 2012; Glenn *et al.* 2016; baddna.uga.edu). Complete protocols are provided in Supporting Information. Briefly, genomic DNA was visualized on an agarose gel to verify high molecular weight and determine a suitable sonication protocol. Samples were then fragmented to a length distribution centered on ~500 bp using a Covaris M220 Focused-ultrasonicator (Covaris, Inc.). Libraries were prepared primarily following the KAPA Hyper Prep Kit protocols, using 250 ng of fragmented DNA per sample, custom indexed adapters, KAPA Pure Beads for bead clean-ups, and 10 cycles in the indexing amplification step. After quantifying libraries by Qubit, equal amounts of sample libraries were

combined in pools of eight, four, two, or a single sample for capture. We only pooled libraries with similar parasitemia values as determined by qPCR in an attempt to obtain even capture success and sequencing coverage across samples within a pool. Capture pools contained 1–2 µg DNA (at least 125 ng per library).

Hybrid enrichment was performed following the MYbaits Version 3.02 protocols with minor modifications as follows. Block Mix 3 was prepared from custom oligos for the iTru dual-indexing system, and Chicken Hybloc DNA (Applied Genetics Laboratories, Inc.) was used as Block Mix 1. We extended the hybridization time to 36–40 hr, as recommended to increase capture efficiency for low-abundance targets. For the post-capture amplification, we used 2X KAPA HiFi HotStart ReadyMix with the bead-bound library and the following thermal profile: 98 °C for 2 min, 16 cycles of 98 °C for 20 sec, 60 °C for 30 sec, and 72 °C for 60 sec, followed by a final extension at 72 °C for 5 min and held at 4 °C. Post amplification, we removed the beads and performed a final 1.2X KAPA Pure Bead clean-up. Captured pools were quantified and characterized by Qubit and an Agilent 2100 Bioanalyzer, and shipped to the Oklahoma Medical Research Foundation (OMRF) Clinical Genomics Center for final qPCR and sequencing. All capture pools were combined and run on a single lane of PE150 Illumina HiSeq 3000.

Bioinformatic processing

Demultiplexed reads for each sample were obtained from the OMRF Clinical Genomics Center. We trimmed adapters and low-quality bases using default settings in Illumiprocessor 2.0.6 (Faircloth 2013), which provides a wrapper for Trimmomatic (Bolger *et al.* 2014). These settings include trimming reads with lengths <40 (MINLEN:40), bases at the start of a read with quality scores <5 (LEADING:5), bases at the end of a read with scores <15 (TRAILING:15), and

bases in a sliding window where four consecutive bases have an average quality <15 (SLIDINGWINDOW:4:15). We then used the automated Target Restricted Assembly Method, aTRAM 2.0 (Allen *et al.* 2015, 2018; http://www.github.com/juliema/aTRAM) to assemble haemosporidian parasite genes using the 498 reference *Haemoproteus* gene sequences. This approach uses local BLAST searches and an iterative approach to produce assemblies for genes of interest from cleaned read data. We used BLAST 2.7.1 (Altschul *et al.* 1990), Trinity 2.0.6 (Grabherr *et al.* 2011) as the assembler, five iterations, and nucleotide reference sequences from *H. tartakovskyi* for all individuals. We also conducted two additional tests to improve locus recovery for individuals with higher divergences from the reference. First, we used amino acid reference sequences instead of nucleotide for aTRAM assemblies, but found that several bird host genes were assembled in place of the haemosporidian genes. Second, we used *H. majoris* as the nucleotide reference, and added the new locus assemblies recovered to the set for further processing.

We next used custom scripts written in Perl and Python (available at https://github.com/juliema/) to keep only the contigs from the last aTRAM iteration, compare and align them to the translated exon sequences for the reference *H. tartakovskyi* using Exonerate 2.2.0 (Slater & Birney 2005), and stitch together any exons that were broken into multiple contigs (as described in Allen *et al.* 2017). Because aTRAM performs iterative assemblies that can extend outward from the original reference sequence, the last iteration is most likely to have the most complete, longest contigs. For samples where the middle of a locus was either not sequenced or not recovered due to low coverage, we used exon stitching to retain both ends of the sequence for alignment. We then conducted a reciprocal best-BLAST check on the stitched exons, and removed any individual-locus combination for which the top match for the assembled

223 locus was not the target locus. For this search, we created a local BLAST database from all 6,436
224 *H. tartakovskyi* genes (downloaded from http://mbio-serv2.mbioekol.lu.se/Malavi/Downloads).
225 Fewer than 0.2% (9 of 4,507) individual-locus combinations were mismatched and therefore
226 removed. Prior to multiple sequence alignment, we added sequences for the three reference
227 species, and used custom Python scripts (available on Dryad) to reformat the aTRAM sequences
228 for the PHYLUCE pipeline (Faircloth 2015).

229 Several PHYLUCE scripts were used to summarize locus information for each individual,
230 produce multiple sequence alignments, and generate concatenated datasets for phylogenetic
231 analysis in RAxML. We considered samples with at least 50 recovered loci to be successful, and
232 generated alignments including only those individuals. We generated edge-trimmed alignments
233 with MAFFT 7.130b (Katoh & Standley 2013), using a threshold parameter of 0.3 (at least 30%
234 of individuals with sequence at the edges) and maximum divergence of 0.4. As one final check,
235 we manually examined all alignments with at least 50% of taxa (404 alignments), and removed
236 25 that were poorly-aligned. Concatenated RAxML alignments were generated allowing different
237 levels of missing data: at least 50, 70, or 90% of individuals per locus.

238

239 *Sequence coverage estimation*

240 To provide an estimate of sequence coverage and its influence on capture success, we
241 used the BLAST-only option in aTRAM to find and count the reads for a parasite mtDNA gene
242 and a host mtDNA gene for each sample. The *H. tartakovskyi* reference sequence was used to
243 estimate coverage for the parasite mtDNA *cytb* gene. Given the availability of host bird
244 sequences on GenBank, we used *ND2* reference sequences (Table S2) to estimate host mtDNA
245 coverage. Host genes were not targeted by the probe kit, but these non-target reads were

sequenced as a by-product because of the large quantity of bird mtDNA in the samples. We estimated per-site coverage based on the length of the reference gene used, assuming 150 base pair read lengths, and compared the ratio of parasite to host per-site coverage between samples that were considered capture successes and failures.

Detection of mixed infections

For each successfully-captured individual, we mapped the cleaned, paired reads to both *COI* and *cytb* mtDNA reference sequences in Geneious to check for possible co-infections. In cases where multiple haplotypes were apparent in the mapped read assembly, we compared the reads to the *cytb* barcode region for the original assigned haplotype to sort out the alternative haplotype and determine the variant frequency.

Downstream analysis

To test for effects of multiple variables on sequence capture success, we used generalized linear models (GLMs) in R (R Core Team 2016). We tested whether parasitemia (qPCR CT value), level of divergence from the nearest reference (% mtDNA sequence divergence), or the number of samples pooled (factor with two categories: 8 or <8), had an effect on the number of loci recovered per sample. We also included the interaction between parasitemia and divergence from the reference. The one *Leucocytozoon* sample was an outlier for mtDNA divergence, thus we repeated analyses with and without this individual.

To determine whether the nuclear loci we recovered improve inferences of haemosporidian relationships, we estimated a phylogeny for the samples with sufficient data. We used PartitionFinder 2 (Lanfear *et al.* 2017) to select appropriate models of nucleotide evolution

and partitioning schemes for each dataset. Given the number of loci, we used the rcluster algorithm with RAxML (Stamatakis 2014; Lanfear *et al.* 2014), linked branch lengths, and AICc for model selection. Phylogenies were estimated for each nuclear dataset (50, 70, 90% complete matrices), for the mtDNA capture data (3,226 bp of COI and *cytb*), and for the original *cytb* barcode data (478 bp). We used the rapid hill-climbing algorithm in RAxML with the GTR+G model and 1000 bootstrap replicates. We also estimated species trees from the nuclear datasets using SVDquartets (Chifman & Kubatko 2014), implemented in PAUP* 4.0a163 (Swofford 2002). We performed an exhaustive search of all quartets and conducted multilocus bootstrapping with 1000 replicates and partitioned loci.

Results

Data summary

We obtained 620,951,640 reads from one sequencing lane, of which 571,186,910 (92%) were sorted by individual barcode. On average, 11.2 million (s.d.: ± 7.2 million) reads were obtained per individual (min–max: 3.7–39.3 million; median: 8.7 million). The number of parasite loci assembled per individual ranged from 491 (99%) to none; eight of the 51 samples resulted in no *Haemoproteus* loci. We considered 15 samples (29%) to be sequence capture successes, with >70 loci obtained. The remaining individuals resulted in 11 or fewer loci. For most of these failures, only one or two mtDNA genes were assembled. For the successful samples, an average of 295 ± 140 (min–max: 71–491, median: 254) loci were recovered with mean locus lengths of 2,428 (min–max: 2,317–2,798) bp. The average number of loci assembled with >1,000 bp in length was 241 ± 108 (min–max: 64–386; median: 212; Fig. S1).

Effects of parasitemia, divergence, and pooling on success

Parasitemia was positively correlated with capture success and locus recovery ($t = 6.15$, $p < 0.0001$; Fig. 2a). Samples with parasitemia values >0.07% (~7 out of 10,000 infected cells) were all successful, and samples with as low as ~0.03% (~3 of 10,000 cells; qPCR CT value ~26) also resulted in >100 loci. Divergence from the nearest reference was negatively correlated with locus recovery ($t = -3.91$, $p = 0.0003$; Fig. 2b). There was also a significant interaction between parasitemia and divergence ($t = -3.76$, $p = 0.0005$). Samples with both sufficient parasitemia and low divergence from the reference had the best locus recovery (>400 loci). The most divergent capture success was the *Leucocytozoon* sample, with 15.9% mtDNA divergence from the nearest reference and 71 loci recovered. GLM results excluding this outlier were qualitatively similar but stronger in magnitude. The number of samples included in a pool did not affect capture success ($t = 0.68$, $p = 0.5$); several individuals in the most cost-effective, 8-sample pools resulted in >100 sequenced loci.

Parasite versus host coverage

The estimated depth of sequence coverage per site for parasite mtDNA showed substantial variation across samples, with a mean of 2,804 reads (min–max: 0.39–45,683). Samples that were considered capture successes had a mean parasite mtDNA per-site coverage of 9,516 (185–45,683) and mean host mtDNA coverage of 856 (1.73–2,540). In contrast, samples that were considered capture failures had a mean parasite mtDNA coverage of only 8.22 (0.39–125), while host mtDNA coverage was similar with a mean of 957 (0.87–3,110). On average, the ratio of parasite to host coverage was 137 (0.09–1,622) for capture successes, and only 0.32 (0.0002–11.15) for capture failures.

315

316 *Detection of mixed infections*

317 We identified three samples co-infected with multiple lineages, two of which had not been
318 previously detected by nested PCR and Sanger sequencing. Sample NK168883 was co-infected
319 with the *Haemoproteus* lineage G009 and a novel *Leucocytozoon* lineage (assigned name G403).
320 Within the MalAvi barcode region, the frequency of the *Leucocytozoon* lineage dominated the
321 reads and ranged from 94.8–98.0%. The concatenated sequence for this sample had >15%
322 sequence divergence from the pure G009 *Haemoproteus* sample (NK168881), indicating that the
323 genes assembled for that sample belonged to *Leucocytozoon*. Sample NK275890 was co-infected
324 with two *Haemoproteus* lineages, SPIPAS01 and SIAMEX01 (the read frequency of SIAMEX01
325 was 11.0–14.2%; mtDNA divergence between the two haplotypes of 6.1%). Sample NK276102
326 was also co-infected with two *Haemoproteus* lineages, VIGIL05 and VIGIL07 (the read
327 frequency of VIGIL07 was 30.3–40.5%; mtDNA divergence 3.6%). Phylogenetic analyses were
328 repeated without the two *Haemoproteus* co-infected samples because the loci extracted by
329 aTRAM may have represented a composite of the two haplotypes that could not be definitively
330 sorted.

331

332 *Phylogenetic resolution*

333 The nuclear datasets substantially improved phylogenetic resolution of *Haemoproteus*
334 parasites (Fig. 3). Most relationships were consistent among datasets with different levels of
335 completeness. The 50% complete matrix included 377 loci and 287,164 bp (Fig. 3a), the 70%
336 matrix included 206 loci and 189,883 bp, and the 90% matrix included 59 loci and 70,611 bp
337 (Fig. 3b). The topologies resulting from RAxML and SVDquartets were identical for the 50% and

70% datasets. The position of *H. tartakovskyi* differed for the 90% matrix RAxML analysis (Fig. 3b), while the SVDquartets topology was consistent with the other datasets. Phylogenetic analyses excluding the *Haemoproteus* mixed infections resulted in similar topologies, except for the uncertain position of TROAED12 (Fig. S2). The majority of nodes had high support for the nuclear datasets; bootstrap values were ≥ 95 for 15 (94%) nodes with the 50% matrix and 13 (81%) nodes with the 90% matrix. In contrast, the two-gene mtDNA dataset and the *cytb* barcode produced poorly-resolved phylogenies for the lineages in our study, inferring very few relationships with any certainty (Fig. 3c,d). Only 6 (38%) and 3 (19%) of the nodes had bootstrap values ≥ 95, respectively.

Discussion

Capture success and parasitemia

The extremely low abundance of parasite DNA compared to host DNA has presented a great challenge for obtaining haemosporidian genomic data from naturally-infected birds. We provide parameters for the successful implementation of a new sequence-capture assay to obtain hundreds of haemosporidian parasite loci from wild bird samples. By quantifying parasitemia using either a standard qPCR protocol or microscopic examination, researchers can select samples above a certain threshold to enable capture success. Based on our results, samples with ~0.07% parasitemia were always successful, and samples with as low as ~0.03% parasitemia also tended to be successful.

This finding is exciting because tens of thousands of avian blood and tissue samples exist in museums and laboratories globally and have already been screened for haemosporidians. Our results indicate that a large portion of these samples (29% in our study) will be suitable for

sequence capture of haemosporidian parasites. One possible improvement for sequence capture of samples with even lower parasitemia is to perform a double enrichment using the haemosporidian probe kit, in order to increase the relative amount of parasite DNA in the samples further. If a single sample is captured at a time, parasite and host DNA could accurately be quantified for each sample with qPCR before and after each enrichment. For a more cost-effective approach, however, we recommend quantifying parasitemia with qPCR prior to capture, and pooling sample libraries with similar values as we have done here, in order to obtain more even sequencing coverage across samples.

Potential improvements and cost

Our probe kit was designed to work broadly across *Haemoproteus* because of our interest in the diversity and host range variation of this genus worldwide (Clark *et al.* 2014; Ellis & Bensch 2018) and the genomic resources available to design probes (Bensch *et al.* 2016; Huang *et al.* 2018). A clear extension of this method is to incorporate probes targeting all avian haemosporidian genera. Although we have not yet tested our probe kit on *Plasmodium*-infected birds, the successful capture and sequencing of >70 loci for *Leucocytozoon* is quite promising because avian *Plasmodium* is less divergent from *Haemoproteus*. Additional tests on *Leucocytozoon*- and *Plasmodium*-infected samples with the current probe kit can be carried out to determine success rates, and the generated sequences can be incorporated into new probe designs along with new and existing transcriptome and genome data for the other genera (Lutz *et al.* 2016; Videvall *et al.* 2017; Böhme *et al.* 2018).

As reference data for more parasite lineages are generated, it will likely be possible to improve certain steps of our bioinformatics pipeline. We tested two different reference species for

aTRAM assemblies, and found that the reference chosen for assembly has some effect on the number of loci recovered; initial locus recovery was better for samples that were less divergent from the *H. tartakovskyi* reference. We assembled >425 loci (>85%) for the samples with the lowest divergence from *H. tartakovskyi* (1.7–3% mtDNA divergence), even though they did not have the highest parasitemia values. For the samples with higher parasitemia but higher divergences from *H. tartakovskyi* (up to 6.2% for *Haemoproteus* sample), we still recovered more than 200 loci. We were also able to add 45–60 more loci for these divergent samples by using *H. majoris* as a reference. In aTRAM, protein sequences can be used instead of nucleotide sequences as references for assembling more divergent loci, but in our case, contamination from host bird DNA resulted in some gene assemblies for the bird instead of the parasite. One potential work-around may be to filter reads by GC content prior to assembly, because avian haemosporidian genomes have lower GC content on average than bird hosts (Galen *et al.* 2018a), but this potential approach will require further testing.

One other challenge for haemosporidian research is that mixed infections are extremely common in nature. We could be fairly confident that the sequences from the co-infected *Leucocytozoon* sample did indeed belong to *Leucocytozoon* because we were able to compare them with a pure infection of the same *Haemoproteus* lineage. The two other co-infected samples in our study were not unambiguously sorted, but with reference sequences from one or both of the lineages in a mixed infection, a step could be added to the pipeline to sort them bioinformatically. One other future improvement to the pipeline will be to incorporate protein-guided multiple sequence alignments. We chose to manually check our MAFFT alignments and we removed 25 genes (6%) that were poorly-aligned. Because we are targeting exons, protein-guided DNA

406 alignments may improve results for more divergent sequences and remove the need for extensive
407 manual checking.

408 The total cost for our study including DNA extraction, qPCR, library preparation, capture,
409 and sequencing was approximately $87 USD per sample, or $0.30 per locus based on the average
410 of 295 loci per successful sample. For just library preparation and sequence capture, with eight
411 libraries pooled before capture, we spent ~$35 per sample, or $0.12 per locus. Costs could be
412 reduced further by ordering larger capture kits or combining more samples together on one
413 higher-output sequencing platform.

414

415 *Future directions for avian haemosporidian research*

416 Avian haemosporidian research has grown at a rapid pace since molecular tools have been
417 applied to the field. To date, more than 3,100 parasite mtDNA haplotypes have been discovered
418 and uploaded to the MalAvi database along with their associated locality and host species. Broad
419 syntheses of these data have provided important insights into global distribution patterns and
420 host-parasite associations (Clark *et al.* 2014, 2017; Ellis & Bensch 2018), but the improved
421 resolution of *Haemoproteus* relationships afforded by our genomic sequence capture data has
422 great potential for moving avian haemosporidian research forward. First, the species limits of
423 haemosporidian parasites are difficult to define. Some lineages differing by a single nucleotide in
424 the *cytb* barcode region are considered to be reproductively isolated, biological species (Nilsson
425 *et al.* 2016), while others are considered to represent intraspecific variants (Outlaw & Ricklefs
426 2014; Hellgren *et al.* 2015). Galen *et al.* (2018b) used seven nuclear loci to show that both
427 phenomena occur in *Leucocytozoon*, confirming the poor resolution of mtDNA for inferring
428 species limits. Second, developing large, multi-locus nuclear DNA sequence datasets is needed to

advance the study of haemosporidian evolutionary dynamics. Robust phylogenies will allow for more accurate estimates of biogeographic history (Hellgren *et al.* 2015), trait evolution (Ellis & Bensch 2018), and transitions between host-generalist and host-specialist strategies (Loiseau *et al.* 2012b). In this way, methods for collecting haemosporidian genomic data will facilitate detailed studies of parasite diversification, host breadth, and distributional limits across the globe. Furthermore, high-resolution determination of haemosporidian species limits will be critical to identify and manage the novel host-parasite interactions that are expected to pose a major threat to host-species persistence during climate warming (Garamszegi 2011; Loiseau *et al.* 2012a).

Acknowledgements

This work was supported by NSF DEB-1146491, a New Mexico Ornithological Society research grant, and an NSF Postdoctoral Fellowship (NSF PRFB-1611710) to LNB. We thank Michael Andersen, Jenna McCullough, and the staff of the UNM Center for Advanced Research Computing.

References

Albert TJ, Molla MN, Muzny DM *et al.* (2007) Direct selection of human genomic loci by microarray hybridization. *Nature methods*, **4**, 903–905.

Allen JM, Boyd B, Nguyen NP *et al.* (2017) Phylogenomics from whole genome sequences using aTRAM. *Systematic Biology*, **66**, 786–798.

Allen JM, Huang DI, Cronk QC, Johnson KP (2015) aTRAM - automated target restricted assembly method: a fast method for assembling loci across divergent taxa from next-generation sequencing data. *BMC bioinformatics*, **16**, 98.

Allen JM, LaFrance R, Folk RA, Johnson KP, Guralnick RP (2018) aTRAM 2.0: An Improved, Flexible Locus Assembler for NGS Data: *Https://Doi.Org/10.1177/1176934318774546*, **14**, 117693431877454.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Atkinson CT, LaPointe DA (2009) Introduced Avian Diseases, Climate Change, and the Future of Hawaiian Honeycreepers. *Journal of Avian Medicine and Surgery*, **23**, 53–63.

Atkinson CT, Lease JK, Drake BM, Shema NP (2001) Pathogenicity, serological responses, and diagnosis of experimental and natural malarial infections in native hawaiian thrushes. *The Condor*, **103**, 209–218.

Bensch S, Canbäck B, DeBarry JD *et al.* (2016) The Genome of *Haemoproteus tartakovskyi* and Its Relationship to Human Malaria Parasites. *Genome Biology and Evolution*, **8**, 1361–1373.

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. *Molecular Ecology Resources*, **9**, 1353–1358.

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. *Proceedings of the Royal Society B: Biological Sciences*, **267**, 1583–1589.

Böhme U, Otto TD, Cotton J *et al.* (2018) Complete avian malaria parasite genomes reveal features associated with lineage specific evolution in birds and mammals. *Genome research*, **28**, 547–560.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Borner J, Pick C, Thiede J *et al.* (2016) Phylogeny of haemosporidian blood parasites revealed by a multi-gene approach. *Molecular Phylogenetics and Evolution*, **94**, 221–231.

Buddenhagen C, Lemmon AR, Lemmon EM *et al.* (2016) Anchored Phylogenomics of Angiosperms I: Assessing the Robustness of Phylogenetic Estimates. *bioRxiv*, **doi: https**.

Chifman J, Kubatko L (2014) Quartet inference from SNP data under the coalescent model. *Bioinformatics*, **30**, 3317–3324.

Clark NJ (2018) Phylogenetic uniqueness, not latitude, explains the diversity of avian blood parasite communities worldwide. *Global Ecology and Biogeography*, 1–12.

Clark NJ, Clegg SM, Lima MR (2014) A review of global diversity in avian haemosporidians (Plasmodium and Haemoproteus: Haemosporida): new insights from molecular data. *International Journal for Parasitology*, **44**, 329–338.

Clark NJ, Clegg SM, Sam K *et al.* (2017) Climate, host phylogeny and the connectivity of host communities govern regional parasite assembly. *Diversity and Distributions*, 1–11.

Ellis VA, Bensch S (2018) Host specificity of avian haemosporidian parasites is unrelated among sister lineages but shows phylogenetic signal across larger clades. *International Journal for Parasitology*.

Faircloth BC (2013) illumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming.

Faircloth BC (2015) PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*, **32**, 786–788.

Faircloth BC (2017) Identifying conserved genomic elements and designing universal bait sets to enrich them. *Methods in Ecology and Evolution*, **doi: 10.11**.

Faircloth BC, Glenn TC (2012) Not all sequence tags are created equal: Designing and validating sequence identification tags robust to indels. *PLoS ONE*, **7**, e42543.

Faircloth BC, McCormack JE, Crawford NG *et al.* (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic biology*, **61**, 717–726.

Falk BG, Glor RE, Perkins SL (2015) Clonal reproduction shapes evolution in the lizard malaria parasite Plasmodium floridense. *Evolution*, **69**, 1584–1596.

Fallon SM, Bermingham E, Ricklefs RE (2005) Host Specialization and Geographic Localization of Avian Malaria Parasites: A Regional Analysis in the Lesser Antilles. *The American Naturalist*, **165**, 466–480.

Fallon SM, Ricklefs RE, Latta SC, Bermingham E (2004) Temporal stability of insular avian malarial parasite communities. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 493–500.

Fallon SM, Ricklefs RE, Swanson BL, Bermingham E (2003) Detecting Avian Malaria: An Improved Polymerase Chain Reaction Diagnostic. *Journal of Parasitology*, **89**, 1044–1047.

Fecchio A, Pinheiro R, Felix G *et al.* (2017) Host community similarity and geography shape the diversity and distribution of haemosporidian parasites in Amazonian birds. *Ecography*.

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. *Royal Society Open Science*, **5**, 171780.

Galen SC, Nunes R, Sweet PR, Perkins SL (2018b) Integrating coalescent species delimitation with analysis of host specificity reveals extensive cryptic diversity despite minimal mitochondrial divergence in the malaria parasite genus Leucocytozoon. , 1–15.

Garamszegi LZ (2011) Climate change increases the risk of malaria in birds. *Global Change Biology*, **17**, 1751–1759.

Glenn TC, Nilsen RA, Kieran TJ *et al.* (2016) Adapterama I: Universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext). *bioRxiv*.

Gnirke A, Melnikov A, Maguire J *et al.* (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotech*, **27**, 182–189.

Grabherr MG, Haas BJ, Yassour M *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**, 644.

Hamilton CA, Lemmon AR, Lemmon EM, Bond JE (2016) Expanding anchored hybrid enrichment to resolve both deep and shallow relationships within the spider tree of life. *BMC Evolutionary Biology*, **16**, 212.

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.

534 Hellgren O, Waldenström J, Bensch S (2004) A new PCR assay for simultaneous studies of
535 Leucocytozoon, Plasmodium, and Haemoproteus from avian blood. *Journal of Parasitology*,
536 **90**, 797–802.
537 Huang X, Hansson R, Palinauskas V *et al.* (2018) The success of sequence capture in relation to
538 phylogenetic distance from a reference genome: a case study of avian haemosporidian
539 parasites. *International Journal for Parasitology*, **48**, 893–900.
540 Ishtiaq F, Rao M, Huang X, Bensch S (2017) Estimating prevalence of avian haemosporidians in
541 natural populations: A comparative study on screening protocols. *Parasites and Vectors*, **10**,
542 1–10.
543 Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7:
544 Improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772–780.
545 Lanfear R, Calcott B, Kainer D, Mayer C, Stamatakis A (2014) Selecting optimal partitioning
546 schemes for phylogenomic datasets. *BMC evolutionary biology*, **14**, 82.
547 Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B (2017) Partitionfinder 2: New
548 methods for selecting partitioned models of evolution for molecular and morphological
549 phylogenetic analyses. *Molecular Biology and Evolution*, **34**, 772–773.
550 Lemmon AR, Emme SA, Lemmon EM (2012) Anchored hybrid enrichment for massively high-
551 throughput phylogenomics. *Systematic Biology*, **61**, 727–744.
552 Loiseau C, Harrigan RJ, Cornel AJ *et al.* (2012a) First Evidence and Predictions of Plasmodium
553 Transmission in Alaskan Bird Populations. *PLoS ONE*, **7**, 7–11.
554 Loiseau C, Harrigan RJ, Robert A *et al.* (2012b) Host and habitat specialization of avian malaria
555 in Africa. *Molecular Ecology*, **21**, 431–441.
556 Lutz HL, Hochachka WM, Engel JI *et al.* (2015) Parasite prevalence corresponds to host life
557 history in a diverse assemblage of afrotropical birds and haemosporidian parasites. *PLoS*
558 *ONE*, **10**.
559 Lutz HL, Marra NJ, Grewe F *et al.* (2016) Laser capture microdissection microscopy and genome
560 sequencing of the avian malaria parasite, Plasmodium relictum. *Parasitology Research*, **115**,
561 4503–4510.
562 Marroquin-Flores RA, Williamson JL, Chavez AN *et al.* (2017) Diversity, abundance, and host
563 relationships in the avian malaria community of New Mexico pine forests. *PeerJ*, **5**, e3700.
564 Martinsen ES, Perkins SL, Schall JJ (2008) A three-genome phylogeny of malaria parasites
565 (Plasmodium and closely related genera): Evolution of life-history traits and host switches.
566 *Molecular Phylogenetics and Evolution*, **47**, 261–273.
567 Nilsson E, Taubert H, Hellgren O *et al.* (2016) Multiple cryptic species of sympatric generalists
568 within the avian blood parasite Haemoproteus majoris. *Journal of Evolutionary Biology*, **29**,
569 1812–1826.
570 Olsson-Pons S, Clark NJ, Ishtiaq F, Clegg SM (2015) Differences in host species relationships
571 and biogeographic influences produce contrasting patterns of prevalence, community
572 composition and genetic structure in two genera of avian malaria parasites in southern
573 Melanesia. *Journal of Animal Ecology*, **84**, 985–998.
574 Outlaw DC, Ricklefs RE (2014) Species limits in avian malaria parasites (Haemosporida): how to
575 move forward in the molecular era. *Parasitology*, **141**, 1223–32.
576 Pacheco MA, Matta NE, Valkiünas G *et al.* (2018) Mode and rate of evolution of haemosporidian
577 mitochondrial genomes: Timing the radiation of avian parasites. *Molecular Biology and*
578 *Evolution*, **35**, 383–403.

Palinauskas V, Dolnik O V, Valkiūnas G, Bensch S (2010) Laser microdissection microscopy and single cell PCR of avian hemosporidians. *The Journal of parasitology*, **96**, 420–424.

Palinauskas V, Križanauskienė A, Iezhova TA *et al.* (2013) A new method for isolation of purified genomic DNA from haemosporidian parasites inhabiting nucleated red blood cells. *Experimental Parasitology*, **133**, 275–280.

Perkins SL (2014) Malaria's Many Mates: Past, Present, and Future of the Systematics of the Order Haemosporida. *The journal of parasitology*, **100**, 11–25.

Quattrini AM, Faircloth BC, Dueñas LF *et al.* (2018) Universal target-enrichment baits for anthozoan (Cnidaria) phylogenomics: New approaches to long-standing problems. *Molecular Ecology Resources*, **18**, 281–295.

R Core Team (2016) R: A language and environment for statistical computing.

Ricklefs RE, Outlaw DC, Svensson-Coelho M *et al.* (2014) Species formation by host shifting in avian malaria parasites. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 14816–21.

van Riper C, van Riper SG, Goff ML, Laird M (1986) The Epizootiology and Ecological Significance of Malaria in Hawaiian Land Birds. *Ecological Monographs*, **56**, 327–344.

RStudio Team (2016) RStudio: Integrated Development for R.

Scheuerlein A, Ricklefs RE (2004) Prevalence of blood parasites in European passeriform birds. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 1363–1370.

Slater GSC, Birney E (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, **6**, 1–11.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Swofford DL (2002) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods).

Valkiunas G (2005) *Avian malaria parasites and other haemosporidia*. CRC Press, Boca Raton, Florida, USA.

Valkiŭnas G, Iezhova T a, Križanauskienė A *et al.* (2008) A comparative analysis of microscopy and PCR-based detection methods for blood parasites. *The Journal of Parasitology*, **94**, 1395–401.

Videvall E, Cornwallis CK, Ahrén D *et al.* (2017) The transcriptome of the avian malaria parasite Plasmodium ashfordi displays host-specific gene expression. *Molecular Ecology*, **26**, 2939–2958.

Waldenström J, Bensch S, Hasselquist D, Östman Ö (2004) A New Nested Polymerase Chain Reaction Method Very Efficient in Detecting Plasmodium and Haemoproteus Infections from Avian Blood. *Journal of Parasitology*, **90**, 191–194.

Zehtindjiev P, Ilieva M, Westerdahl H *et al.* (2008) Dynamics of parasitemia of malaria parasites in a naturally and experimentally infected migratory songbird, the great reed warbler Acrocephalus arundinaceus. *Experimental Parasitology*, **119**, 99–110.

Data Accessibility

Specimen information is available from the Arctos database (arctosdb.org). Parasite sequences are available on MalAvi and GenBank (Accession Numbers in Table S2). Protocols and probe sequences are included as supporting information. Scripts, sequence data, and alignments will be made available on Dryad.

Author Contributions

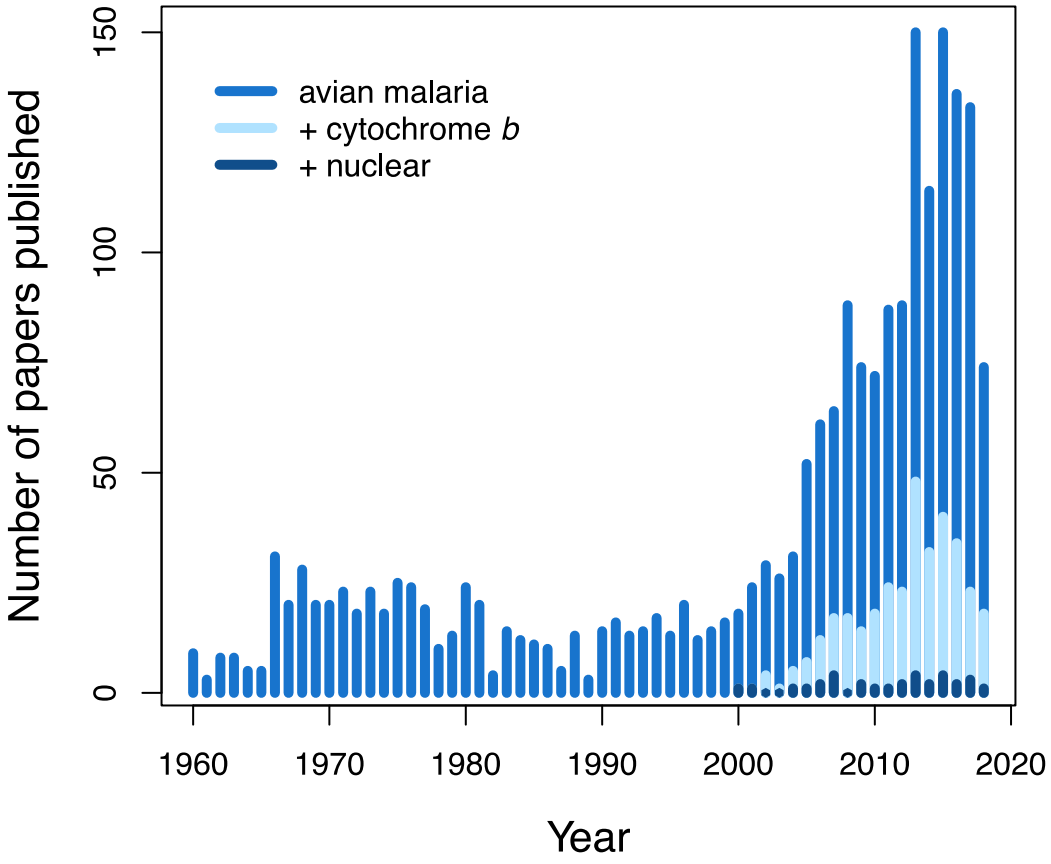
LNB designed the study, collected the data, and led the writing; LNB and JMA analyzed the data; XH, SB, and CCW contributed genomic data and resources; all authors edited the manuscript.

631 **Tables**
632
633 Table 1 Sampling information for haemosporidian parasite lineages included in the study. All are
634 *Haemoproteus* except one *Leucocytozoon* (*Leuc*). Additional information for each specimen is
635 provided in Table S2.
636

Parasite lineage	Study region	N samples	N host species	% mtDNA divergence from *H. tartakovskyi*	Nearest reference	% mtDNA divergence from nearest reference
G001	Peru	18	18	4.1	*H. majoris*	3.77
G002	Peru	5	3	5.57	*H. tartakovskyi*	5.57
G003	Peru	2	2	3.64	*H. tartakovskyi*	3.64
G004	Peru	2	2	6.2	*H. majoris*	5.23
G005	Peru	2	2	2.57	*H. tartakovskyi*	2.57
G006	Peru	2	2	3.43	*H. tartakovskyi*	3.43
G007	Peru	6	5	3.21	*H. tartakovskyi*	3.21
G009	Peru	4	4	3.43	*H. tartakovskyi*	3.43
G011	Peru	3	3	3.0	*H. tartakovskyi*	3.0
G403 (*Leuc*)	Peru	1	1	18.0	*H. nucleocondensus*	15.9
CHOGRA01	New Mexico	1	1	1.71	*H. tartakovskyi*	1.71
SPIPAS01	New Mexico	2	2	1.71	*H. tartakovskyi*	1.71
SPISAL01	New Mexico	1	1	6.0	*H. nucleocondensus*	4.81
TROAED12	New Mexico	1	1	5.9	*H. majoris*	3.77
VIGIL05	New Mexico	1	1	4.5	*H. majoris*	3.98

637
638
639
640
641

642 **Figures**
643
644 Fig. 1 Papers published on 'avian malaria' by year since 1960 (n = 2,066) based on a Web of
645 Science search (Clarivate Analytics, August 2018). Since 2000, many more 'avian malaria'
646 papers include 'mitochondrial' (not shown, n = 218) or the barcode gene 'cytochrome *b*' (n =
647 339) than include 'nuclear' (n = 32).
648
649
650



651
652

Fig. 2 Number of parasite loci sequenced out of 498 total. (a) Locus recovery increased with parasitemia, shown as % infected cells estimated by microscopic examination of the qPCR standard. Colors depict % divergence from the nearest reference. (b) Locus recovery decreased with increasing % divergence from references, but with sufficient parasitemia, divergent samples were successful. Colors depict parasitemia as the qPCR CT value. The *Leucocytozoon* sample (15.9% divergent) is not shown in (b) to improve visualization. In both plots, shapes depict the number of samples in a capture pool, which did not influence locus recovery.

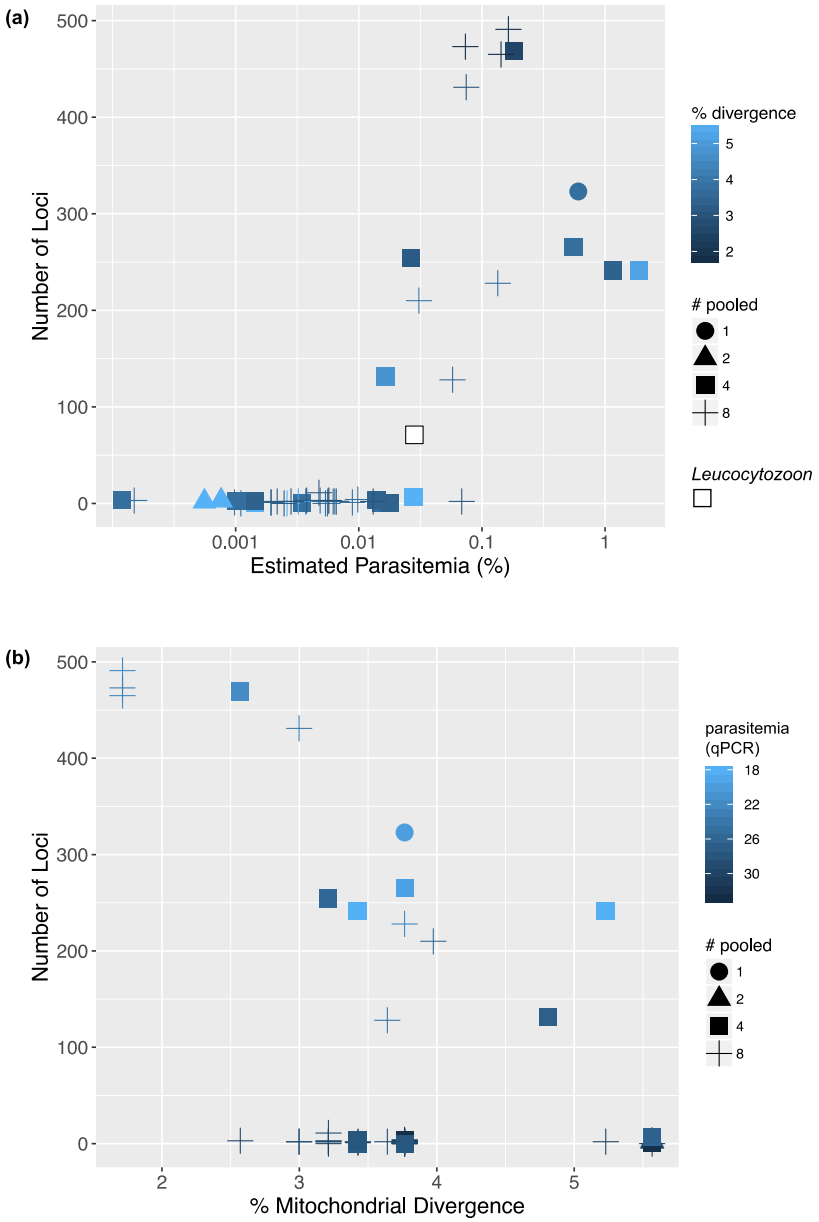


Fig. 3 Phylogenies estimated with nuclear (a, b) and mitochondrial (c, d) datasets. Blue tip labels indicate reference samples. RAxML best trees are shown with branch lengths in substitutions per site and RAxML bootstrap support above branches. (a, b) SVDquartets support values are shown in bold below branches. Red tip labels indicate samples with mixed infections. (c, d) Branches with support values <50 were collapsed. Branch lengths for G403 (*Leucocytozoon*) were reduced to improve visualization.