Received: 20 September 2018 | Revised: 26 November 2018 | Accepted: 27 November 2018

DOI: 10.1111/1755-0998.12977

RESOURCE ARTICLE

WILEY MOLECULAR ECOLOGY RESOURCES

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

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

[3]Department of Biology, Molecular Ecology and Evolution Laboratory, Lund University, Lund, Sweden

Correspondence
Lisa N. Barrow, Museum of Southwestern Biology and Department of Biology, MSC03 2020, 1 University of New Mexico, Albuquerque, NM.
Email: lnbarrow@unm.edu

Funding information
National Science Foundation, Grant/Award Number: DEB-1146491 and PRFB-1611710; New Mexico Ornithological Society

Abstract

Avian malaria and related haemosporidians (*Plasmodium*, *[Para]Haemoproteus* and *Leucocytoozoon*) represent an exciting multihost, multiparasite system in ecology and evolution. Global research in this field accelerated after the publication in 2000 of PCR protocols to sequence a haemosporidian mitochondrial (mtDNA) barcode and 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 ~0.02% (2 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 sequences and describe our bioinformatics pipeline using ATRAM 2.0, PHYLUCE and custom Perl/Python scripts. This approach can be applied to thousands of avian samples that have already been found to have haemosporidian infections of at least moderate intensity, greatly improving our understanding of parasite speciation, biogeography and evolutionary dynamics.

KEYWORDS

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

1 | INTRODUCTION

Multihost, multiparasite 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 (Fallon, Bermingham, & Ricklefs, 2005; Fallon, Ricklefs, Latta, & Bermingham, 2004; Fecchio et al., 2017; Olsson-Pons, Clark, Ishtiaq, & Clegg, 2015), to host-switching and diversification across long evolutionary timescales (Galen, Borner, et al., 2018; Martinsen, Perkins, & Schall, 2008; Pacheco et al., 2018; Ricklefs et al., 2014). 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 (Bensch, Hellgren, & Pérez-Tris, 2009; Clark, Clegg, & Lima, 2014; Lutz et al., 2015; Scheuerlein & Ricklefs, 2004). 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 (Atkinson & LaPointe, 2009; van Riper, van Riper, Goff, & Laird, 1986).

The detection and description of avian blood parasites greatly accelerated with the application of molecular methods. While microscopy of thin blood smears remains essential for morphological verification and detailed species descriptions (Valkiūnas, 2005; Valkiūnas et al.., 2008), the field benefited substantially from the development of PCR primers for avian haemosporidians (Bensch et al., 2000; Figure 1). Subsequent nested PCR protocols based on these primers (Hellgren, Waldenström, & Bensch, 2004; Waldenström, Bensch, Hasselquist, & Östman, 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, 2018; Clark et al., 2014, 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 multilocus nuclear data of avian haemosporidian parasites (Figure 1). Studies demonstrate that the *cytb* barcode provides limited resolution; a single *cytb* haplotype can include multiple cryptic species (Falk, Glor, & Perkins, 2015; Galen, Nunes, Sweet, & Perkins, 2018), and phylogenies estimated from multiple nuclear loci substantially improve inferences of evolutionary relationships (Borner et al., 2016; Galen, Borner, et al., 2018). Several challenges, however, have previously prevented any large-scale efforts to obtain genomic data

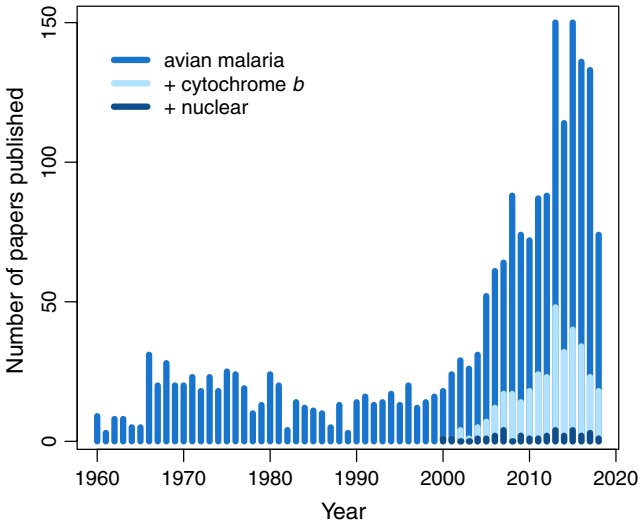

FIGURE 1 Papers published on "avian malaria" by year since 1960 (n = 2,066) based on a Web of Science search (Clarivate Analytics, August 2018). Since 2000, many more "avian malaria" papers include "mitochondrial" (not shown, n = 218) or the barcode gene "cytochrome *b*" (n = 339) than include "nuclear" (n = 32) [Colour figure can be viewed at wileyonlinelibrary.com]

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, Dolnik, Valkiūnas, & Bensch, 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 (a) work on any sample that contained preserved DNA, including the tens of thousands of frozen blood and tissue samples that have been screened for avian haemosporidians, (b) be broadly applicable across a diverse group of haemosporidian parasites and (c) 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 (Buddenhagen et al., 2016; Faircloth, 2017; Faircloth et al., 2012; Hamilton, Lemmon, Lemmon, & Bond, 2016; Lemmon, Emme, & Lemmon, 2012; Quattrini et al., 2018). These techniques allow for the enrichment of genomic regions of interest by hybridizing oligonucleotide probes to genomic samples and removing nontarget 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 (Huang et al., 2018; Lemmon et al., 2012). 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 avian 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 <0.1% (Atkinson, Lease, Drake, & Shema, 2001; Ishtiaq, Rao, Huang, & Bensch, 2017; Zehtindjiev et al., 2008). 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 (a) describe the relationship between parasitemia levels and sequence capture success, (b) test how sequence capture success is affected by per cent divergence from the reference sequences used for probe design and (c) compare strategies for pooling samples before capture to increase cost-effectiveness. To facilitate the 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, Huang, Cronk, & Johnson, 2015; Allen, LaFrance, Folk, Johnson, & Guralnick, 2018) for locus assembly and PHYLUCE (Faircloth, 2015) and custom PERL and Python scripts for downstream processing.

2 | MATERIALS AND METHODS

2.1 | 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 (Supporting information Table S1) 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 cut-offs 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; Supporting information 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.

2.2 | 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*. 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. All collection and export/import were conducted under permits issued by the management authorities of Peru (076-2006-INRENA-IFFS-DCB, 004-2007-INRENAIFFS-DCB, 135-2009-AG-DGFFS-DGEFFS, 0377-2010-AGDGFFS-DGEFFS, 008843-AG-INRENA, 011283-AG-INRENA, 011829-AG-INRENA, 0000082-SERFOR, CITES Certificate Nos. 9213, 10644, 11387, 000487, and 14PE000016/SP) and the United States (USFWS-MB094297-0, NMDGF-3217, USDA-APHIS-102547), respectively. 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; Supporting information 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, Ricklefs, Swanson, & Bermingham, 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 s and 57°C for 1 min, followed by a melt curve analysis (47°C to

TABLE 1 Sampling information for haemosporidian parasite lineages included in the study

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

Note. All are *Haemoproteus* except one *Leucocytozoon* (*Leuc*). Additional information for each specimen is provided in Supporting information Table S2.

95°C at 0.5°C and 5 s 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.

2.3 | 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 centred 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 postcapture 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 s, 60°C for 30 s and 72°C for 60 s, followed by a final extension at 72°C for 5 min and held at 4°C. Postamplification, 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.

2.4 | 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, Lohse, & Usadel, 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, Gish, Miller, Myers, & Lipman, 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 to keep only the contigs from the last ATRAM iteration (available at https://github.com/juliema/phylogenomics_scripts; "getlastiteration.pl"), 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; https://github.com/juliema/exon_stitching). 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 locus was not the target locus. For this search, we created a local BLAST database from all 6,436 *H. tartakovskyi* genes (downloaded from http://mbio-serv2.mbioekol.lu.se/Malavi/Downloads). Fewer than 0.2% (9 of 4,507) individual-locus combinations were mismatched and therefore removed. Prior to multiple sequence alignment, we added sequences for the three reference species and used custom Python scripts (available on Dryad: https://doi.org/10.5061/dryad.8030nq9) to reformat the ATRAM sequences for the PHYLUCE pipeline (Faircloth, 2015).

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

2.5 | Sequence coverage estimation

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

2.6 | Detection of mixed infections

For each successfully captured individual, we mapped the cleaned, paired reads to two haemosporidian mtDNA genes, *COI* and *cytb*, to check for possible coinfections. We used the *H. tartakovskyi* sequences as references to map the paired reads in Geneious and visually inspected assemblies. 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.

2.7 | 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 avian haemosporidian relationships, we estimated a phylogeny for the samples with sufficient data. We used PARTITIONFINDER 2 (Lanfear, Frandsen, Wright, Senfeld, & Calcott, 2017) to select appropriate models of nucleotide evolution and partitioning schemes for each data set. Given the number of loci, we used the rcluster algorithm with RAXML (Lanfear, Calcott, Kainer, Mayer, & Stamatakis, 2014; Stamatakis, 2014), linked branch lengths and chose AICc for model selection. Phylogenies were estimated for each nuclear data set (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 1,000 bootstrap replicates. We also estimated species trees from the nuclear data sets 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 1,000 replicates and partitioned loci.

3 | RESULTS

3.1 | 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 (*SD* ±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; Supporting information Figure S1).

3.2 | Effects of parasitemia, divergence and pooling on success

Parasitemia was positively correlated with capture success and locus recovery ($t = 6.15$, $p < 0.0001$; Figure 2a). Samples with parasitemia values >0.07% (~7 out of 10,000 infected cells) were all successful, and samples with as low as 0.017% (~2 of 10,000 cells; qPCR CT value 26.6) also resulted in >100 loci (Figure 2a; Supporting information Table S2). Divergence from the nearest reference was negatively correlated with locus recovery ($t = -3.91$, $p = 0.0003$; Figure 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.

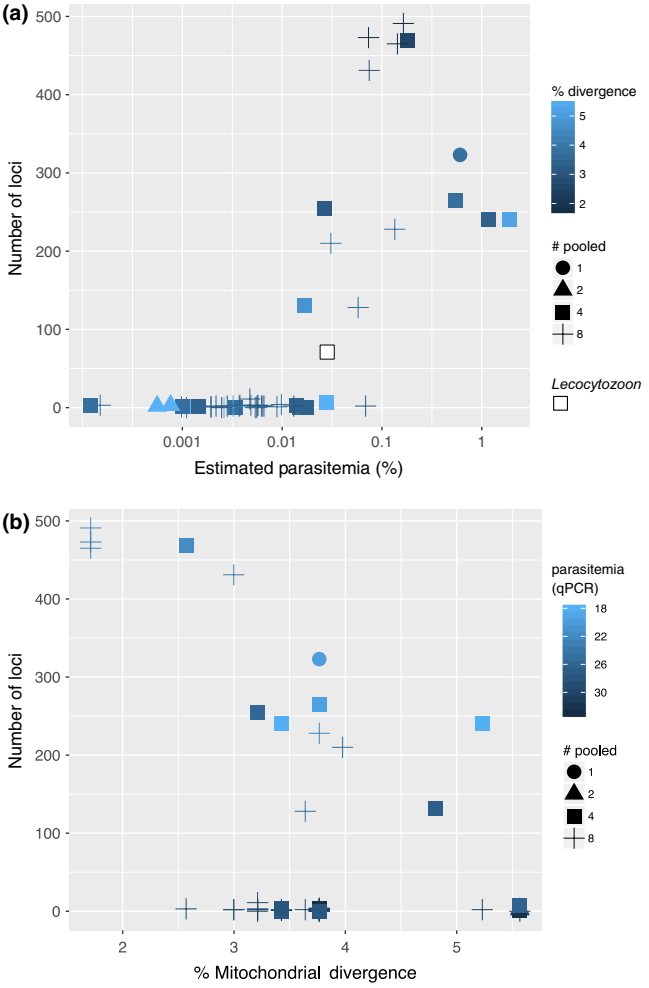

FIGURE 2 Number of parasite loci sequenced out of 498 total. (a) Locus recovery increased with parasitemia, shown as % infected cells estimated by microscopic examination of the qPCR standard. Colours depict % divergence from the nearest reference. (b) Locus recovery decreased with increasing % divergence from references, but with sufficient parasitemia, divergent samples were successful. Colours 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 [Colour figure can be viewed at wileyonlinelibrary.com]

3.3 | 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.

3.4 | Detection of mixed infections

We identified three samples coinfected with multiple lineages, two of which had not been previously detected by nested PCR and Sanger sequencing. Sample NK168883 was coinfected with the *Haemoproteus* lineage G009 and a novel *Leucocytozoon* lineage (assigned name G403). Within the MalAvi barcode region, the frequency of the *Leucocytozoon* lineage dominated the reads and ranged from 94.8% to 98.0%. The concatenated sequence for this sample had >15% sequence divergence from the pure G009 *Haemoproteus* sample (NK168881), indicating that the genes assembled for that sample belonged to *Leucocytozoon*. Sample NK275890 was coinfected with two *Haemoproteus* lineages, SPIPAS01 and SIAMEX01

(the read frequency of SIAMEX01 was 11.0%–14.2%; mtDNA divergence between the two haplotypes of 6.1%). Sample NK276102 was also coinfected with two *Haemoproteus* lineages, VIGIL05 and VIGIL07 (the read frequency of VIGIL07 was 30.3%–40.5%; mtDNA divergence 3.6%). Phylogenetic analyses were repeated without the two *Haemoproteus* coinfected samples because the loci extracted by ATRAM may have represented a composite of the two haplotypes that could not be definitively sorted.

3.5 | Phylogenetic resolution

The nuclear data sets substantially improved phylogenetic resolution of *Haemoproteus* parasites (Figure 3). Most relationships were consistent among data sets with different levels of completeness. The 50% complete matrix included 377 loci and 287,164 bp (Figure 3a), the 70% matrix included 206 loci and 189,883 bp, and the 90% matrix included 59 loci and 70,611 bp (Figure 3b). The topologies resulting from RAXML and SVDquartets were identical for the 50% and 70% data sets. The position of *H. tartakovskyi* differed for the 90% matrix RAXML analysis (Figure 3b), while the SVDquartets topology was consistent with the other data sets. Phylogenetic analyses excluding the *Haemoproteus* mixed infections resulted in similar topologies, except for the uncertain position of TROAED12 (Supporting information Figure S2). The majority of nodes had high support for the nuclear data sets; 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 data set and the *cytb* barcode produced poorly resolved phylogenies for the lineages in our study, inferring very few relationships with any certainty (Figure 3c,d). Only 6 (38%) and 3 (19%) of the nodes had bootstrap values ≥95, respectively.

4 | DISCUSSION

4.1 | 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% or higher parasitemia were always successful, and samples with as low as ~0.02% 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 moderate portion of these samples (29% in our study) will be suitable for sequence capture of avian 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.

4.2 | 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* are 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 (Böhme et al., 2018; Lutz et al., 2016; Videvall et al., 2017).

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 workaround 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, Borner, et al., 2018), 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 coinfected *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 coinfected samples in our study were not

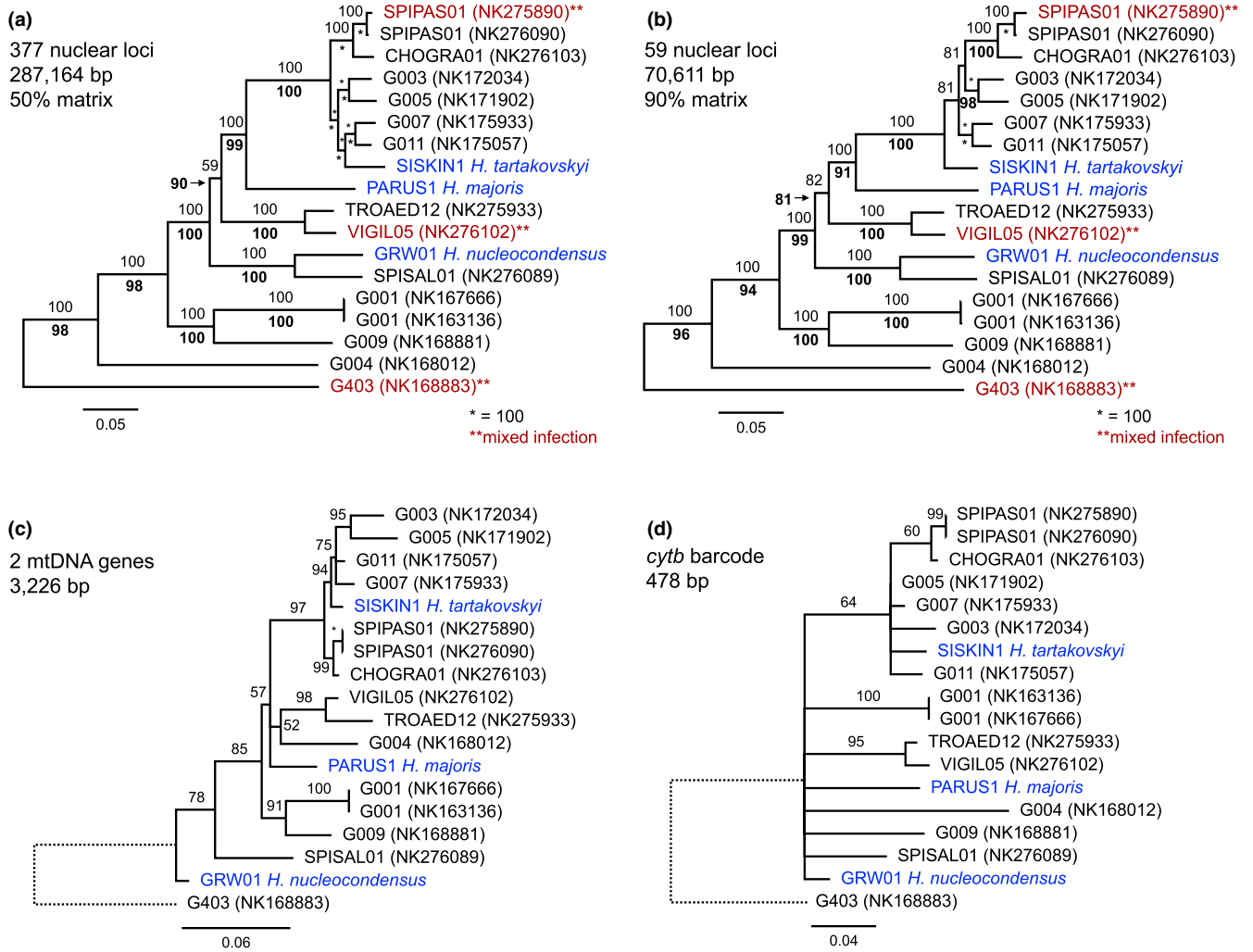

FIGURE 3 Phylogenies estimated with nuclear (a, b) and mitochondrial (c, d) data sets. Blue tip labels indicate reference samples. RAXML best trees are shown with branch lengths in substitutions per site and RAXML bootstrap support above branches. (a, b) SVDquartets support values are shown in bold below 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 [Colour figure can be viewed at wileyonlinelibrary.com]

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

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

4.3 | Future directions for avian haemosporidian research

Avian haemosporidian research has grown at a rapid pace since molecular tools have been applied to the field. To date, more than 3,100 parasite mtDNA haplotypes have been discovered and uploaded to the MalAvi database along with their associated locality and host species. Broad syntheses of these data have provided important insights into global distribution patterns and host–parasite associations (Clark et al., 2014, 2018; Ellis & Bensch, 2018), but the improved resolution of *Haemoproteus* relationships afforded by our genomic sequence capture data has great potential for moving avian haemosporidian research forward. First, the species limits of haemosporidian parasites are difficult to define. Some lineages differing by a single nucleotide in the *cytb* barcode region are considered to be reproductively isolated, biological species (Nilsson et al., 2016), while others are considered to represent intraspecific

variants (Hellgren et al., 2015; Outlaw & Ricklefs, 2014). Galen, Nunes, et al. (2018) used seven nuclear loci to show that both phenomena occur in *Leucocytozoon*, confirming the poor resolution of mtDNA for inferring species limits. Second, developing large, multilocus nuclear DNA sequence data sets 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, Harrigan, Robert, et al., 2012). 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, Harrigan, Cornel, et al., 2012).

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, Sara Brant, Mariel Campbell, Rosario Marroquin-Flores, George Rosenberg, Jenna McCullough and the staff of the UNM Center for Advanced Research Computing.

AUTHOR CONTRIBUTIONS

The study was designed, the data were collected, and the manuscript was written by L.N.B.; the data were analysed by L.N.B. and J.M.A.; genomic data and resources were contributed by X.H., S.B. and C.C.W.; all authors edited the manuscript.

DATA ACCESSIBILITY

Specimen information is available from the Arctos database (arctosdb.org). Parasite sequences are available on MalAvi and GenBank (Accession Numbers in Supporting information Table S2). Protocols and probe sequences are included as supporting information. Scripts, sequence data and alignments are available on Dryad https://doi.org/10.5061/dryad.8030nq9.

ORCID

Lisa N. Barrow https://orcid.org/0000-0001-7081-2432
Julie M. Allen https://orcid.org/0000-0002-7625-3050
Xi Huang https://orcid.org/0000-0001-6515-261X
Staffan Bensch https://orcid.org/0000-0002-0082-0899
Christopher C. Witt https://orcid.org/0000-0003-2781-1543

REFERENCES

Albert, T. J., Molla, M. N., Muzny, D. M., Nazareth, L., Wheeler, D., Song, X., … Gibbs, R. A. (2007). Direct selection of human genomic loci by microarray hybridization. *Nature Methods*, *4*(11), 903–905. https://doi.org/10.1038/nmeth1111

Allen, J. M., Huang, D. I., Cronk, Q. C., & Johnson, K. P. (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. https://doi.org/10.1186/s12859-015-0515-2

Allen, J. M., LaFrance, R., Folk, R. A., Johnson, K. P., & Guralnick, R. P. (2018). ATRAM 2.0: An improved, flexible locus assembler for NGS data. *Evolutionary Bioinformatics*, *14*, 1–4. https://doi.org/10.1177/1176934318774546

Allen, J. M., Boyd, B., Nguyen, N. P., Vachaspati, P., Warnow, T., Huang, D. I., … Johnson, K. P. (2017). Phylogenomics from whole genome sequences using aTRAM. *Systematic Biology*, *66*(5), 786–798. https://doi.org/10.1093/sysbio/syw105

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

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

Atkinson, C. T., Lease, J. K., Drake, B. M., & Shema, N. P. (2001). Pathogenicity, serological responses, and diagnosis of experimental and natural malarial infections in native Hawaiian thrushes. *The Condor*, *103*(2), 209–218. https://doi.org/10.1650/0010-5422(2001)103[0209:psrado]2.0.co;2

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*(5), 1353–1358. https://doi.org/10.1111/j.1755-0998.2009.02692.x

Bensch, S., Stjernman, M., Hasselquist, D., Orjan, O., Hannson, B., Westerdahl, H., & Pinheiro, R. T. (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*(1452), 1583–1589. https://doi.org/10.1098/rspb.2000.1181

Bensch, S., Canbäck, B., DeBarry, J. D., Johansson, T., Hellgren, O., Kissinger, J. C., … Valkiūnas, G. (2016). The genome of *Haemoproteus tartakovskyi* and its relationship to human malaria parasites. *Genome Biology and Evolution*, *8*(5), 1361–1373. https://doi.org/10.1093/gbe/evw081

Böhme, U., Otto, T. D., Cotton, J., Steinbiss, S., Sanders, M., Oyola, S. O., … Berriman, M. (2018). Complete avian malaria parasite genomes reveal features associated with lineage specific evolution in birds and mammals. *Genome Research*, *28*, 547–560. https://doi.org/10.1101/gr.218123.116

Bolger, A. M., Lohse, M., & Usadel, B. (2014). TRIMMOMATIC: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Borner, J., Pick, C., Thiede, J., Kolawole, O. M., Kingsley, M. T., Schulze, J., … Burmester, T. (2016). Phylogeny of haemosporidian blood parasites revealed by a multi-gene approach. *Molecular Phylogenetics and Evolution*, *94*, 221–231. https://doi.org/10.1016/j.ympev.2015.09.003

Buddenhagen, C., Lemmon, A. R., Lemmon, E. M., Bruhl, J., Cappa, J., Clement, W. L., … Mast, A. (2016). Anchored phylogenomics of angiosperms I: Assessing the robustness of phylogenetic estimates. *BioRxiv*, https://doi.org/10.1101/086298

Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics*, *30*(23), 3317–3324. https://doi.org/10.1093/bioinformatics/btu530

Clark, N. J. (2018). Phylogenetic uniqueness, not latitude, explains the diversity of avian blood parasite communities worldwide. *Global Ecology and Biogeography*, *27*(6), 744–755. https://doi.org/10.1111/geb.12741

Clark, N. J., Clegg, S. M., & Lima, M. R. (2014). A review of global diversity in avian haemosporidians (Plasmodium and Haemoproteus: Haemosporida): New insights from molecular data. *International Journal for Parasitology*, *44*(5), 329–338. https://doi.org/10.1016/j.ijpara.2014.01.004

Clark, N. J., Clegg, S. M., Sam, K., Goulding, W., Koane, B., & Wells, K. (2018). Climate, host phylogeny and the connectivity of host communities govern regional parasite assembly. *Diversity and Distributions*, *24*(1), 13–23. https://doi.org/10.1111/ddi.12661

Ellis, V. A., & 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*, *48*(12), 897–902. https://doi.org/10.1016/j.ijpara.2018.05.005.

Faircloth, B. C. (2013). ILLUMIPROCESSOR: a trimmomatic wrapper for parallel adapter and quality trimming. https://doi.org/10.6079/J9ILL

Faircloth, B. C. (2015). PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*, *32*(5), 786–788. https://doi.org/10.1093/bioinformatics/btv646.

Faircloth, B. C. (2017). Identifying conserved genomic elements and designing universal bait sets to enrich them. *Methods in Ecology and Evolution*, *8*(9), 1103–1112. https://doi.org/10.1111/2041-210X.12754.

Faircloth, B. C., & Glenn, T. C. (2012). Not all sequence tags are created equal: Designing and validating sequence identification tags robust to indels. *PLoS ONE*, *7*(8), e42543. https://doi.org/10.1371/journal.pone.0042543

Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, *61*(5), 717–726. https://doi.org/10.1093/sysbio/sys004

Falk, B. G., Glor, R. E., & Perkins, S. L. (2015). Clonal reproduction shapes evolution in the lizard malaria parasite *Plasmodium floridense*. *Evolution*, *69*(6), 1584–1596. https://doi.org/10.1111/evo.12683

Fallon, S. M., Bermingham, E., & Ricklefs, R. E. (2005). Host specialization and geographic localization of avian malaria parasites: A regional analysis in the Lesser Antilles. *The American Naturalist*, *165*(4), 466–480. https://doi.org/10.1086/428430

Fallon, S. M., Ricklefs, R. E., Latta, S. C., & Bermingham, E. (2004). Temporal stability of insular avian malarial parasite communities. *Proceedings of the Royal Society B: Biological Sciences*, *271*(1538), 493–500. https://doi.org/10.1098/rspb.2003.2621

Fallon, S. M., Ricklefs, R. E., Swanson, B. L., & Bermingham, E. (2003). Detecting avian malaria: An improved polymerase chain reaction diagnostic. *Journal of Parasitology*, *89*(5), 1044–1047. https://doi.org/10.1645/GE-3157

Fecchio, A., Pinheiro, R., Felix, G., Faria, I. P., Pinho, J. B., Lacorte, G. A., … Weckstein, J. D. (2017). Host community similarity and geography shape the diversity and distribution of haemosporidian parasites in Amazonian birds. *Ecography*, *41*(3), 505–515. https://doi.org/10.1111/ecog.03058.

Galen, S. C., Borner, J., Martinsen, E. S., Schaer, J., Austin, C. C., West, C. J., & Perkins, S. L. (2018). The polyphyly of *Plasmodium*: Comprehensive phylogenetic analyses of the malaria parasites (order Haemosporida) reveal widespread taxonomic conflict. *Royal Society Open Science*, *5*(5), 171780. https://doi.org/10.1098/rsos.171780

Galen, S. C., Nunes, R., Sweet, P. R., & Perkins, S. L. (2018). Integrating coalescent species delimitation with analysis of host specificity reveals extensive cryptic diversity despite minimal mitochondrial divergence in the malaria parasite genus *Leucocytozoon*. *BMC Evolutionary Biology*, *18*, 128. https://doi.org/10.1186/s12862-018-1242-x

Garamszegi, L. Z. (2011). Climate change increases the risk of malaria in birds. *Global Change Biology*, *17*(5), 1751–1759. https://doi.org/10.1111/j.1365-2486.2010.02346.x

Glenn, T. C., Nilsen, R. A., Kieran, T. J., FingerJr, J. W., Pierson, T. W., Bentley, K. E., … Faircloth, B. C. (2016). Adapterama I: Universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext). *BioRxiv*, https://doi.org/10.1101/049114

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., … Nusbaum, C. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology*, *27*(2), 182–189. https://doi.org/10.1038/nbt.1523

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*, 644. https://doi.org/10.1038/nbt.1883

Hamilton, C. A., Lemmon, A. R., Lemmon, E. M., & Bond, J. E. (2016). Expanding anchored hybrid enrichment to resolve both deep and shallow relationships within the spider tree of life. *BMC Evolutionary Biology*, *16*(1), 212. https://doi.org/10.1186/s12862-016-0769-y

Hellgren, O., Waldenström, J., & Bensch, S. (2004). A new PCR assay for simultaneous studies of *Leucocytozoon*, *Plasmodium*, and *Haemoproteus* from avian blood. *Journal of Parasitology*, *90*(4), 797–802. https://doi.org/10.1645/GE-184R1

Hellgren, O., Atkinson, C. T., Bensch, S., Albayrak, T., Dimitrov, D., Ewen, J. G., … Marzal, A. (2015). Global phylogeography of the avian malaria pathogen *Plasmodium relictum* based on MSP1 allelic diversity. *Ecography*, *38*(8), 842–850. https://doi.org/10.1111/ecog.01158

Huang, X., Hansson, R., Palinauskas, V., Valkiūnas, G., Hellgren, O., & Bensch, S. (2018). The success of sequence capture in relation to phylogenetic distance from a reference genome: A case study of avian haemosporidian parasites. *International Journal for Parasitology*, *48*(12), 947–954. https://doi.org/10.1016/j.ijpara.2018.05.009

Ishtiaq, F., Rao, M., Huang, X., & Bensch, S. (2017). Estimating prevalence of avian haemosporidians in natural populations: A comparative study on screening protocols. *Parasites and Vectors*, *10*, 127. https://doi.org/10.1186/s13071-017-2066-z

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Lanfear, R., Calcott, B., Kainer, D., Mayer, C., & Stamatakis, A. (2014). Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evolutionary Biology*, *14*(1), 82. https://doi.org/10.1186/1471-2148-14-82

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2017). PARTITIONFINDER 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, *34*(3), 772–773. https://doi.org/10.1093/molbev/msw260

Lemmon, A. R., Emme, S. A., & Lemmon, E. M. (2012). Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, *61*(5), 727–744. https://doi.org/10.1093/sysbio/sys049

Loiseau, C., Harrigan, R. J., Robert, A., Bowie, R. C. K., Thomassen, H. A., Smith, T. B., & Sehgal, R. N. M. (2012). Host and habitat specialization of avian malaria in Africa. *Molecular Ecology*, *21*(2), 431–441. https://doi.org/10.1111/j.1365-294X.2011.05341.x

Loiseau, C., Harrigan, R. J., Cornel, A. J., Guers, S. L., Dodge, M., Marzec, T., … Sehgal, R. N. M. (2012). First evidence and predictions of Plasmodium transmission in Alaskan bird populations. *PLoS ONE*, *7*(9), e44729. https://doi.org/10.1371/journal.pone.0044729

Lutz, H. L., Marra, N. J., Grewe, F., Carlson, J. S., Palinauskas, V., Valkiūnas, G., & Stanhope, M. J. (2016). Laser capture microdissection microscopy and genome sequencing of the avian malaria parasite, *Plasmodium relictum*. *Parasitology Research*, *115*(12), 4503–4510. https://doi.org/10.1007/s00436-016-5237-5

Lutz, H. L., Hochachka, W. M., Engel, J. I., Bell, J. A., Tkach, V. V., Bates, J. M., … Weckstein, J. D. (2015). Parasite prevalence corresponds to host life history in a diverse assemblage of afrotropical birds and haemosporidian parasites. *PLoS ONE*, *10*(4), e0121254. https://doi.org/10.1371/journal.pone.0121254

Marroquin-Flores, R. A., Williamson, J. L., Chavez, A. N., Bauernfeind, S. M., Baumann, M. J., Gadek, C. R., … Barrow, L. N. (2017). Diversity, abundance, and host relationships in the avian malaria community of New Mexico pine forests. *PeerJ*, *5*, e3700. https://doi.org/10.7717/peerj.3700

Martinsen, E. S., Perkins, S. L., & Schall, J. J. (2008). A three-genome phylogeny of malaria parasites (Plasmodium and closely related genera): Evolution of life-history traits and host switches. *Molecular Phylogenetics and Evolution*, *47*(1), 261–273. https://doi.org/10.1016/j.ympev.2007.11.012

Nilsson, E., Taubert, H., Hellgren, O., Huang, X., Palinauskas, V., Markovets, M. Y., … Bensch, S. (2016). Multiple cryptic species of sympatric generalists within the avian blood parasite Haemoproteus majoris. *Journal of Evolutionary Biology*, *29*(9), 1812–1826. https://doi.org/10.1111/jeb.12911

Olsson-Pons, S., Clark, N. J., Ishtiaq, F., & Clegg, S. M. (2015). Differences in host species relationships and biogeographic influences produce contrasting patterns of prevalence, community composition and genetic structure in two genera of avian malaria parasites in southern Melanesia. *Journal of Animal Ecology*, *84*(4), 985–998. https://doi.org/10.1111/1365-2656.12354

Outlaw, D. C., & Ricklefs, R. E. (2014). Species limits in avian malaria parasites (Haemosporida): How to move forward in the molecular era. *Parasitology*, *141*, 1223–1232. https://doi.org/10.1017/S0031182014000560

Pacheco, M. A., Matta, N. E., Valkiūnas, G., Parker, P. G., Mello, B., Stanley, C. E., … Escalante, A. A. (2018). Mode and rate of evolution of haemosporidian mitochondrial genomes: Timing the radiation of avian parasites. *Molecular Biology and Evolution*, *35*(2), 383–403. https://doi.org/10.1093/molbev/msx285

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*(2), 420–424. https://doi.org/10.1645/GE-2247.1

Palinauskas, V., Križanauskienė, A., Iezhova, T. A., Bolshakov, C. V., Jönsson, J., Bensch, S., & Valkiūnas, G. (2013). A new method for isolation of purified genomic DNA from haemosporidian parasites inhabiting nucleated red blood cells. *Experimental Parasitology*, *133*(3), 275–280. https://doi.org/10.1016/j.exppara.2012.12.003

Perkins, S. L. (2014). Malaria's many mates: Past, present, and future of the systematics of the Order Haemosporida. *The Journal of Parasitology*, *100*(1), 11–25. https://doi.org/10.1645/13-362.1.

Quattrini, A. M., Faircloth, B. C., Dueñas, L. F., Bridge, T. C. L., Brugler, M. R., Calixto-Botía, I. F., … McFadden, C. S. (2018). Universal target-enrichment baits for anthozoan (Cnidaria) phylogenomics: New approaches to long-standing problems. *Molecular Ecology Resources*, *18*(2), 281–295. https://doi.org/10.1111/1755-0998.12736

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ricklefs, R. E., Outlaw, D. C., Svensson-Coelho, M., Medeiros, M. C. I., Ellis, V. A., & Latta, S. (2014). Species formation by host shifting in avian malaria parasites. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(41), 14816–14821. https://doi.org/10.1073/pnas.1416356111

Scheuerlein, A., & Ricklefs, R. E. (2004). Prevalence of blood parasites in European passeriform birds. *Proceedings of the Royal Society B: Biological Sciences*, *271*(1546), 1363–1370. https://doi.org/10.1098/rspb.2004.2726

Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, *6*, 1–11. https://doi.org/10.1186/1471-2105-6-31

Stamatakis, A. (2014). RAXML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Swofford, D. L. (2002). *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sunderland, MA: Sinauer Associates.

Valkiūnas, G. (2005). *Avian malaria parasites and other haemosporidia*. Boca Raton, FL: CRC Press.

Valkiūnas, G., Iezhova, T., Križanauskienė, A., Palinauskas, V., Sehgal, R. N. M., & Bensch, S. (2008). A comparative analysis of microscopy and PCR-based detection methods for blood parasites. *The Journal of Parasitology*, *94*(6), 1395–1401. https://doi.org/10.1645/GE-1570.1.

van Riper, C., van Riper, S. G., Goff, M. L., & Laird, M. (1986). The epizootiology and ecological significance of malaria in Hawaiian land birds. *Ecological Monographs*, *56*(4), 327–344. https://doi.org/10.2307/1942550.

Videvall, E., Cornwallis, C. K., Ahrén, D., Palinauskas, V., Valkiūnas, G., & Hellgren, O. (2017). The transcriptome of the avian malaria parasite *Plasmodium ashfordi* displays host-specific gene expression. *Molecular Ecology*, *26*(11), 2939–2958. https://doi.org/10.1111/mec.14085

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*(1), 191–194. https://doi.org/10.1645/GE-3221RN

Zehtindjiev, P., Ilieva, M., Westerdahl, H., Hansson, B., Valkiunas, G., & Bensch, S. (2008). Dynamics of parasitemia of malaria parasites in a naturally and experimentally infected migratory songbird, the great reed warbler *Acrocephalus arundinaceus*. *Experimental Parasitology*, *119*(1), 99–110. https://doi.org/10.1016/j.exppara.2007.12.018

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Barrow LN, Allen JM, Huang X, Bensch S, Witt CC. Genomic sequence capture of haemosporidian parasites: Methods and prospects for enhanced study of host–parasite evolution. *Mol Ecol Resour.* 2019;19:400–410. https://doi.org/10.1111/1755-0998.12977