**Avian haemosporidians (*Plasmodium* and *Haemoproteus*) in selected bird groups (Old world Flycatchers, Warblers, Babblers, and Thrushes) of India and their phylogenetic relationships with other lineages of the world**

Vipin[1], Ashutosh Singh[2,4], Rajnikant Dixit[3], Narinder Kumar Tripathi[1], Bhim Singh[2], Vinita Sharma[1], Chandra Prakash Sharma[2], Dhananjai Mohan[2], Sandeep Kumar Gupta[2]*

[1]*Department of Zoology, Central University of Jammu, Rahya-Suchani (Bagla), District Samba, Jammu-181143, Jammu and Kashmir, India.*

[2]*Wildlife Forensic and Conservation Genetics Cell, Wildlife Institute of India, Dehradun-248001, Uttarakhand, India.*

[3]*Laboratory of Host-Parasite Interaction Studies, ICMR-National Institute of Malaria Research, Dwarka, New Delhi.*

[4]*Salim Ali Centre for Ornithology and Natural History, Anaikatty (PO), Coimbatore, Tamil Nadu-641108, India.*

Email: vipinsharma_24@yahoo.com, as25lu@gmail.com, dixitrk@mrcindia.org, narindertripathi@gmail.com, bhim1992singh@gmail.com, vinita302003@gmail.com, cpsharma@wii.gov.in, dmohan@wii.gov.in, skg@wii.gov.in

*** Address for correspondence**

Dr. S. K. Gupta

Scientist E

Wildlife Institute of India,
Chandrabani, Dehra Dun
248 001 (U.K.), India
E-mail: skg@wii.gov.in,
Telephone: +91-135-2646343
Fax No: +91-135-2640117

**Abstract**

We examined the Avian haemosporidians (*Plasmodium* and *Haemoproteus*) status in selected bird groups (Old world Flycatchers, Warblers, Babblers, and Thrushes) of India and their phylogenetic relationships with other known lineages of the world. The use of common genetic markers with information on the geographic distribution of parasite and host samples makes the data of avian haemosporidians genetic studies more comparable and infection source surveillance ready. The prevalence of Avian haemosporidians, in above bird groups from northern, eastern, and southern parts of India and the phylogenetic genetic analysis of genetic lineages was carried out at world level to know the possible source of infection. The MCC tree revealed six Haemosporidian lineages in which one was common (H_MSP01) and five were unique (H_CYOPOL01, H_CHD01, H_CYORUB01, H_EUMTHA01, and P_GEOCIT01). The avian host richness Index was 2.0852. 9.9%, prevalence of *Haemosporidian* infection was found in 111 DNA samples belonging to 6 host species. The *Haemoproteus* prevalence was found to be 9.0% across five host species (*Phylloscopus trochiloides, Cyornis poliogenys, C. hainanus dialilaemus, C. rubeculoides, Eumiyas thalassinus*) and *Plasmodium* prevalence was 0.9% in *Geokichla citrina*. The spatial phylogeny map showed possible source populations of hosts having *H. pallidus* lineages COLL2 and PFC1 infections in parts of Africa, Europe, North America, Malaysia, and Philippines. The *Plasmodium* lineage (P_GEOCIT01) showed a weak (93.89%) similarity with PADOM16 in Egypt. The statistical analysis suggested that the haemosporidian's host species' distribution range was directly and significantly associated with the altitude, minimum temperature, and relative humidity. H_MSP01 distribution was in accordance with *H. pallidus* having a wide geographic and host range.

**Keywords:** *Plasmodium*, *Haemoproteus*, Status, Birds, Groups, India

## Introduction

*Plasmodium* and *Haemoproteus* (Apicomplexa) are common and widely distributed vector-borne blood parasites, which occur in the bird species (Atkinson and Van 1991; Valkiunas 2005). While studying the avian haemosporidians at the genetic level, the lack of common genetic markers and information on the geographic distribution of parasite and host samples makes it exceedingly difficult to compare the data (Bensch et al. 2009). Further, the uncommon lineage naming and use of a common length of a standard genetic marker by researchers make it very difficult to carry out a fair comparison of data at the end (Ricklefs and Fallon 2002; Beadell et al. 2004). Therefore, many studies, these days, are making use of the MalAvi database, which is available at http://mbio-serv4.mbioekol.lu.se/avianmalaria/index.html for download (Marzal et al. 2011; Jia et al. 2018; Gil-Vergas and Sedano-Cruz 2019). This database is developed by a partial fragment of cytochrome b gene of 479 base pairs using HAEMF and HAEMR2 primer pair for *Plasmodium*, *Haemoproteus* and Leucocytozoon parasites. It contains information on these avian blood parasites and their hosts' geographic locations (Bensch et al. 2000). Thus, removing the above problems in the present study. Very few studies on avian haemosporidians have been performed in India utilizing the MalAvi database that were not either confined to a single bird species (Ishtiaq 2017) or particular mountain ranges (Gupta et al. 2019).

The Old-world Flycatchers studied in the current study belonged to the families of Muscicapidae, Timaliidae, Stenostirdae, Phylloscopidae and Turdidae. The Muscicapidae and Turdidae families have a worldwide distribution, are strongly migratory, and winter in South-east Asia. The Warblers belong to the family Phylloscopidae are distributed in Eurasia, Wallacea, Africa, and Alaska and winter in Southeastern Asia. The Babblers belong to the family Timaliidae and are mostly distributed in the Indian Sub-continent and in Southeast Asia. So, with the help of the information retrieved from the MalAvi database, we tried to locate the source of the above two Haemosporidian infections into the current samples and their hosts through spatial phylogenetics along with other predicting parameters. The present study took into account the vast distributional range of hosts, from the extreme north to south and east of India. We examined *Plasmodium* and *Haemoproteus's* status in selected bird groups (Old world Flycatchers, Warblers, Babblers, and Thrushes) in India as these are the two most common genera of parasites found in avian blood (Marzal et al. 2011). The sampling regions represent a large range of altitudinal variation, relative humidity, and temperature which may be affecting

the host and parasite distribution. We also attempted to explain any correlation that exists among these parasites.

## Materials and Methods

*Sample collection*

DNA from blood and tissue samples of birds (n=111), belonging to the group of Old world Flycatchers (2 Families and 14 species), Warblers (1 Family and 1 species), Babblers (1 Family and 2 species), and Thrushes (1 Family and 1 species), collected from the western Himalayas, Northeastern Hill states and Eastern Ghats of India between the period of 2014 to 2018 for a different project at Wildlife Institute of India, Dehradun were used for the current study. The distribution and sampling details about each sample are provided in Fig. 1 and Table S1.

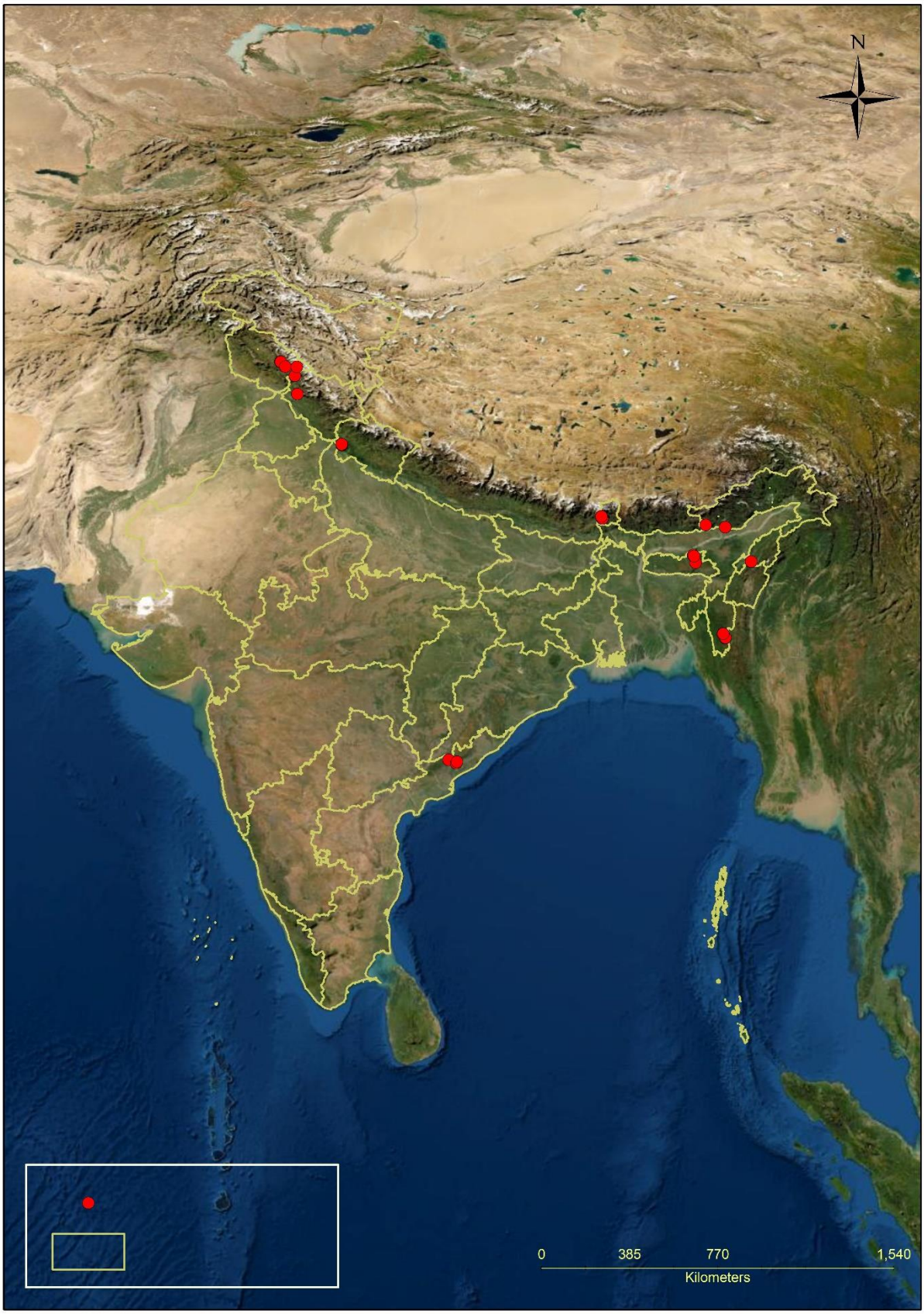


Figure 1. The distribution details of collected bird samples.

Table 1. Lineages of *Plasmodium* and *Haemoproteus* found in different host species in the present study.

| S. No. | Lineage found | Sample code | Host species name | Host family | Host group | Host location |
|---|---|---|---|---|---|---|
| 1 | H_MSP01 | LM1 | *Phylloscopus Trochiloides* | Phylloscopidae | Warbler | Lanka pakala, Andhra Pradesh |
| 2 | H_MSP01 | LM2 | *Phylloscopus Trochiloides* | Phylloscopidae | Warbler | Lambasingi, Andhra Pradesh |
| 3 | H_MSP01 | LM3 | *Phylloscopus Trochiloides* | Phylloscopidae | Warbler | Lambasingi, Andhra Pradesh |
| 4 | H_MSP01 | LM4 | *Phylloscopus Trochiloides* | Phylloscopidae | Warbler | Lambasingi, Andhra Pradesh |
| 5 | **H_CYOPOL01** | LM6 | *Cyornis poliogenys* | Muscicapidae | Flycatcher | Lanka pakala, Andhra Pradesh |
| 6 | H_MSP01 | LM10 | *Cyornis poliogenys* | Muscicapidae | Flycatcher | Lanka pakala, Andhra Pradesh |
| 7 | **P_GEOCIT 01** | LM16 | *Geokichla citrina* | Turdidae | Thrush | Lambasingi, Andhra Pradesh |
| 8 | **H_CHD 01** | AS86 | *Cyornis hainanus dialilaemus* | Muscicapidae | Flycatcher | Nongkhylem WLS, Meghalaya |
| 9 | H_MSP01 | AS89 | *Cyornis rubeculoides* | Muscicapidae | Flycatcher | Jotsama, Nagaland |
| 10 | **H_CYORUB 01** | AS91 | *Cyornis rubeculoides* | Muscicapidae | Flycatcher | Jotsama, Nagaland |
| 11 | **H_EUMTHA_01** | AS102 | *Eumyias thalassinus* | Muscicapidae | Flycatcher | Jammu & Kashmir |

The lineage prefixes 'H' and 'P' denotes *Haemoproteus* and *Plasmodium*, respectively. The lineages in bold are unique.

*DNA extraction and host species identification*

The genomic DNA was extracted using DNeasy blood and tissue kit (Qiagen, Hilden, Germany) as described by Singh et al (2020), and species identification was done as described previously by Singh et al (2017). All DNA extraction steps were carried out in a sterile and dedicated laboratory to avoid any contamination. We also included one negative extraction control to check any cross-contamination in the kit reagents.

*Identification of Plasmodium and Haemoproteus parasite lineages*

To identify the lineages of *Plasmodium* and *Haemoproteus*, the 479 bp region of the Cytochrome *b* gene was amplified using the primers HAEMF and HAEMR2 (Bensch et al. 2000). These primers were chosen because the region between them has been proven to show the almost haplotypes that could be present in the entire cytochrome *b* gene (Hellgren et al. 2007). The *Plasmodium berghei* DNA was used as positive control while performing all PCR reactions. The PCR amplification was carried out in a 10-µl reaction volume containing 8 µl of Taq PCR Master Mix (Qiagen, Hilden, Germany), 10 µM of each primer (1 µl), and 1 µl of the genomic DNA. The cycling conditions were initial denaturation at 95 °C for 5 min, followed by 35 cycles of denaturation at 95 °C for 40 s, annealing at 53 °C for 45 s, and extension at 72 °C for 45 s, with a final extension for 10 min at 72 °C in a Veriti Thermal Cycler (Applied Biosystems, Foster City, CA). We also amplified the *Plasmodium* genus-specific fragment of

cytochrome b gene (596 bp) with primer pair L1 and L2 (Tan et al. 1997) in all DNA samples to cross check the amplification failure by HAEMF and HAEMR2 primers. The reaction volumes and PCR conditions for L1 and L2 were as described by the Tan et al (1997). The amplified PCR products were visualized on 2% agarose gel. Before sequencing, the PCR products were treated using exonuclease I and shrimp alkaline phosphatase enzyme to remove the residual primer and dNTPs. The sequencing was carried out in both the directions using BigDye RR kit 3.1 (Applied Biosystems, Foster City, CA) in a 3500xL Genetic Analyzer (Applied Biosystems, Foster City, CA). The generated sequences were cleaned manually on Sequencing Software v5.1 (Applied Biosystems, Foster City, CA). The sequences were aligned using BioEdit Sequence Alignment Editor Software v7.2.5 (15) (Hall 1999). The PCR and DNA sequencing were carried out at Wildlife Institute of India, Dehradun. The sequences were BLAST searched on the MalAvi database for avian haemosporidian parasites (Zheng et al. 2000). A parasite sequence was defined as a unique lineage if a single nucleotide difference was found while comparing it with the sequences available at MalAvi database (Bensch et al., 2000; Hellgren, 2005; Chagas et al., 2017)

*Phylogenetic analysis*

The Bayesian evolutionary analysis of *Plasmodium* and *Haemoproteus* lineages was done using BEAST2 v2.6.2 (Bouckaert et al. 2019) to know the level of genetic relatedness among them. The host phylogenetic tree was constructed using the generated cytochrome b sequences. The phylogenetic relationships of the lineages, from the current study, at the global level were analyzed by obtaining the cytochrome *b* sequences of a total of 2862 *Plasmodium* and *Haemoproteus* lineages, of which 1287 lineages were of *Plasmodium* and 1575 lineages were of *Haemoproteus* from the MalAvi database for avian haemosporidian parasites, as of 24-05-2020, in FASTA format (Table S2A). Three lineages of Leucocytozoon (L_CLAHYE01, L_CLAHYE02, L_CLAHYE03) were used as outgroups. The HKY substitution model was selected with Gamma Category Count 4. The empirical frequencies were selected for getting a good fit to the data. The strict molecular clock model was selected for the data. The Yule Model of speciation was selected for the tree prior as it has been considered more appropriate when using different species sequences. The priors for rest parameters were set as default. The MCMC chain length was set to 6 million for this data set. The trace log and tree log frequencies were set to 1000 and the screen log frequencies were set to 10000. Assessment of adequate sampling and other of BEAST output parameters was done with TRACER (Rambaut et al. 2018). To obtain the phylogenetic tree, the time trees produced by the BEAST were summarized using TreeAnnotator program v1.7.2. (Drummond et al., 2012). The

TreeAnnotator burns the first 10% of the total 6000 trees generated by BEAST and the maximum clade credibility (MCC) tree with the highest posterior probability of all nodes was generated. The annotated tree was visualized on the FigTree program v1.4.4 (Rambaut 2018).

*Spatial phylogenetic analysis of Plasmodium and Haemoproteus lineages*

The Bayesian phylogeographic analysis of *Plasmodium* and *Haemoproteus* lineages was carried out to know the phylogenetic relationships with other known lineages of these two genera at the global level, if any, using Spherical Biogeography package with BEAST v2.6.2 (Bouckaert et al. 2016). For this, the cytochrome *b* sequences (n=54) of all lineages (excluding not having geographical coordinates information) of *Plasmodium* (n=1) and *Haemoproteus* (n=53) (Table S3), showing maximum similarity percentage with current study lineages, were used from the MalAvi database (obtained on 24-05-2020) for comparison (Table S4). The Geo Sphere package having a phylogeographic model was installed in the BEAST. The tip dates were set and the HKY substitution model was selected with empirical frequencies. A strict clock model was set. The Coalescent with Constant Population model was chosen for the tree prior. The information for latitude and longitude of each lineage was entered in the Spherical Geography model. The clock rate was set to 2e-5 for the data set the partition to speed up the convergence. For the geography partition, we selected a relaxed clock with a log-normal distribution. The MCMC chain length was set to 1 million. The screen log and file log were set to 10000 and 1000, respectively. The tree annotation was done as mentioned above, except the first 10% of the total 1000 trees were burned. The maximum clade credibility file was KML rendered using SpreaD3 (Spatial Phylogenetic Reconstruction of Evolutionary Dynamics) program v.0.9.7.1rc (Bielejec et al. 2016) for spatial map construction on Google Earth Pro. The DNA sequence similarity percentage was calculated through Multiple Sequence Alignment using MUSCLE (Edgar 2004).

*Species richness, the correlation between physical parameters with host and parasite distribution, if any*

Margalef Richness Index for Avian host species was calculated using the formula = (S - 1) / Log (n), where S = total number of species and n = total number of individuals in the sample. Further statistical analysis were carried out by taking the host species Groupwise i.e., Old world Flycatcher, Warblers, and Thrushes. The values of elevation for each sample were recorded by GPS as described by Kumar et al (2016). Other physical parameters viz. relative humidity, minimum, and maximum temperature for each collection site and date were obtained from NASA's meteorological data sets available from Data Access Viewer

(https://power.larc.nasa.gov/data-access-viewer/) (Table S1). The elevation range (minimum and maximum) for each avian host found infected with parasites was obtained from the IUCN Red List of Threatened Species (https://www.iucnredlist.org) (Table S1). The linear correlation, regression, and Karl Pearson's Correlation were carried out for physical parameters of the study area from where avian host samples were found infected with blood parasites using SPSS statistics software version 19 (IBM Corp 2010).

## Results

*Parasite lineages and their phylogeny analysis at world level*

Of the 111 host DNA samples, the cytochrome *b* gene's sequencing was done for 17 parasite positive samples out of which only 11 sample's (LM1, LM2, LM3, LM4, LM6, LM10, LM16, AS86, AS89, AS91, and AS102) sequences belonging to six species (*Phylloscopus Trochiloides, Cyornis poliogenys, Geokichla citrina, Cyornis hainanus dialilaemus, Cyornis rubeculoides* and *Eumyias thalassinus*) were usable (Table 1). The cytochrome *b* gene fragment of 596 base pairs was also amplified with L1 and L2 primers for the above 17 samples. The ten sequences representing five lineages were matched with *Haemoproteus* and one sequence with *Plasmodium* and thus a total of six lineages were found (Table 1). The naming of these six lineages was done as suggested by Bensch et al (2009), which were as follows H_MSP01, H_CYOPOL01, H_CHD01, H_CYORUB01, H_EUMTHA01, and P_GEOCIT01. The maximum clade credibility tree of six lineages has been given in Fig. 2. Out of 479 base pairs of cytochrome *b* gene the consensus sequences of 230 base pairs for the 6 lineages could be drawn after the alignment. The aligned consensus sequences of these 6 lineages were compared for their relationships with all available lineages of *Haemoproteus* and *Plasmodium* on the MalAvi database.

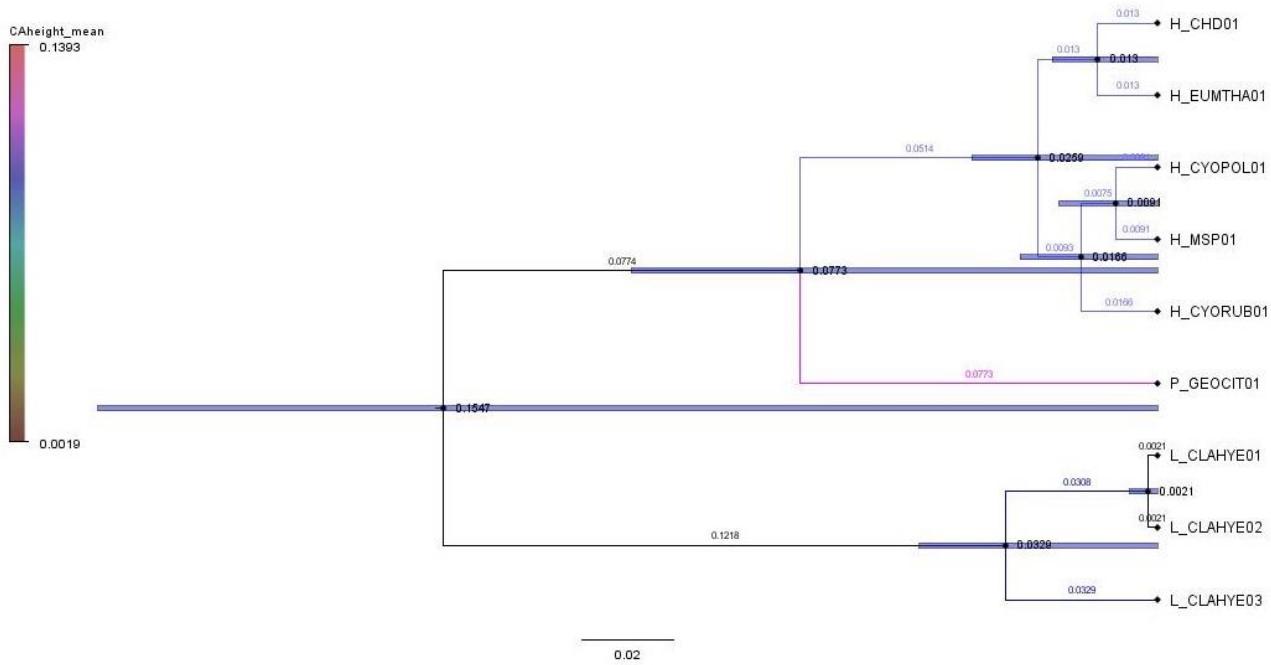


Figure 2. The maximum clade credibility tree showing six lineages.

The phylogenetic analysis at the global level was carried out by making a consensus sequence of 230 base pairs of cytochrome *b* sequences of a total of 2686 *Plasmodium* and *Haemoproteus* lineages, of which 1230 lineages were of *Plasmodium* and 1456 lineages were of *Haemoproteus* (Table S2B). The world phylogenetic analysis of five *Haemoproteus* lineages revealed that the MSP01 lineage clustered with COLL2 and PFC1 lineages, which belong to *Haemoproteus pallidus* (Fig. 3). The MSP01 shared 100% sequence similarity with COLL2 and PFC1 (Table S4). Lineages CYOPOL01, EUMTHA01, and CYORUB01 clustered in the clade containing the COLL2 and PFC1 lineages while CHD01 in a different clade (Fig. 3). The GEOCIT clustered in the *Plasmodium* clade having PADOM16 and shared maximum sequence similarity with it at a global level i.e., 93.89% (Fig. 3 and Table S4, respectively). Excluding MSP01, the rest five lineages matched 93% to 99% with the MalAvi database, whose details have been given in Table S4. So, all lineages in this study except MSP01 were unique. The short sequence of 230 base pairs do not hinder, here, in defining a new lineage as enough variations within this fragment, different from all known lineages available at MalAvi database, were found.

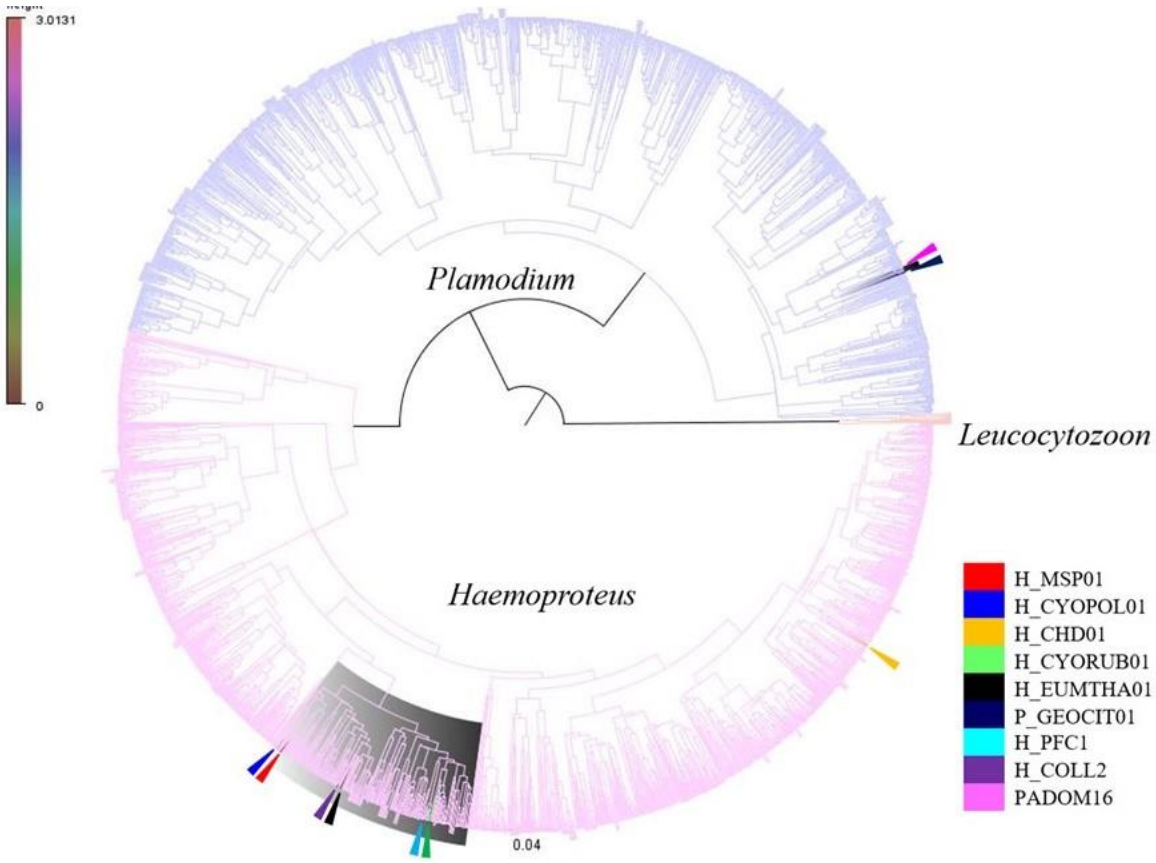


Figure 3. The maximum clade credibility tree showing World phylogeny of all *Plasmodium* and *Haemoproteus* lineages, including from this study.

*Parasite prevalence and host-parasite association*

Of the 111 DNA samples belonging to the eighteen host species, haemosporidian infection was found in 11 hosts belonging to the six species, which showed a 9.9% prevalence. *Haemoproteus* prevalence was found to be 9.0% across five host species (*Phylloscopus trochiloides, Cyornis poliogenys, C. hainanus dialilaemus, C. rubeculoides, Eumiyas thalassinus*) (Table 1). *Plasmodium* prevalence was 0.9% in one species of bird (*Geokichla citrina*). The *Haemoproteus* haplotype MSP01 was more common in warbler's species *Phylloscopus trochiloides* (N=4) than in two species of Old-World Flycatchers *C. poliogenys* and *C. rubeculoides* (N=2). The *Haemoproteus* lineage CYOPOL01, CHD01, CYORUB01, and EUMTHA01 were unique to *C. poliogenys, C. hainanus dialilaemus, C. rubeculoides, and Eumiyas thalassinus, respectively*. The *Plasmodium* lineage GEOCIT01 was unique to the species *Geokichla citrina*.

*Correlation between physical parameters with host and parasite distribution*

The Margalef Avian host richness Index was calculated to be 2.0852. The correlation between relative humidity and altitudinal height was found to be positive ($R^2$ = 0.77) (Table 2). The regression model predicted that the altitude and minimum temperature values showed $R^2$ = 0.77 and significant ($p < 0.003$) (Fig. 4). The correlation coefficients among different physical parameters are shown in Table 3. A significantly high negative correlation (87.8%) was observed between altitude and minimum temperature, suggesting that this parameter is directly related to the host's distribution range. Using Pearson's correlation analysis test, different physical parameters were associated with the occurrence of host species in different locations sampled. The 3D scatter plot summarizes all the above parameters collectively (Fig.5).

Table 2. Model Summary showing simple corelation and total variation (humidity and altitude and minimum temperature).

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | 0.878[a] | 0.771 | 0.714 | 2.97889 |

a. Predictors: (Constant), RHUM, Altitude

Table 3. Significant correlation (Carlpearson's) between different physical parameters from where blood samples of host species were collected.

Correlations

| | | Altitute | MINTEMP | RHUM |
|---|---|---|---|---|
| Altitute | Pearson Correlation | 1 | -.878** | .347 |
| | Sig. (2-tailed) | | .000 | .296 |
| | N | 11 | 11 | 11 |
| MINTEMP | Pearson Correlation | -.878** | 1 | -.282 |
| | Sig. (2-tailed) | .000 | | .401 |
| | N | 11 | 11 | 11 |
| RHUM | Pearson Correlation | .347 | -.282 | 1 |
| | Sig. (2-tailed) | .296 | .401 | |
| | N | 11 | 11 | 11 |

**. Correlation is significant at the 0.01 level (2-tailed).

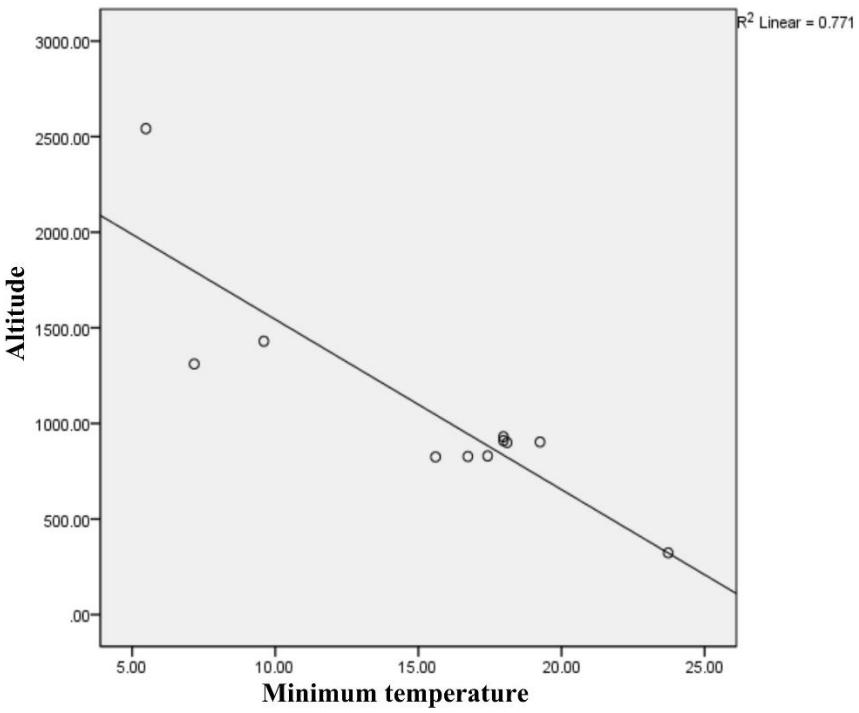


Figure 4. The regression model between altitude and minimum temperature.

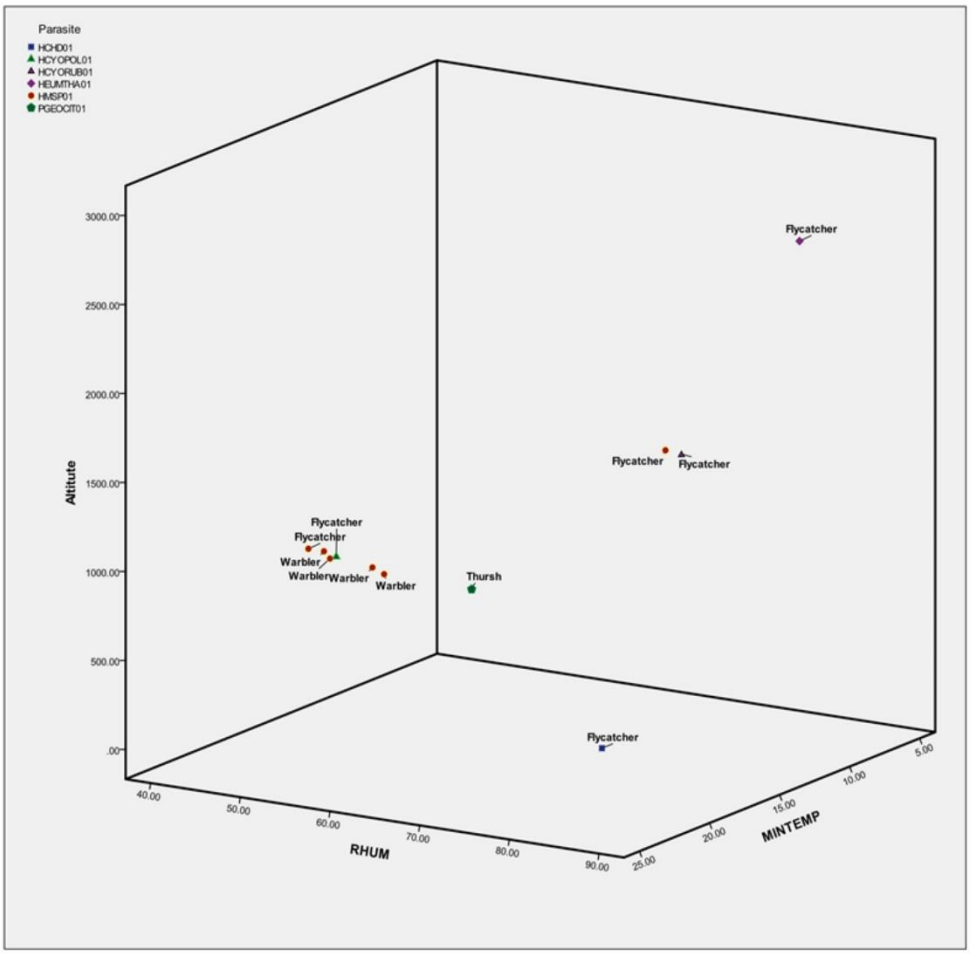


Figure 5. Distribution of host species in relation to altitude, minimum temperature, and relative humidity (RH).

*Spatial phylogenetics*

The spatial phylogeny map was constructed to understand the possible source parasite (*Haemoproteus* and *Plasmodium*) lineages and host species, causing infection to the bird species studied in three different parts of the India (Fig. 6). The chances of parasites being evolved in-situ or locally are equally likely, however, their maximum sequence similarity with lineages, anywhere in the world, makes them the candidates for a source population which has been well taken care off, here. The possible source populations have been shown connected through the colored tree branches on the globe and the polygon area represents the proportional lineage numbers maintained in that geographical area. The lineage MSP01 which was detected in host species *Phylloscopus trochiloides* (LM1, LM2, LM,3, LM4), *Cyornis poliogenys* (LM10) and *Cyornis rubeculoides* (AS89) belonging to the families of Phylloscopidae and Muscicapidae i.e. the groups of Warblers and Flycatchers, respectively, are connected through orange tree branches at places between Nagaland and Andhra Pradesh (Fig. 6 A). The *Phylloscopus trochiloides* are strongly migratory, which breeds in northeastern Europe, temperate to subtropical parts of Asia, and winters in India (Katti and Price 2003). *Cyornis rubeculoides are* native to India, Lao People's Democratic Republic, Myanmar, Thailand, and Viet Nam (BirdLife 2017) and *Cyornis poliogenys* are native to Bangladesh, Bhutan, China, India, Myanmar, and Nepal (BirdLife 2018). At the global level, the COLL2 and PFC1 infections were observed in 23 species belonging to the eight families of birds (Muscicapidae, Tyrannidae, Turdidae, Phylloscopidae, Parulidae, Locustellidae, Sylviidae, Ptilonorhynchidae) in the group of Flycatchers, Thrushes, and Warblers and more than 60 percent species out of them were found to be migratory (MalAvi database and present study). So, *Cyornis poliogenys* and *Cyornis rubeculoides* may have acquired the *Haemoproteus* infections from species like *Phylloscopus trochiloides* or other which are migratory (more detail in the discussion section) and frequently visiting India from the northern hemisphere. The spatial map shows possible source populations of hosts having COLL2 and PFC1 infections, connected through orange tree branches, in different parts of the world (Africa, Europe, North America, Malaysia, and Philippines) to India. The lineages CYOPOL01, EUMTHA01, CYORUB01, and CHD01 have been shown connected to maximum similarity *Haemoproteus* lineages through blue color tree branches (Fig. 6A, B, C). The *Plasmodium* lineage GEOCIT01 is connected with PADOM16, through a red branch (Fig. 6A), which was found in *Passer domesticus* from Egypt; however, their DNA sequence similarity is incredibly low i.e., 93.89 % (Table S4).

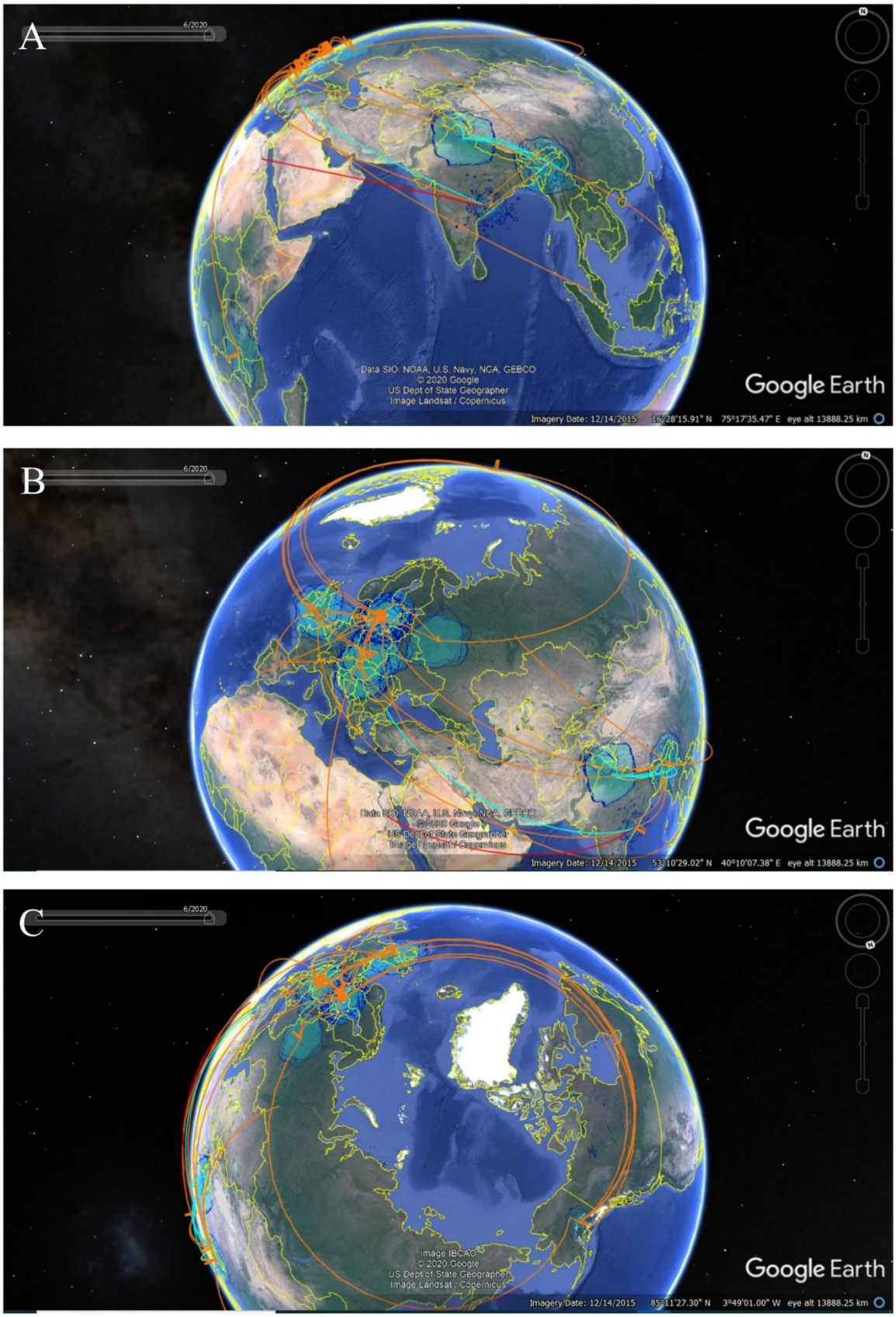


Figure 6. The spatial phylogeny map showing possible source *Haemoproteus* and *Plasmodium* lineages and host species which may be causing infection to the bird species studied in the current study.

## Discussion

The MSP01 lineage of *Haemoproteus* was detected in *Cyornis poliogenys* (Pale-chinned Flycatcher), *Cyornis rubeculoides* (Blue-throated Blue Flycatcher), and *Phylloscopus trochiloides* (Greenish Warbler) in India. The *C. poliogenys and C. rubeculoides* have been collected from Lanka pakala, Andhra Pradesh and Jotsama, Nagaland, respectively, and both these species are residents of Northeast India (BirdLife 2017 and 2018). *P. trochiloides* is migratory (Clement, 2015) and is found breeding in Afghanistan, Austria, Belarus, Bhutan, China, Czechia, Estonia, Finland, Germany, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Nepal, Poland, Russian Federation, Slovakia, Sweden, Tajikistan and Ukraine (BirdLife 2017). The MSP01 sequence is 100% (386 bp) similar to COLL2 lineage of *Haemoproteus*, so it may be possible that *P. trochiloides* had got infection, initially, from hosts having COLL2 infection during migration. Other migratory host species with COLL2 infection from Asia, Africa, Europe, and North America vising Indian parts may also be the source of MSP01 infection in the above three species. The phylogeography maximum clade credibility tree shows different host species with COLL2 infection in Asia, Africa, Europe and North America connected to MSP01 detected *C. poliogenys, C. rubeculoides* and *P. trochiloides* in Northeast and Andhra Pradesh parts of India through orange color clades (Fig. 6A).

The comparison of MSP01 cytochrome *b* sequence (387bp) with sequences of morphologically identified *Haemoproteus* species available at the MalAvi database revealed 100% similarity with *Haemoproteus pallidus* (Table S4). PFC1 was also a lineage of *H. pallidus* and MSP01 sequence was also matched with PFCI in the present study. Therefore, hosts having PFC1 infection in different parts of the world may also be the source of *H. pallidus* infection in Indian hosts showing MSP01 infection. However, we could not examine the blood of host species for morphological identification of blood parasites because samples were collected on Whatman filter paper.

The MalAvi database revealed that HCOLL2 lineage is primarily prevalent in 58% of Flycatchers belonging to the two families (Muscicapidae and Tyrannidae), 30% Thrushes family Turdidae, and 12% in Warblers belonging to the four families (Phylloscopidae, Parulidae, Locustellidae and Sylviidae) (Table S5). The PFC1 has been reported from Flycatchers belonging to the family of Muscicapidae (Table S5). The present study results are also in compliance with these as MSP01 is reported in birds belonging to Flycatchers and Warblers (Table S5 and Fig. 5). So, if there is any host specificity factor in these two *Haemoproteus* haplotypes is a topic of future research. It is an interesting fact that the HCOLL2 and PFC1 have only been found associates with Flycatchers, Thrushes, and Warblers in

literature, and as the sequence of MSP01 is matching completely with these two lineages, the fact goes same with the current study also. Hence, if there is any host specificity for these three lineages is a matter of future research. The information about the host species with *Haemoproteus* COLL2 and PFC1 infection in different parts of the world has been given in Table S6.

Four unique lineages CYOPOL01, CHD01, CYORUB01, and EUMTHA01 of *Haemoproteus* have been found in *Cyornis poliogenys, Cyornis hainanus dialilaemus, Cyornis rubeculoides,* and *Eumyias thalassinus*, respectively. The DNA sequences of CYOPOL01 is matching more than 99 percent with COLL2 and PFC1; however, it is very difficult to say that it belonged to *Haemoproteus pallidus* until a detailed morphological study is carried out. The same was the case with CYORUB01, of which DNA matched more than 98 percent with COLL2. The other *Haemoproteus* (CHD01 and EUMTHA01) and *Plasmodium* (PADOM16) lineages are connected with distantly related lineages in different parts of the world through other than orange tree branches in the spatial phylogeny map and may not indicate the source, host and parasite populations of infection. The record of EUMTHA01 in *Eumyias thalassinus* is the first of any *Haemoproteus* lineage from the Union Territory (UT) of Jammu and Kashmir. Therefore, a detailed study of blood parasites in Flycatcher and other bird species should be carried out in this UT region. It is very important to discuss here that while making a spatial phylogeny map, we have only used the maximum sequence similarity lineages from the MalAvi database, which had the geographical coordinate information (69.23 percent, Table S6). So, there are equal chances that the infection source may be related to the rest of the hosts and parasite populations and the spatial phylogeny map might be indicating the least sources of infections in host samples of the current study.

The low parasite prevalence seen in the present study might be attributed to multiple factors i.e., PCR failure, host-parasite association, vector distribution, host migration and low sample size. We could only amplify the *Plasmodium* genus-specific fragment of cytochrome b gene (596 bp) using primers L1 and L2 (Tan et al. 1997) in DNA samples, which also amplified 478 bp fragment using the primers HEAMF and HAEMR2. So, the DNA samples which could not amplify HEAMF and HAEMR2 primers suggested that those were negative for the presence of parasites and there was no PCR failure. The host species' distribution range was found to be directly associated with the parameters viz. altitude, minimum temperature, and relative humidity (Table 3 and Fig. 5). How the host-parasite association and host migration may affect the parasite prevalence have been explained above in para third and four, respectively. Though we have not collected the parameters responsible for vector distribution

in the present study, a lot of information already available in the literature can be utilized for this purpose. The geographic and climatic diversity in India influences the distribution of malaria vectors and parasite species (Das et al. 2012). So, malaria endemicity (in the human) is ten times or even more in eastern and southern parts than the extreme northern parts of India (Das et al. 2012). No detection of *Plasmodium* from the samples collected from Jammu and Kashmir may be due to the low sample size (n = 11) and the distribution status of *Anopheles* mosquito species. The records of *Anopheles* species between the altitudinal heights of 1550 to 3900 m in Jammu and Kashmir are very few over the last century (Adie and Adie 1912; Dar 2018). The vectors for avian Haemoproteidae belong to Hippoboscidae and Ceratopogonidae families representing hippoboscid flies and biting midges, respectively (Marzal et al. 2011; Valkiunas 2006). The records of both the hippoboscid flies (Rani et al. 2011) and biting midges have been found throughout India (Gosh et al. 2009). Therefore, we will suggest analysing all these factors in future studies.

In conclusion, the *Haemoproteus* lineage P_MSP01 was found to be more distributed in the samples of this study. Though, the lineage P_MSP01 was genetically closer to the *H. pallidus* generalist lineages H_COLL2 and PFC1 but morphological analysis is required for the final species confirmation. The *H. pallidus* is known to have a wide geographic distribution and host species range (Ortiz-Catedral et al., 2019) and our results were also supporting this. The physical factors (altitude, minimum temperature, and relative humidity) were associated with the host species' distribution range.

**Acknowledgments**

The current study was part of the project under UGC-Dr. D.S. Kothari Postdoctoral Fellowship Scheme, UGC awarded to the first author (BL/16-17/0151) from March 2017 to March 2020. We are thankful to National Institute of Malaria Research, Dwarka, New Delhi, for sparing the positive *Plasmodium* DNA sample for the standardization work. We acknowledge the support of the Director, Indian Institute of Integrative Medicine, CSIR, Jammu and Dr. Sumit Gandhi, Principal Scientist, for helping us to carry out the standardization work of this project. We sincerely thank Professor Trevor Price, University of Chicago for his generous support. We thank Forest department of Union Territory of Jammu & Kashmir, Himachal Pradesh, Uttarakhand, Sikkim, Arunachal Pradesh, Mizoram, Andhra Pradesh and Meghalaya for granting permission for sample collection.

## References


Adie, MRS. and Adie, JR. (1912). Note of an inquiry into malaria and mosquitoes in the Kashmir valley. The Indian Medical Gazette.1913. Page 341.

Atkinson, C.T., Van, Riper C. (1991). Pathogenicity and epizootiology of avian haematozoa: Plasmodium, Leucocytozoon, and Haemoproteus. In: Loye JE, Zuk M `(Eds.) Bird-parasite interactions: ecology, evolution, and behaviour. Oxford, Oxford University Press, pp. 19–48.

Beadell, J.S., Gering, E., Austin, J., Dumbacher John P., Peirce Mike, A. Thane K, Pratt Atkinson Carter, T. and Fleischer Robert C. (2004). Prevalence and differential host-specificity of two avian blood parasite genera in the Australo-Papuan region. Molecular Ecology.13, 3829–3844.

Bensch, S., Hellgren, O., Perez-Tris, J. (2009). MalAvi: a public database of malaria parasites and related haemosporidians in avian hosts based on mitochondrial cytochrome b lineages. Mol. Ecol. Resour. 2009; 9, 1353–1358. (doi:10.1111/j.1755-0998.2009.02692.x).

Bensch, S., Stjernman, M., Hasselquist, D. O. Ostman, B. Hansson, H. Westerdahl, and 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 of London B, Biological Sciences. 267, 1583–1589.

Bielejec Filip, Baele Guy, Vrancken Bram, Suchard Marc A, Rambaut Andrew and Lemey Philippe. (2016). SpreaD3: interactive visualisation of spatiotemporal history and trait evolutionary processes. Mol. Biol. Evol. Vol.33 Number 8, 2167—2169. Doi 10.1093/molbev/msw082.

BirdLife International. (2018). *Cyornis poliogenys*. *The IUCN Red List of Threatened Species*. e.T22709527A131953768. https://dx.doi.org/10.2305/IUCN.UK.2018-2.RLTS.T22709527A131953768.en. Downloaded on 25 July 2020.

BirdLife International. (2017). *Cyornis rubeculoides* (amended version of 2016 assessment). *The IUCN Red List of Threatened Species* 2017: e.T103761873A111163294. https://dx.doi.org/10.2305/IUCN.UK.2017-1.RLTS.T103761873A111163294.en. Downloaded on 25 July 2020.

Bouckaert, R., Vaughan, T. G., Barido-Sottani J., Duchêne, S., Fourment, M., Gavryushkina, A., Joseph Heled, Graham Jones, Denise Kühnert, Nicola De Maio, Michael Matschiner, Fábio K., Mendes, Nicola, F., Müller, Huw, A., Ogilvie, Louis du Plessis, Alex Popinga, Andrew Rambaut, David Rasmussen, Igor Siveroni, Marc A. Suchard, Chieh-Hsi Wu,Dong Xie,Chi Zhang,Tanja Stadler, Alexei, J. Drummond. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS computational biology. 15(4), e1006650.

Clement, P. Greenish Warbler (*Phylloscopus trochiloides*). (2015). In: del Hoyo, J., Elliott, A., Sargatal, J., Christie, D.A. and de Juana, E. (eds), *Handbook of the Birds of the World Alive*, Lynx Edicions, Barcelona.

Dar, T. H. (2018). Faunistic studies on the diversity and distribution of mosquitoes of the high altitude Himalayan Region Jammu and Kashmir. A Ph. D thesis submitted to the Pondicherry University. URI: http://hdl.handle.net/10603/284347.

Das Aparup, Anupkumar R Anvikara, Lauren J Catorb, Ramesh C Dhimana, Alex Eapenc, Neelima Mishraa, Bhupinder N Nagpala, Nutan Nandaa, Kamaraju Raghavendraa, Andrew F Readd, Surya K Sharmae, Om P Singha, Vineeta Singha, Photini Sinnisf, Harish C

Srivastavag, Steven A Sullivanh, Patrick L Suttonh, Matthew B Thomasb, Jane M Carltonh and Neena Valecha. (2012). Malaria in India: The Center for the Study of Complex Malaria in India (2012). Acta Trop. March; 121(3): 267–273. doi:10.1016/j.actatropica.2011.11.008.

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29: 1969–1973

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 32(5):1792-1797. PMID: 15034147. DOI: 10.1093/nar/gkh340

Ghosh Sikha, Majumdar Uttaran, Majumdar Abhijit and Chaudhuri Prasanta K. (2009). Biting flies of the genus *Homohelea* of India (Diptera: Ceratopogonidae). Folia Heyrovskyana, series A. vol. 16(4): 91-106, 2008, ed.: March 31, 2009. ISSN 1801-7142.

Gil-Vargas, D L, Sedano-Cruz R E. (2019). Genetic variation of avian malaria in the tropical Andes: a relationship with the spatial distribution of hosts. *Malar J.* 18, 129. https://doi.org/10.1186/s12936-019-2699-9

Gupta, P., Vishnudas, C.K., Ramakrishnan, U., Robin, V.V. and Dharmarajan, G. (2019). Geographical and host species barriers differentially affect generalist and specialist parasite community structure in a tropical skyisland archipelago. Proc. R. Soc. B. 286: 20190439. http://dx.doi.org/10.1098/rspb.2019.0439.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41:95-98.

Hellgren, O., Krizanauskiene, A., Valkiunas, G. and Bensch, S. (2007). Diversity and phylogeny of mitochondrial cytochrome b lineages from six morphospecies of avian Haemoproteus (Haemosporida, Haemoproteidae). Journal of Parasitology. 93, 889–896

IBM, Corp. (2010). IBM SPSS Statistics for Windows, Version 19.0, IBM Corp., Armonk, NY.

Ishtiaq Farah (2017). Exploring host and geographical shifts in transmission of haemosporidians in a Palaearctic passerine wintering in India. J Ornithol. 158:869–874. DOI 10.1007/s10336-017-1444-9.

Jia, T., Huang, X., Valkiūnas, G., Minghai Yang, Changming Zheng, Tianchun, Pu., Yanyun Zhang, Lu Dong, Xun Suo and Chenglin Zhang. (2018). Malaria parasites and related haemosporidians cause mortality in cranes: a study on the parasites diversity, prevalence and distribution in Beijing Zoo. *Malar J.* 17, 234. https://doi.org/10.1186/s12936-018-2385-3

Katti, M. and Price, T. D. (2003). Latitudinal trends in body size among over-winter- ing leaf warblers (genus Phylloscopus). Ecography. 26:69-79.

Kumar, R. S., Singh, P., Mohan, D., Gupta, S. K., Rana, S. K. and Singh, A. (2016). Causes of avian diversity gradients along the Himalayas. Wildlife Institute of India, Dehradun (DST Project Completion Report) Pp. 136.

Marzal A, Ricklefs RE, Valkiu¯ nas G, Albayrak T, Arriero E, Camille Bonneaud,Gábor A. Czirják,John Ewen,Olof Hellgren,Dita Hořáková,Tatjana A. Iezhova,Henrik Jensen,Asta Križanauskienė,Marcos R. Lima,Florentino de Lope,Eyðfinn Magnussen,Lynn B. Martin,Anders P. Møller,Vaidas Palinauskas,Péter L. Pap,Javier Pérez-Tris,Ravinder N. M. Sehgal,Manuel Soler,Eszter Szöllősi,Helena Westerdahl,Pavel Zetindjiev,Staffan Bensch (2011). Diversity, Loss, and Gain of Malaria Parasites in a Globally Invasive Bird. PLoS ONE. 6(7): e21905. doi:10.1371/journal.pone.0021905

Ortiz-Catedral, L., Brunton, D., Stidworthy, Hany M. Elsheikha, Tom Pennycott, Christoph Schulze, Michael Braun, Michael Wink, Helga Gerlach, Helene Pendl, Achim D. Gruber, John Ewen, Javier Pérez-Tris, Gediminas Valkiūnas & Philipp Olias. (2019). Haemoproteus minutus is highly virulent for Australasian and South American parrots. *Parasites Vectors.*12, 40. https://doi.org/10.1186/s13071-018-3255-0

Rambaut, A., Drummond, A.J., Xie, D., Baele, G. and Suchard, M.A. (2018). Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*. **syy032**. doi:10.1093/sysbio/syy032.

Rambaut, A (2018). FigTree v1.4.4, A Graphical Viewer of Phylogenetic Trees. Available from 〈http://tree.bio.ed.ac.uk/software/figtree/〉.

Rani Puteri, Azaziah Megat Abd., Coleman, Glen T., Irwin, Peter J. and Traub, Rebecca J. (2011). Hippobosca longipennis - a potential intermediate host of a species of Acanthocheilonema in dogs in northern India. Parasites & Vectors. 4:143.

Ricklefs, R.E. and Fallon, S.M. (2002). Diversification and host switching in avian malaria parasites. Proceedings of the Royal Society of London. Series B, Biological Sciences. 269, 885–892.

Singh, A., Gupta, S. K, Alström, P., Mohan D., Hooper, D. M., Kumar, R. S., Bhatt, D., Singh, P. and Price, T. D. (2020). Taxonomy of cryptic species in the Cyornis rubeculoides complex in the Indian subcontinent. Ibis. 162(3), pp.924-935. https://doi.org/10.1111/ibi.12735.

Singh, A., Kumar, A., Kumar, R. S., Bhatt, D. and Gupta, S. K. (2017). Amplification of mtDNA control region in opportunistically collected bird samples belonging to nine families of the order Passeriformes. Mitochondrial DNA Part B. 2(1), pp.99-100. https://doi.org/10.1080/23802359.2017.1289342

Tan, T. M. C., Nelson, J. S., H. C. Ng, R C Y Ting, and U A K Kara. (1997). Direct PCR amplification and sequence analysis of extra chromosomal P*lasmodium* DNA from dried blood spots. Acta Trop. 68:105–114.

Valkiunas, G. (2005). Avian malaria parasites and other haemosporidia. CRC Press, Boca Raton Florida. 932 p.

Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller. (2000). "A greedy algorithm for aligning DNA sequences", J Comput Biol. 7(1-2):203-14.