**Bird migration** **may connect parasite lineages but does not raise local prevalence and richness of avian haemosporidian parasites**

Daniela de Angeli Dutra¹\*, Antoine Filion¹, Alan Fecchio², Érika Martins Braga³, Robert Poulin¹

[danideangeli@live.com\*](mailto:danideangeli@live.com*) https://orcid.org/0000-0003-2341-2035

[afilion90@gmail.com](mailto:afilion90@gmail.com) https://orcid.org/0000-0003-1198-3017

[alanfecchio@gmail.com](mailto:alanfecchio@gmail.com) https://orcid.org/0000-0002-7319-0234

[embraga@icb.ufmg.br](mailto:embraga@icb.ufmg.br) <https://orcid.org/0000-0001-5550-7157>

[robert.poulin@otago.ac.nz](mailto:robert.poulin@otago.ac.nz) https://orcid.org/0000-0003-1390-1206

1.Department of Zoology, University of Otago, Dunedin, New Zealand

2.Programa de Pós-graduação em Ecologia e Conservação da Biodiversidade, Universidade Federal de Mato Grosso, Cuiabá, MT 78060-900, Brazil

3.Departamento de Parasitologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Brazil

**\*Correspondence:**

Daniela de Angeli Dutra

danideangeli@live.com

2. Methods

2.1 Dataset

All analyses were performed using a dataset comprising ~13200 bird blood samples accounting for 916 species from 63 different localities sampled from 2005 to 2018 in South America, previously described in Lacorte et al. 2013, Ferreira et al. 2017, Fecchio et al. 2019 and supplemented with new, previously unpublished data. In addition, haemosporidian lineages from the MalAvi database (<http://130.235.244.92/Malavi/>, Bensch et al. 2009) were included from South American regions (Figure 1, Supplementary material). Combining both datasets, we obtained a total of ~2800 sequenced parasites representing 668 distinct lineages collected from 506 different host species and 156 localities (all lineages belonging to one of these three genera: *Plasmodium*, *Haemoproteus* and *Leucocytozoon*). Each locality was assigned to a biome based on the classification of Turchetto-Zolet et al. 2013. The parasite prevalence per bird species and locality was estimated using PCR diagnostic protocols described by Hellgren et al. 2004 and Fallon et al. 2003. The parasite lineages were identified by the PCR protocol described by Hellgren et al. 2004. This protocol produces a *cyt b* fragment of 478 bp. The birds present in each locality were classified into three ecological classes: (1) resident; (2) partial migrant and (3) full migrant, according to the Brazilian Committee of Ornithology Records - CRBO 2014, Somenzari et al. 2018 and BirdLife International (<https://www.birdlife.org/>).

2.2 Potential correlates of prevalence and richness

*Spatial and temporal correlation*

All analyses were conducted in R (R Core Team, 2019). We determined whether there was significant spatial autocorrelation among localities for prevalence and parasite richness in our dataset by calculating the Moran Index value. In order to estimate this index, we combined the coordinates data into a matrix and employed the function “Moran.I” from the “Ape” package (Paradis and Schliep 2018). Temporal correlation analyses were performed using linear models, to determine whether prevalence or richness estimates varied throughout the sampling period (2005–2018). For parasite prevalence, we conducted a mixed linear model using the package “lme4” and the function “lmer” (Bates et al. 2015). Firstly, we grouped the data by year and location. Then, we compared the prevalence among years of collection considering number of birds collected and location as random variables. In order to test for a temporal correlation for parasite richness, we performed a simple linear model using the “lm” function.

*Phylogenetic Signal*

In order to estimate the phylogenetic signal among prevalence and richness estimates for the bird species in our dataset, we downloaded the file AllBirdsHackett1.tre from <https://birdtree.org/> website. Using the “treeman” package (Bennett et al. 2017), we created a treeman file containing all trees from the original file. Then, we randomly selected 100 trees to account for phylogenetic uncertainty. This new file was converted from treeman to a phylo file, from which we extracted one single random tree. We grouped our data per species and eliminated all bird species from the phylo tree which were not present in our dataset. Using the “match” function from the “picante” package (Kembel et al. 2010), we matched the species between the tree and our dataset. Then, we calculated Pagel’s lambda (λ) to evaluate the phylogenetic signal among bird species in our dataset, for both haemosporidian prevalence and parasite richness. Values of λ can range between 0 (no phylogenetic signal) and 1 (strong phylogenetic signal). In order to estimate lambda (λ), we applied the “phylosig” function from the “phytools” package (Revell 2012).

*Climate variables*

We used mean precipitation seasonality, and annual mean temperature (ºC) as predictors in the mixed models. We used R to extract these climate variables from the Worlclim database (<https://worldclim.org/version2>). Using the package “raster”, we extracted the data using the “getData” function, then we selected only the data from the 63 localities included in our original dataset since climate variables were applied only in mixed model analyses, for which the MalAvi data was not employed.

2.3 Statistical Analyses

The spatial and temporal autocorrelation analyses revealed there was no substantial effect of time or space on parasite richness, however, for prevalence, we observed a Moran Index effect of 0.15, and for this reason, locality was used as a random effect in our second mixed model to control for idiosyncratic characteristics of localities. Likewise, considerable phylogenetic signals were observed among bird species for prevalence (0.49) and parasite richness (0.17). Considering this, we analysed the prevalence using species as a fixed factor in the second mixed model.

*Bayesian model*

In order to determine whether migratory birds spread parasite lineages along their migratory routes and to evaluate the parasite connectivity among localities due to migratory behavior, we used multi-level modeling (MLM) with the “brms” package (Bürkner 2017) to evaluate the percentage of localities in which haemosporidian lineages occurred depending on whether they were found only in resident birds, only in partial migrant and fully migrant birds, or in both residents and migrants. We decided to use this approach as it allows us to statistically estimate the percentage of localities among which lineages are distributed according to their host status.

Firstly, we applied the “get\_priors” function to fit the priors for our model. We considered as independent and dependent variables bird migratory categories and percentage of localities in which each lineage was present, respectively, and lineages present only resident birds as reference category. We consider host richness and number of bird individuals infected by each lineage as fixed variables. As our Moran Index value for spatial autocorrelation of parasite richness among localities was low (-0.0008), we did not consider locality as a variable in our model and also did not use model correction for locality coordinates. Thus, we ran the model applying the “Beta” family, 4 chains with 2000 total iterations per chain and 1000 of warmup interactions. The model results were plotted using the “conditional\_effects” function to visualize the predictions of the population-level effects. We ran three models: one for all three parasite genera combined, one for *Plasmodium* lineages only, and one for *Haemoproteus* lineages only. Finally, we ran the “loo\_model\_weights“ to account for the effect of host richness (here, meaning the number of host species infected by each haemosporidian lineage) and number of hosts infected per lineage in our dataset.

*Mixed models*

Two mixed models were performed to estimate whether localities with more migratory birds have greater prevalence and richness of haemosporidian lineages. We employed the “glmer” function from the “lme4” package (Bates et al. 2015) applying Poisson and binomial distributions, respectively. In both models, we firstly created previous models including all variables that presented significant correlation with our dependent variable, and then selected the best model among them using “AIC” function in R. We tested the following variables as fixed factors: local host richness, local parasite richness, local prevalence across all birds sampled, local percentage of migratory species (based on birds caught), local number of migrant individuals, local temperature and precipitation.

In the first model, we considered parasite richness as the dependent variable and percentage of migratory bird individuals (i.e., percentage of migratory individuals out of all individual birds sampled in a locality) as the independent variable. According to our analyses, we employed local host richness (i.e., number of bird species sampled per locality), prevalence across all birds sampled, percentage of migratory species and number of migrant individuals as fixed variables. Biome was set as random variable. In this model, we did not use data from the MalAvi database, but only our dataset described above since it provides more information regarding the localities, such as prevalence data and host richness. We ran three models: one for all three parasite genera combined, one for *Plasmodium* lineages only, and one for *Haemoproteus* lineages only.

In the second model, we analysed the prevalence of infection in each bird species among localities. For this, we considered local prevalence in each bird species as our dependent variable and local percentage of migratory bird individuals as our independent variable. Following our previous analyses, only temperature was retained as a fixed factor. Further, we considered locality and biome as a random variables. In this model, we filtered our data in order to include only species with 10 or more bird individuals analysed. For this second model we again used only our dataset described above and excluded data from the MalAvi database, since the latter presents only positive and sequenced samples. Again, we ran three models: one for all three parasite genera combined, one for *Plasmodium* lineages only, and one for *Haemoproteus* lineages only.