

WITHIN-LINEAGE DIVERGENCE OF AVIAN HAEMOSPORIDIANS: A CASE STUDY TO REVEAL THE ORIGIN OF A WIDESPREAD *HAEMOPROTEUS* PARASITE

Xi Huang¹, Paulius Rapševičius¹, Leonardo Chapa-Vargas², Olof Hellgren¹, and Staffan Bensch¹

¹ Molecular Ecology and Evolution Lab, Department of Biology, Lund University, Sölvegatan 37, SE-22362, Sweden.

² Instituto Potosino de Investigación Científica y Tecnológica A.C. Camino a la Presa San José 2055, Lomas 4ª Sección, San Luis Potosí, SLP. C.P. 78216, Mexico.

Correspondence should be sent to Xi Huang (<https://orcid.org/0000-0001-6515-261X>) at: xi.huang43@gmail.com

KEY WORDS ABSTRACT

Haemoproteus tartakovskyi
Passerine
Haemosporidian Parasites
Cytochrome b Lineage
Genomic Divergence
Sequence Capture
Nuclear Genes
Evolutionary Origin
Haplotype Network
Geographical Distribution

Avian haemosporidian parasites are particularly diverse and widespread. To date, more than 3,000 distinct cytochrome b lineages have been recorded, of which some present extremely wide geographical distributions, even including multiple continents. Whether these isolates represent one or several cryptic species remains unknown. Here we carried out a case study of SISKIN1, a common haemosporidian parasite lineage belonging to the morphologically described species *Haemoproteus tartakovskyi*. To shed light on its evolutionary origin, we investigated the divergence between SISKIN1 isolates obtained from siskins and redpolls in Europe (Russia and Sweden) and house finches in North America (Mexico). First, we used sequence capture of a small data set (2 Russian isolates and 1 Mexican isolate) to investigate the genetic structure based on the full-length mitochondrial genome and ~1,000 genes. The mitochondrial genomes of Russian isolates were identical with each other but differed from the Mexican one at 6 positions. The nuclear divergence between Russian and Mexican isolates was on average 2.8%, close to what has been observed between 2 species of malaria parasites that respectively infect humans (*Plasmodium falciparum*) and gorillas (*Plasmodium praefalciparum*). Second, we used the expanded data set (15 samples in total) to investigate the genetic structure in 3 genes known to be involved in host invasion. The European isolates were identical across all sequenced genes, whereas the Mexican isolates were highly diverse. The lack of shared alleles between European and Mexican populations suggests that they might have diverged in isolation without gene flow. From the MalAvi database we examined the lineages most similar to the SISKIN1 barcode fragment (part of the *cyt b* gene) and found that most of them had been recorded in North and South America. This suggests that the lineage SISKIN1 originated in North America and subsequently spread to Europe. Our analyses support that the *cyt b* gene barcoding region is a useful marker for identification of avian haemosporidian lineages that can classify them into clusters of closely related parasites, but to further investigate species limits and evolutionary history, molecular data from multiple faster-evolving genes are required.

Haemosporidian parasites are transmitted by blood-sucking dipteran insects to their vertebrate hosts and are particularly diverse and widespread in birds (Valkiunas, 2005). For more than 100 yr, avian haemosporidians have served as a model system for research on evolutionary patterns and ecological interactions of host-parasite associations (Pérez-Tris et al., 2005; Rivero and Gandon, 2018).

The identification of avian haemosporidians is normally performed by either morphological or molecular analyses of blood samples from infected birds. By using the traditional microscopic method for examinations of blood smears, more than 250 species of avian haemosporidians have been described, mainly based on morphological characters of gametocytes and meronts

(Valkiunas, 2005). During the past 20 yr various molecular identification methods have been developed, of which most are targeting a fragment of the cytochrome b (*cyt b*) gene located in the mitochondrial genome (Bensch et al., 2000; Perkins and Schall, 2002; Beadell et al., 2004; Hellgren et al., 2004; Waldenström et al., 2004). Parasites that differ by as little as 1 base pair in the partial 479 bp *cyt b* sequence are defined as unique lineages, and to date, more than 3,000 distinct *cyt b* lineages have been detected (Bensch et al., 2009).

Infections with parasite lineages that show similar morphological characters are typically defined as the same morphological species. Hellgren et al. (2007) showed that a *cyt b* distance of >5% generally corresponds to different morphological species.



However, the MalAvi database has compiled some morphologically divergent parasite lineages that differ in *cyt b* by less than 0.3% (e.g., *Haemoproteus minutus* and *Haemoproteus pallidus* differ by 0.21%; *Haemoproteus balmorali* and *Haemoproteus attenatus* differ by 0.22%). Molecular analyses based on nuclear genes have demonstrated that lineages with similar or no notable morphological characters in some cases correspond to cryptic species (Hellgren et al., 2015; Nilsson et al., 2016). Unlike in most eukaryotes, the mitochondrial genes of haemosporidians evolve slowly (Ricklefs and Outlaw, 2010; Bensch et al., 2013), with a divergence rate that is only 10% of the rate observed in nuclear genes (Nilsson et al., 2016). With the limited information generated from sequences of partial *cyt b* gene, we cannot determine whether similar lineages represent different biological species or within-species diversity.

Among the thousands of the avian haemosporidian lineages recorded to date, a surprising number have been encountered on multiple continents (Ellis et al., 2018) and appear to have wider distributions than many of the bird species they infect. The dispersal of those parasite lineages might be mediated by migratory birds (Ricklefs et al., 2017); however, some of the distributions cannot be explained by the current migration pattern. As the barcoding region in mitochondrial *cyt b* gene is only 479 bp in length, it may not provide sufficient information for resolving phylogenetic relationships between haemosporidian parasites (Ricklefs et al., 2004). Additionally, parasites that are not closely related might become coincidentally identical in the short *cyt b* fragment as a result of parallel- or back-mutations.

The lineage SISKIN1, which is morphologically described as *Haemoproteus (Parahaemoproteus) tartakovskyi* (Valkiunas et al., 2008), has been recorded in different resident bird species in both Eurasia and North America. By comparing the genomic variation between isolates originating from different geographical locations, we tested whether they represent identical parasites or are just coincidentally similar in the short *cyt b* fragment for lineage definition.

In this study, we applied the sequence capture method (Huang et al., 2018) on a limited number of samples to compare the complete mitochondrial genome and ~1,000 nuclear genes between SISKIN1 isolates collected from Russia and Mexico. In an expanded data set, we used nested PCR and Sanger sequencing of 3 nuclear genes in the host invasion pathway to further investigate the genetic structure between the populations in Europe (Russia and Sweden) and North America (Mexico).

The aim of this study was to answer the following questions: Can the identical sequence in the *cyt b* barcoding region be explained by coincidental parallel- or back-mutations, or is the similarity distributed across the mitochondrial genome, supporting that they share a recent ancestor? Is the mitochondrial similarity between isolates obtained from different continents reflected in the nuclear genome? If the European and North American SISKIN1 isolates have a common ancestor, where and when did they diverge?

MATERIALS AND METHODS

Data collection

We downloaded the barcoding *cyt b* sequences of all *Haemoproteus* lineages from the MalAvi database and calculated their identity to the lineage SISKIN1 in Geneious version 8.1.9.

The 20 lineages that were most identical (>98% similarity) with SISKIN1 as well as 4 other lineages morphologically described as *H. tartakovskyi* were selected for further analysis. A Bayesian tree based on the partial *cyt b* gene (479 bp) was constructed in MrBayes (Ronquist and Huelsenbeck, 2003) with the lineage PARUS1 (*Haemoproteus majoris*), which was 95.8% identical with SISKIN1 in barcoding sequence as an outgroup for rooting. The GTR+G nucleotide substitution model was selected as the best model according to the corrected Akaike Information Criterion using jModelTest version 2.1.4 (Darriba et al., 2012). Two independent Markov chain Monte Carlo analyses, each with 6 chains, were run simultaneously for 3 million generations and sampled every 1,000 steps. The first 750 trees were discarded as burn-in.

To estimate the likely divergence history, we extracted the information of host range and geographical distribution of the 25 selected lineages from the “Hosts and Sites Table” in the MalAvi database and mapped them to the phylogenetic tree using the phyloseq R package (McMurdie and Holmes, 2013).

Molecular analysis

We used a targeted sequence capture protocol based on probes designed from the published *H. tartakovskyi* genome (Bensch et al., 2016) to sequence up to 2,565 exons in 1,000 genes from 3 SISKIN1 isolates (Huang et al., 2018). Two of the isolates were from siskins (*Spinus spinus*) in Russia (Kaliningrad), and 1 from a house finch (*Haemorhous mexicanus*) in Mexico. To ensure sequencing success, infection intensities were assessed using a general qPCR protocol amplifying a partial 16S rRNA gene, and these 3 samples were selected because of their relatively high infection intensities (1.73% and 0.88% for the Russian isolates and 6.06% for the Mexican one), as described in Huang et al. (2018). Sequence capture was performed using the SureSelect^{XT} Target Enrichment System kits (Agilent Technologies, Santa Clara, California) following the manufacturer’s protocol (version B4). Briefly, DNA samples were fragmented, end-repaired, and ligated to Illumina adaptors before hybridization with the biotin-labeled probes. The bound probe and targeted DNA were then captured with streptavidin-coated magnetic beads and separated from the nontargeted background DNA. After that, each sample was indexed with a unique barcode and pooled for multiplex sequencing on an Illumina MiSeq platform with Generate FASTQ workflow (Illumina, San Diego, California) using the paired-end method with a mean length of 300 bp per read. The obtained sequences were trimmed for quality using default settings in Trimmomatic (Bolger et al., 2014) and mapped to the reference genome. The coverage and divergence in each exon were obtained from a custom automated pipeline, as described at www.github.com/Roland-Hansson.

Exons shorter than 100 bp were excluded in statistical analyses. For 14 nuclear genes that have been confirmed to be conserved across the genera *Plasmodium* and *Haemoproteus* (Huang et al., 2018), we calculated the nucleotide divergence and nonsynonymous substitution ratio (Ka/Ks) between the Russian and Mexican isolates and compared it with the divergence between *Plasmodium falciparum* and its closest gorilla malaria parasite, *Plasmodium praefalciparum* (Otto et al., 2018).

To verify the differences detected by sequence capture, we selected 11 additional samples with known SISKIN1 infections

Table I. Summary of sequenced samples containing SISKIN1 infections. The sequence capture method was used to sequence 3 of the samples, and Sanger sequencing was carried out for the rest. Sequences with heterozygous sites (mix) as well as low-quality ones (–) were excluded, and the alleles were defined using the Pegas package in R.

Sample ID	Host species	Country	Location	Sequencing method	CTR _P allele	AMA1 allele	MSP1 allele
SISKIN1_ref	<i>Spinus spinus</i>	Russia	Kaliningrad	Genome	I	I	I
102-11C	<i>Spinus spinus</i>	Russia	Kaliningrad	SeqCap	I	I	I
126-11C	<i>Spinus spinus</i>	Russia	Kaliningrad	SeqCap	I	I	I
BT48280	<i>Acanthis flammea</i>	Sweden	Ammarnäs	Sanger	I	I	I
D00495	<i>Haemorrhous mexicanus</i>	Mexico	Charcas	Sanger	II	V	–
D00485	<i>Haemorrhous mexicanus</i>	Mexico	Charcas	Sanger	IV	IV	–
D00439	<i>Haemorrhous mexicanus</i>	Mexico	Charcas	Sanger	II	III	VII
D00405	<i>Haemorrhous mexicanus</i>	Mexico	Charcas	Sanger	Mix	II	II
C00484	<i>Haemorrhous mexicanus</i>	Mexico	Charcas	Sanger	III	Mix	VI
C00434	<i>Haemorrhous mexicanus</i>	Mexico	Charcas	Sanger	II	II	II
C00434	<i>Haemorrhous mexicanus</i>	Mexico	Charcas	SeqCap	II	II	II
C00413	<i>Haemorrhous mexicanus</i>	Mexico	Charcas	Sanger	II	Mix	V
C00425	<i>Haemorrhous mexicanus</i>	Mexico	Charcas	Sanger	II	II	II
B00417	<i>Haemorrhous mexicanus</i>	Mexico	Charcas	Sanger	II	Mix	IV
B00338	<i>Haemorrhous mexicanus</i>	Mexico	Charcas	Sanger	–	–	III
B00320	<i>Haemorrhous mexicanus</i>	Mexico	Charcas	Sanger	II	–	Mix

for nuclear gene analyses (Table I), 10 from Mexico and 1 from Sweden. SISKIN1 infections were confirmed using a nested PCR protocol (Waldenström et al., 2004) amplifying a partial *cyt b* gene located on the mitochondrial genome of the parasites. For these 11 samples, we sequenced 3 nuclear genes (Table II) linked to the host invasion pathway using nested PCR protocols and Sanger sequencing. Primers were designed based on the alignments obtained from sequence capture data. To ensure the specific amplification, we tested the protocols on one of the samples (C00434) included in the previous sequence capture experiment and compared the obtained sequences from both assays.

All PCR reactions were performed in 2 steps each with a volume of 25 µl. The first step includes 25 ng of DNA template, 1 mM of each primer, 2.5 mM of each nucleotide, 1.5 mM of MgCl₂, 0.5 units of Taq DNA polymerase (Invitrogen, Carlsbad, California), and 10× PCR buffer (Invitrogen). The PCR reactions start from a 2 min incubation at 94 C, followed by 20 cycles of 30 sec at 94 C, 30 sec at 52–55 C (as described in Table II), and 45 sec at 72 C, and end up with a 10 min extension at 72 C. Once the first

step is done, 1 µl of the PCR product is used as a template in the second step, with a similar profile except that the amplification cycles were set to 35. Positive amplifications were confirmed by the presence of bands of the expected length on a 2% agarose gel before sequencing.

PCR products were precipitated and sequenced using the BigDye® Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, California). All samples were sequenced from both directions on an ABI 3130xl sequencer (Applied Biosystems). Sequences with low-quality or heterozygous sites that could not be resolved were excluded in further analyses. All obtained sequences were aligned together with the reference genome in Geneious version 8.1.9. Haplotype networks for each gene were constructed to visualize the genetic structure using the Pegas R package (Paradis, 2010), based on a modified maximum parsimony algorithm described by Templeton and Sing (1993). Nucleotide diversity (π) and pairwise divergence (dxy) between populations were calculated in Dnasp version 5.10.01 (Librado

Table II. Primer combinations for amplification of each nuclear gene.

Gene name	Sequence length (bp)	PCR step	Primer name	Primer sequence (5'-3')	Annealing temperature (C)
Circumsporozoite and thrombospondin-related protein (CTR _P)	518	1	CTR _P _F1	GTGTTGTTCCATTGACGCA	55
			CTR _P _R1	GTTCTTTACGTTTGTATATACTCCAG	
		2	CTR _P _F2	ACTTAAAGATTGGTGAATAATGAAATACA	55
			CTR _P _R2	AAACACTCTTTACCTGTACATTCTG	
Apical membrane antigen 1 (AMA1)	416	1	AMA1_F1	AATTCCCAATCCTACCGAATGC	56
			AMA1_R1	ATGCATTTCTTCACAGTTTCCATC	
		2	AMA1_F2	TTGGAAAAGAAATAACTATATCAAATTC	52
			AMA1_R2	ATTTTCTATATTTTTGCGTGGAC	
Merozoite surface protein 1 (MSP1)	415	1	MSP1_F1	CACATTCCAGTTCAGATTCAAATG	55
			MSP1_R1	GCTGCTGTATTTGTACTGTAGTTAC	
		2	MSP1_F2	TTCAAATGTTTCAGATGTAGCAGA	55
			MSP1_R2	GTTACATTTGTATTTGCAGGTGTC	

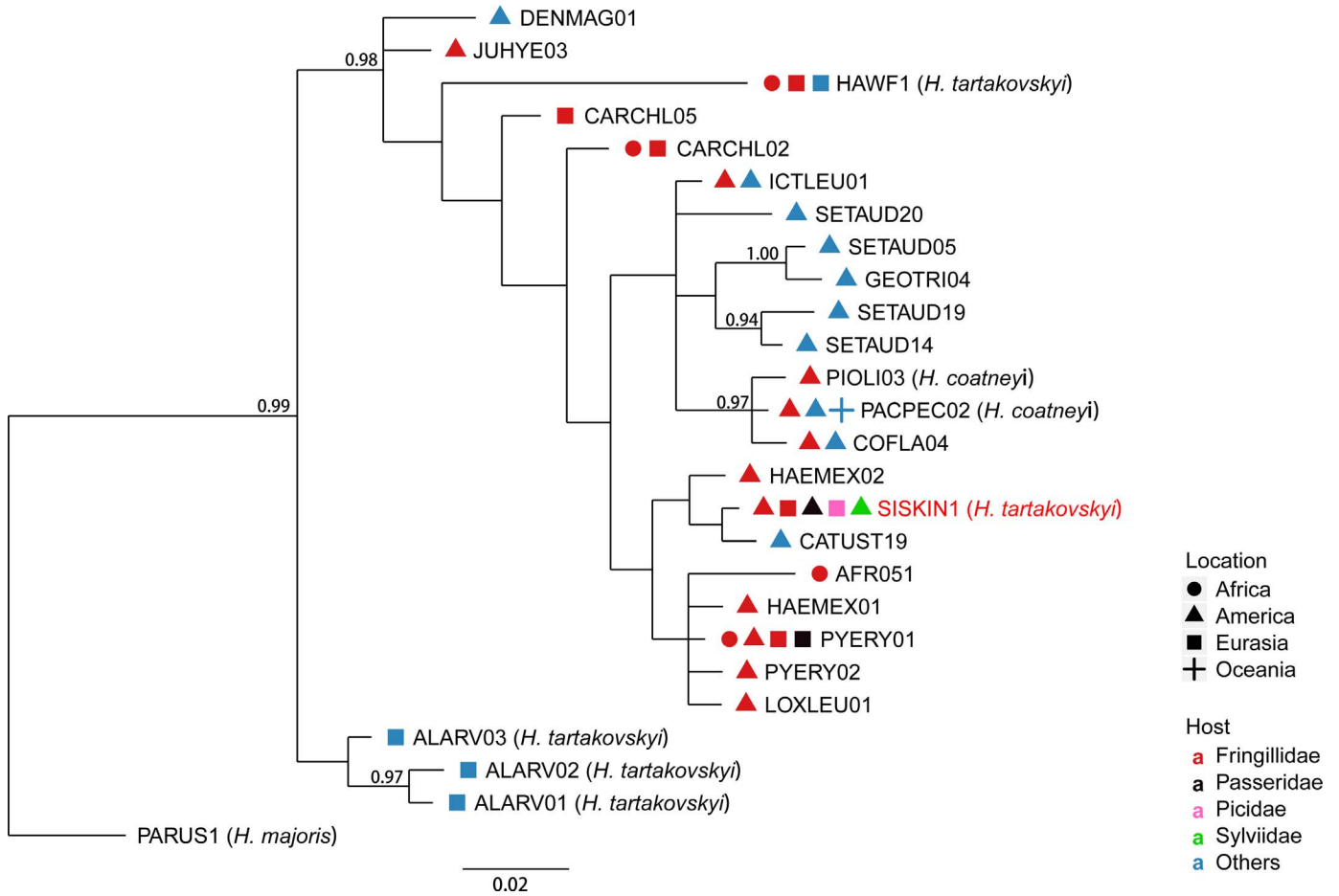


Figure 1. Host range and geographical distribution of lineages that are most identical to SISKIN1, molecularly or morphologically. The phylogenetic tree was constructed based on the MalAvi barcoding fragment (479 bp from the *cyt b* gene), and posterior probability values higher than 0.9 are shown next to the nodes. Host families without records of SISKIN1 infections are presented as “others.” A color version is available online.

and Rozas, 2009). All sequences were submitted to GenBank (accession numbers MH628537–MH628562).

RESULTS

Host range and geographical distribution of closely related lineages

For phylogenetic analysis, we selected 20 lineages from the MalAvi database, which had a sequence difference of <2% to SISKIN1, and 4 lineages that had been described as the same morphospecies, i.e., *H. tartakovskiy* in previous studies (Suppl. Table 1). The host range and geographical distribution of these 25 lineages (including SISKIN1) were mapped to a Bayesian tree based on the partial *cyt b* gene (Fig. 1). Although most of the nodes had low support, it is clear that SISKIN1 lineage belongs to a cluster of lineages primarily recorded in North or South America. Among the lineages that were similar to SISKIN1 in the barcoding region, 17 of 21 have previously been detected in the Americas, whereas only 4 have been recorded in Africa and Eurasia. Fourteen of them have been recorded in birds of the family Fringillidae, which is also the main host family for SISKIN1. The vast majority of the lineages tend to be host

specialists (except PACPEC02), as they have been detected in only a few closely related host species.

Of the 4 lineages (ALARV01, 02, 03, and HAWF1) that are morphologically similar with SISKIN1 and thus previously described as *H. tartakovskiy*, only the lineage HAWF1 was in the clade of lineages with high sequence similarity to SISKIN1. However, HAWF1 was not among the lineages most similar to SISKIN1, although it shares the main host family (Fringillidae) with other lineages in this clade.

Genome-wide variation between isolates

With the sequence capture protocol designed based on a reference genome obtained from a Russian SISKIN1 isolate (SISKIN1_ref), we successfully obtained sequences from 1,875 exons belonging to 970 genes (including 3 mitochondrial genes and 967 nuclear genes, total length 2,259,888 bp) in 3 SISKIN1 isolates. The mean proportion of exons recovered (in term of length) was slightly higher for the Russian isolates (99.80%) than for the Mexican one (98.25%), although the Mexican isolate presented a higher infection intensity.

The 2 Russian isolates were identical to the reference across the whole mitochondrial genome, whereas the single Mexican isolate differed at 6 different sites, all located outside the MalAvi

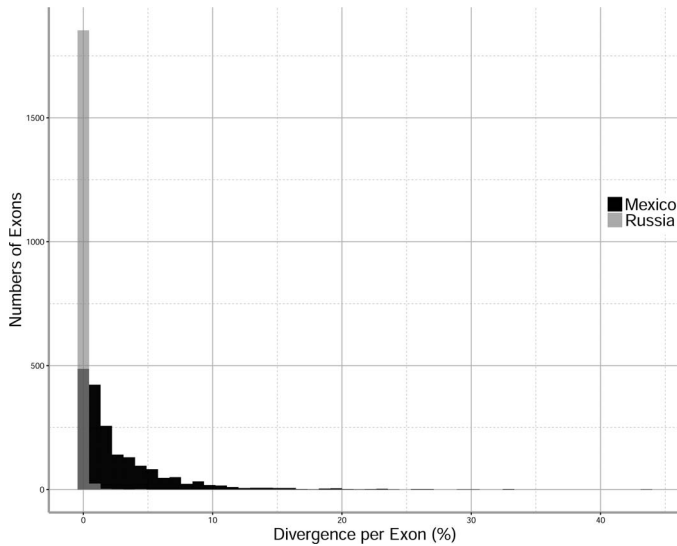


Figure 2. Divergence rate in each exon between the sequence capture reference (SISKIN1 from Russia) and the SISKIN1 isolates obtained from Russia and Mexico. Only nuclear genes were included in this analysis.

barcoding segment (2 in *cyt b*, 1 in *COI*, 1 in *COIII*, and 2 in the noncoding region), corresponding to a sequence similarity of 99.89%.

A much higher divergence was found across the nuclear genes obtained from the sequence capture experiment (Fig. 2). Among all the obtained nuclear gene sequences from the Russian isolates, a total of 770 SNPs were detected from 424 exons when compared to the reference genome, corresponding to a divergence of 0.04%. The sequence divergence between the Mexican isolate and the reference was 2.85%, with 27,359 SNPs detected from 1,725 exons.

Table III. Comparison of single-gene divergence and the ratio of nonsynonymous to synonymous substitution (Ka/Ks) between 2 pairs of parasites, Russian and Mexican isolates of SISKIN1 vs. the human malaria parasite *Plasmodium falciparum* and the gorilla parasite *Plasmodium praefalciparum*. The comparison includes 14 conserved nuclear genes and 3 genes related to host invasion.

Gene no.	Length*	Gene name	Divergence (%)		Ka/Ks ratio	
			SISKIN1	<i>Plasmodium</i>	SISKIN1	<i>Plasmodium</i>
HtGene0168	1,182	26S protease regulatory subunit s10b	0.59	0.17	0	0
HtGene0526	678	14-3-3 protein	0.17	0	0	–
HtGene0575	402	Histone H2A	0.25	0	0	–
HtGene0764	420	40S ribosomal protein S10	0.24	0	0	–
HtGene1012	456	Ribosomal protein S15	0.66	0.22	0	0
HtGene1080	2,169	Heat shock protein 86	0.51	0.63	0	0.37
HtGene1229	546	Proteasome subunit alpha type	0.55	0	0	–
HtGene1356	2,433	Elongation factor 2	0.77	0.08	0.84	0
HtGene1519	657	60S ribosomal protein L10	0.91	0.1	0.06	0
HtGene1772	1,356	26S protease regulatory subunit	0.37	0.08	0	0
HtGene1888	789	40S ribosomal protein S3a	0.48	0.32	–	0
HtGene2958	525	Histamine-releasing factor	1.14	0	0	–
HtGene3475	429	DIM1 protein homolog	1.63	0.71	0	0
HtGene3631	1,110	MORN repeat-containing protein 1	0.45	0.09	0	–
HtGene4136_4137	5,229	CTRP	1.63	1.99	0.33	0.34
HtGene0852	1,809	AMA1	2.16	1.55	0.71	1.11
HtGene0562	5,749	MSP1	4.64	28.11	1.20	0.74

* Length as in the genome of *Haemoproteus tartakovskyi* (Bensch et al., 2016).

In order to estimate the evolutionary distance between the SISKIN1 isolates from Russia and Mexico, we compared their divergence level with that between 2 closely related malaria parasites infecting primates (*P. falciparum* and *P. praefalciparum*). We investigated 14 conserved genes and 3 presumed fast-evolving genes involved in host invasion pathway (Table III). The genes in the host invasion pathway (*CTRP*, *AMA1*, and *MSP1*) showed much higher divergence and Ka/Ks ratios for both the bird and primate haemosporidians comparing with the 14 conserved nuclear genes. For all 17 genes, the 2 Russian isolates presented identical sequences with the reference. The 14 conserved genes were slightly more divergent between the SISKIN1 isolates (0.62% on average) than between the primate parasites (0.17% on average), while the gene *MSP1* was notably much more divergent between *P. falciparum* and *P. praefalciparum* (28.11%) than between the SISKIN1 isolates (4.64%).

Genetic structure and phylogeographical pattern

To confirm and further investigate the genomic diversity of the lineage SISKIN1 from different geographical locations, we selected the above-mentioned 3 nuclear genes for partial gene sequencing of 12 samples with SISKIN1 infections, including the sample C00434, which was involved in the previous sequence capture experiment (Table I). Reassuring, the sequences obtained from C00434 by both methods were consistent. When sequencing these genes in the additional 11 samples, we failed to sequence all 3 genes in some of them, probably due to low infection intensities and mixed infections. Among the successfully sequenced samples, we identified 4 alleles in the gene *CTRP* ($\pi = 0.012$), 5 in *AMA1* ($\pi = 0.034$), and 7 alleles in *MSP1* ($\pi = 0.069$).

The haplotype networks based on these 3 genes all showed a similar pattern (Fig. 3). The European (Russian and Swedish) isolates shared the same allele, which was absent among the Mexican isolates, while the latter presented multiple unique and

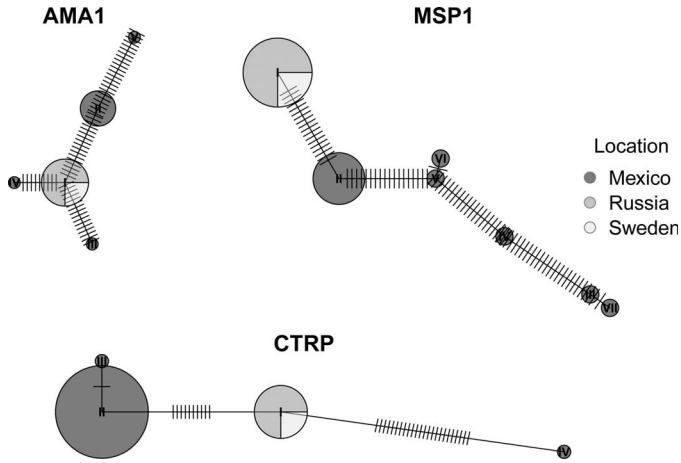


Figure 3. Haplotype networks of SISKIN1 isolates from different geographical locations. The number of mutations are indicated as hatch marks on the links between the haplotype pie charts.

distantly related alleles, with at least 9 mutational steps from the European allele. The pairwise divergence between the European and Mexican populations for the sequenced region of the *CTRP* gene (0.023 ± 0.01) was close to the average exon divergence (0.028) but much higher for the genes *AMA1* (0.042 ± 0.02) and *MSP1* (0.088 ± 0.04). Of the 3 genes, *MSP1* presented the highest divergence (maximum 51 mutation sites in 415 bp) and also contained several insertions and deletions of up to 18 contiguous nucleotides (6 amino acids) in the sequence alignment (Fig. 4).

DISCUSSION

In this study, we found that isolates of a common avian haemosporidian parasite, SISKIN1 (*H. tartakovskyi*), were almost identical across the whole mitochondrial genome. This result indicates that SISKIN1 isolates obtained from the Old and the New World likely share a recent ancestor, rather than coincidentally having identical sequences in the barcoding region of the *cyt b* gene due to convergent evolution or back-mutations. However, when analyzing nearly 1,000 nuclear genes, we found a substantial difference between the Russian and Mexican isolates. This pattern was confirmed by sequencing more samples for 3 host invasion genes. Hence, identical *cyt b* lineages obtained from different zoogeographic regions may not represent the same strain of a widespread parasite but instead represent variants that are substantially divergent in their nuclear genome. Moreover, the lineages that are closely related to SISKIN1 were recorded mostly from the Americas, suggesting that the lineage SISKIN1 may have originated from North America and subsequently spread to Eurasia.

The analyses of the 3 protein-coding genes involved in host invasion for the expanded dataset (Table I) showed that the Mexican isolates were highly diverse, whereas the European isolates shared the same allele. This was despite the Mexican isolates being collected from the same host species in 1 location, while the European isolates were collected from 2 bird species in different countries. Hence, despite the broader sampling, the European isolates were much more similar to each other than were the Mexican ones. Among the 3 sequenced genes, the *CTRP* gene, which is expressed in motile ookinetes infecting the vector's

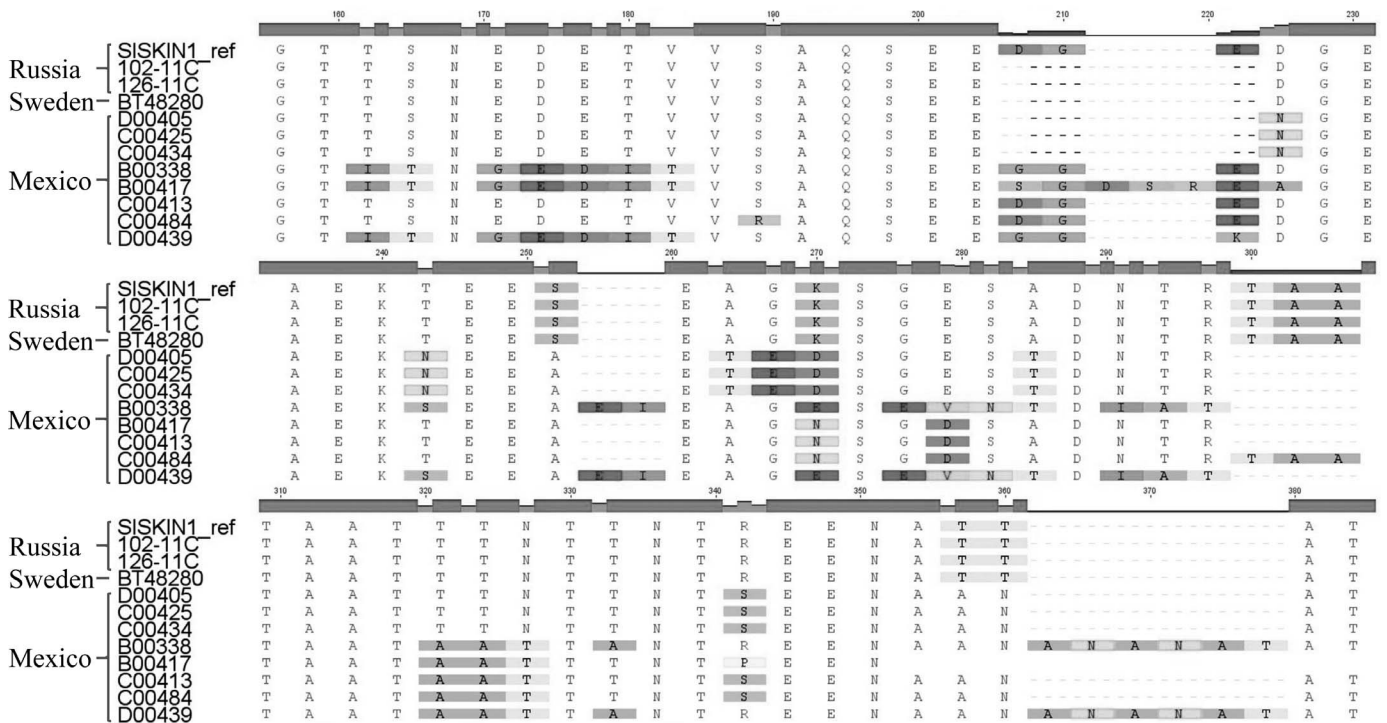


Figure 4. Amino acid alignment of the partial *MSP1* gene sequence showing a region with multiple indels. Among the 12 SISKIN1 isolates from 3 different geographical locations, 8 distinct alleles were identified. The alignment was obtained from Geneious version 8.1.9 with the substitutions highlighted. For details of the different samples, see Sample ID in Table I.

midgut (Morahan et al., 2009), showed a similar divergence rate to the average divergence of 1,000 genes throughout the genome, and the genes *MSP1* (Conway et al., 2000) and *AMA1* (Remarque et al., 2008), which are expressed during the erythrocyte-invading stage in vertebrate hosts, appeared to be 1.5–3 times more divergent than average. Therefore, these fast-evolving genes could be useful molecular markers to investigate the phylogenetic relationship between closely related haemosporidian parasites.

A previous study investigating variation in the *MSP1* gene of the lineage SGS1 (*P. relictum*) found high diversity among isolates from eastern Asia, whereas the European isolates appeared to be monomorphic, indicating recent expansion of the parasite (Hellgren et al., 2015). Similarly, the low genetic diversity in European SISKIN1 isolates might be a result of a founder effect. However, as the European and North American samples were taken from different host species, we cannot tell whether the divergence in host invasion genes between the 2 populations was a result of adaptation to different hosts or resulting from geographical isolation. Future work on samples collected from different host species in more populations is needed before we can reveal the evolutionary history of this lineage.

Although the lineage SISKIN1 has multiple recorded host species, it was mostly detected in birds belonging to the family Fringillidae, with only 1 case in Passeridae (Marzal et al., 2011), 1 in Picidae (Nourani et al., 2018), and 1 in Sylviidae (S. C. Galen, unpubl. data). In all 3 of these cases, only 1 individual was found infected, which might be spillover infections from sympatric Fringillidae hosts. Because of the lack of infection intensity information, we do not know if these were abortive infections.

Most of the lineages closely related to SISKIN1 appear to be host specialist and restricted to the Americas, further supporting that SISKIN1 originated from America. Given the diverse records of this lineage in both migratory and resident bird species in Alaska, it may have spread to Eurasia from North America via Alaska. The geographical range of a parasite often expands with its mobile host (Poulin, 2011), in the case of SISKIN1, it might have spread with the common redpoll (*Acanthis flammea*), which has been recorded with SISKIN1 infections in both Europe and North America. This species seems to lack population genetic structure across its Holarctic breeding range (Mason and Taylor, 2015). Although strictly speaking, the common redpoll is not a migratory species, it is highly nomadic, demonstrated by 7 ringing recoveries (birds ringed and later recaptured or found dead) connecting northwestern Europe with eastern China (<https://www.turnstones.org/chinese-mealy-redpoll>). There is also 1 record of a Michigan-ringed bird recaptured in eastern Russia (Troy, 1983). Hence, it is tempting to speculate that SISKIN1 reached Russia from Alaska together with the nomadic common redpolls. Alternatively, the long-distance dispersal might be induced by dipteran vectors carried by wind (Santiago-Alarcon et al., 2012) or along with introduced species (Woodworth et al., 2005; Jarvi et al., 2013).

To roughly estimate the divergence history, we compared pairwise divergence between the different SISKIN1 isolates and between the human malaria parasite *P. falciparum* and its close relative *P. praefalciparum*, which infects gorillas (Liu et al., 2010). In the mitochondrial DNA sequences (3,428 bp including 3 protein-coding genes), the European and American isolates are

slightly less divergent (0.14%) than the 2 *Plasmodium* parasites (0.34%). In contrast to the mitochondrial genes, the 14 nuclear genes that have been confirmed to be phylogenetically conserved across *Haemoproteus* and *Plasmodium* parasites (Huang et al., 2018) showed a higher divergence between the SISKIN1 isolates (average 0.662%) than between the 2 *Plasmodium* parasites (average 0.175%). The divergence rate between the SISKIN1 isolates was also higher for the *AMA1* gene, but not for the genes *CTRP* and *MSP1*, where the *Plasmodium* parasites are more divergent.

Otto et al. (2018) estimated that *P. falciparum* and *P. praefalciparum* diverged 40,000–60,000 yr ago. This estimate is based on a number of assumptions (e.g., accurate estimates of generation time and mutation rate, etc.) that we cannot verify to hold true for *Haemoproteus* parasites. However, the Russian and Mexican SISKIN1 isolates seem to be similarly divergent from each other as the human and gorilla malaria parasites that are proposed to be different species. Although closely related to *P. falciparum*, the gorilla parasite *P. praefalciparum* seems unable to establish repeated infection in humans (Liu et al., 2010). To confidently conclude that the European and North American SISKIN1 isolates represent different species, we need to verify their divergence in host repertoire. Future work is required to investigate the host range of both European and Mexican isolates, as well as genomic divergence between those infecting different host species.

It is notable that the 4 other lineages defined as *H. tartakovskyi* are unlikely to be close relatives with SISKIN1. Despite the similar morphological characters, they appeared to be phylogenetically distant from SISKIN1 and colonize bird hosts from different families, suggesting that they are more likely to be reproductively isolated species. The traditional morphological identification method is based on a certain set of characters and sometimes may not be valid enough to define cryptic species (Ricklefs et al., 2004; Hellgren et al., 2015). The similarity between these lineages might be a result of convergent evolution in similar habitats, or they may differ in some other morphological or life-cycle characters that have not yet been detected.

Our study has supported that the *cyt b* barcoding region is a useful marker for identification of haemosporidian mitochondrial lineages that can be used for classifying them into clusters of closely related parasites. Therefore, it should be a useful approach in surveys of haemosporidian abundance and diversity in a given host population or community. However, substantial genomic variation can exist below the level of *cyt b* lineages, and isolates belonging to the same lineage may sometimes even represent different species. To obtain robust phylogenetic patterns and estimate evolutionary history of avian haemosporidians, additional molecular data from multiple fast-evolving genes are required.

ACKNOWLEDGMENTS

This work was supported by the Swedish Research Council (grant 621-2013-4839 to SB), the Mexican Council for Science and Technology (grant PN 2015-1628 to LCH), and the China Scholarship Council (Ph.D. scholarship to XH). We thank Gediminas Valkunas and Vaidas Palinauskas for providing the samples from Kaliningrad, Russia.

LITERATURE CITED

- BEADELL, J. S., E. GERING, J. AUSTIN, J. P. DUMBACHER, M. A. PEIRCE, T. K. PRATT, C. T. ATKINSON, AND R. C. FLEISCHER. 2004. Prevalence and differential host-specificity of two avian blood parasite genera in the Australo-Papuan region. *Molecular Ecology* 13: 3829–3844.
- BENSCH, S., B. CANBÄCK, J. D. DEBARRY, T. JOHANSSON, O. HELLGREN, J. C. KISSINGER, V. PALINAUSKAS, E. VIDEVALL, AND G. VALKIUNAS. 2016. The genome of *Haemoproteus tartakovskyi* and its relationship to human malaria parasites. *Genome Biology and Evolution* 8: 1361–1373.
- BENSCH, S., O. HELLGREN, A. KRIZANAUSKIENE, V. PALINAUSKAS, G. VALKIUNAS, D. OUTLAW, AND R. E. RICKLEFS. 2013. How can we determine the molecular clock of malaria parasites? *Trends in Parasitology* 29: 363–369.
- BENSCH, S., O. HELLGREN, AND J. PÉREZ-TRIS. 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., M. STJERNMAN, D. HASSELQUIST, Ö. ÖRJAN, B. HANSSON, H. WESTERDAHL, AND R. T. PINHEIRO. 2000. Host specificity in avian blood parasites: A study of *Plasmodium* and *Haemoproteus* mitochondrial DNA amplified from birds. *Proceedings of the Royal Society of London, Series B: Biological Sciences* 267: 1583–1589.
- BOLGER, A. M., M. LOHSE, AND B. USADEL. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.
- CONWAY, D. J., D. R. CAVANAGH, K. TANABE, C. ROPER, Z. S. MIKES, N. SAKIHAMA, K. A. BOJANG, A. M. J. ODUOLA, P. G. KREMSNER, D. E. ARNOT, ET AL. 2000. A principal target of human immunity to malaria identified by molecular population genetic and immunological analyses. *Nature Medicine* 6: 689–692.
- DARRIBA, D., G. L. TABOADA, R. DOALLO, AND D. POSADA. 2012. jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods* 9: 772–772.
- ELLIS, V. A., E. H. SARI, D. R. RUBENSTEIN, R. C. DICKERSON, S. BENSCH, AND R. E. RICKLEFS. 2018. The global biogeography of avian haemosporidian parasites is characterized by local diversification and intercontinental dispersal. *Parasitology* 146(2): 1–7.
- HELLGREN, O., C. T. ATKINSON, S. BENSCH, T. ALBAYRAK, D. DIMITROV, J. G. EWEN, K. S. KIM, M. R. LIMA, L. MARTIN, AND V. PALINAUSKAS. 2015. Global phylogeography of the avian malaria pathogen *Plasmodium relictum* based on MSP1 allelic diversity. *Ecography* 38: 842–850.
- HELLGREN, O., A. KRIZANAUSKIENE, G. VALKIUNAS, AND S. BENSCH. 2007. Diversity and phylogeny of mitochondrial cytochrome b lineages from six morphospecies of avian *Haemoproteus* (Haemosporida : Haemoproteidae). *Journal of Parasitology* 93: 889–896.
- HELLGREN, O., J. WALDENSTRÖM, AND S. BENSCH. 2004. A new PCR assay for simultaneous studies of *Leucocytozoon*, *Plasmodium*, and *Haemoproteus* from avian blood. *Journal of Parasitology* 90: 797–802.
- HUANG, X., R. HANSSON, V. PALINAUSKAS, G. VALKIUNAS, O. HELLGREN, AND S. BENSCH. 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: 947–954.
- JARVI, S. I., M. E. FARIAS, D. A. LAPOINTE, M. BELCAID, AND C. T. ATKINSON. 2013. Next-generation sequencing reveals cryptic mtDNA diversity of *Plasmodium relictum* in the Hawaiian Islands. *Parasitology* 140: 1741–1750.
- LIBRADO, P., AND J. ROZAS. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
- LIU, W., Y. LI, G. H. LEARN, R. S. RUDICELL, J. D. ROBERTSON, AND B. F. KEELE. 2010. Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. *Nature* 467: 420–425.
- MARZAL, A., R. E. RICKLEFS, G. VALKIUNAS, T. ALBAYRAK, E. ARRIERO, C. BONNEAUD, G. A. CZIRJAK, J. EWEN, O. HELLGREN, D. HORAKOVA, ET AL. 2011. Diversity, loss, and gain of malaria parasites in a globally invasive bird. *PLoS ONE* 6: e21905. doi:10.1371/journal.pone.0021905.
- MASON, N. A., AND S. A. TAYLOR. 2015. Differentially expressed genes match bill morphology and plumage despite largely undifferentiated genomes in a Holarctic songbird. *Molecular Ecology* 24: 3009–3025.
- MCMURDIE, P. J., AND S. HOLMES. 2013. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* 8: e61217. doi:10.1371/journal.pone.0061217.
- MORAHAN, B. J., L. WANG, AND R. L. COPPEL. 2009. No TRAP, no invasion. *Trends in Parasitology* 25: 77–84.
- NILSSON, E., H. TAUBERT, O. HELLGREN, X. HUANG, V. PALINAUSKAS, M. MARKOVETS, G. VALKIUNAS, AND S. BENSCH. 2016. Multiple cryptic species of sympatric generalists within the avian blood parasite *Haemoproteus majoris*. *Journal of Evolutionary Biology* 29: 1812–1826.
- NOURANI, L., M. ALIABADIAN, O. MIRSHAMSI, AND N. DINPARAST DJADID. 2018. Molecular detection and genetic diversity of avian haemosporidian parasites in Iran. *PLoS ONE* 13: e0206638. doi:10.1371/journal.pone.0206638
- OTTO, T. D., A. GILBERT, T. CRELLIN, U. BÖHME, C. ARNATHAU, M. SANDERS, S. O. OYOLA, A. P. OKOUGA, L. BOUNDENGA, AND E. WILLAUME. 2018. Genomes of all known members of a *Plasmodium* subgenus reveal paths to virulent human malaria. *Nature Microbiology* 3: 687–697.
- PARADIS, E. 2010. Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics* 26: 419–420.
- PÉREZ-TRIS, J., D. HASSELQUIST, O. HELLGREN, A. KRIZANAUSKIENE, J. WALDENSTRÖM, AND S. BENSCH. 2005. What are malaria parasites? *Trends in Parasitology* 21: 209–211.
- PERKINS, S. L., AND J. J. SCHALL. 2002. A molecular phylogeny of malarial parasites recovered from cytochrome b gene sequences. *Journal of Parasitology* 88: 972–978.
- POULIN, R. 2011. *Evolutionary ecology of parasites*. Princeton University Press, Princeton, New Jersey, 331 p.
- REMARQUE, E. J., B. W. FABER, C. H. M. KOCKEN, AND A. W. THOMAS. 2008. Apical membrane antigen 1: A malaria vaccine candidate in review. *Trends in Parasitology* 24: 74–84.
- RICKLEFS, R. E., S. M. FALLON, AND E. BERMINGHAM. 2004. Evolutionary relationships, cospeciation, and host switching in avian malaria parasites. *Systematic Biology* 53: 111–119.
- RICKLEFS, R. E., M. MEDEIROS, V. A. ELLIS, M. SVENSSON-COELHO, J. G. BLAKE, B. A. LOISELLE, L. SOARES, A. FECCHIO,

- D. OUTLAW, AND P. P. MARRA. 2017. Avian migration and the distribution of malaria parasites in New World passerine birds. *Journal of Biogeography* 44: 1113–1123.
- RICKLEFS, R. E., AND D. C. OUTLAW. 2010. A molecular clock for malaria parasites. *Science* 329: 226–229.
- RIVERO, A., AND S. GANDON. 2018. Evolutionary ecology of avian malaria: Past to present. *Trends in Parasitology* 34: 712–726.
- RONQUIST, F., AND J. P. HUELSENBECK. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.
- SANTIAGO-ALARCON, D., V. PALINAUSKAS, AND H. M. SCHAEFER. 2012. Diptera vectors of avian haemosporidian parasites: Untangling parasite life cycles and their taxonomy. *Biological Reviews* 87: 928–964.
- TEMPLETON, A. R., AND C. F. SING. 1993. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics* 134: 659–669.
- TROY, D. M. 1983. Recaptures of redpolls: Movements of an irruptive species. *Journal of Field Ornithology* 54: 146–151.
- VALKIUNAS, G. 2005. Avian malaria parasites and other haemosporidia. CRC Press, Boca Raton, Florida, 947 p.
- VALKIUNAS, G., T. A. IEZHOVA, A. KRIZANAUSKIENE, V. PALINAUSKAS, AND S. BENSCH. 2008. In vitro hybridization of *Haemoproteus* spp.: An experimental approach for direct investigation of reproductive isolation of parasites. *Journal of Parasitology* 94: 1385–1394.
- WALDENSTRÖM, J., S. BENSCH, D. HASSELQUIST, AND O. OSTMAN. 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.
- WOODWORTH, B. L., C. T. ATKINSON, D. A. LAPOINTE, P. J. HART, C. S. SPIEGEL, E. J. TWEED, C. HENNEMAN, J. LEBRUN, T. DENETTE, R. DEMOTS, ET AL. 2005. Host population persistence in the face of introduced vector-borne diseases: Hawaii amakihi and avian malaria. *Proceedings of the National Academy of Sciences of the United States of America* 102: 1531–1536.