Electronic Supplementary Material, Biology Letters

Genetic color polymorphism is associated with avian malaria infections

Laura Gangoso, Rafael Gutiérrez-López, Josué Martínez-de la Puente, Jordi Figuerola

- Methods: Extended methods on DNA extraction and blood parasites identification
- Methods: Extended statistical analyses. Random sampling approach and R code
- Results: Table S1: detailed information on the blood parasite lineages found in this study

Methods: DNA extraction and blood parasites identification

Genomic DNA was isolated from birds' blood samples using the semi-automatic Maxwell®16 LEV system Research kit (Promega, Madison, WI) (Gutiérrez-López *et al*. 2015). To determine the prevalence and linage identity of blood parasites, a fragment of the mitochondrial cytochrome b gene of *Plasmodium*, *Haemoproteus*, and *Leucocytozoon* parasites was amplified following Hellgren *et al*. (2004). This procedure is based on a first PCR using primers HaemNFI (5′-CATATATTAAGAGAAITATGGAG-3′) and HaemNR3 (5′-ATAGAAAGATAAGAAATACCATTC-3′). A nested PCR were used to amplify *Haemoproteus*/*Plasmodium* parasites with primers HaemF (5′-ATGGTGCTTTCGATATATGCATG-3′) and HaemR2 (5′-GCATTATCTGGATGTGATAATGGT-3′), while primers HaemFL (5′-ATGGTGTTTTAGATACTTACATT-3′) and HaemR2L (5′-

CATTATCTGGATGAGATAATGGIGC-3′) were used to amplify *Leucocytozoon* parasites.

The presence of amplicons was verified in 1.8% agarose gels. All negative samples in a first screening were repeated to minimize the presence of false negatives. Positive amplifications were sequenced using the Macrogen sequencing service (Macrogen Inc., The Netherlands). Sequences were edited using the Sequencher™ v 4.9 software (Gene Codes Corp., © 1991-2009, Ann Arbor, MI 48108). Parasite lineages were identified by comparing the obtained sequences with those deposited in GenBank database (National Center for Biotechnology Information, Blast 2008). Lineages were named according to MalAvi, a database for avian haemosporidian parasites (http://mbio-serv2.mbioekol.lu.se/Malavi/; Bensch *et al*. 2009). The sequence of the new lineage identified in this study (FALELE01, see table S1) was deposited in Genbank database (accession number: KX649994).

Methods: Statistical analyses

We analyzed 209 blood samples from 183 individuals: 151 pale morphs (91 females and 60 males) and 32 dark morphs (22 females and 10 males). Although our sample size is limited, especially with regard to dark individuals, Jovani and Tella (2006) showed that uncertainty about the real prevalence (i.e. population prevalence) rather than being a linear relationship, rapidly decreases as sample size increases up to 10–20 individuals, but not much more with increasing sample sizes. These authors stated that a sample size around 15 could be used to estimate prevalence as a reasonable trade-off between not loosing too much data from analyses and maintaining acceptable levels of uncertainty. In our case, the minimum sample size was 32 dark morph individuals. The dark morph naturally occurs at a low frequency (about 30%, see

Gangoso *et al*. 2011) and our sampling reflects quite well the actual population' frequencies. Therefore, we believe our sample sizes (32 dark and 151 pale individuals) are large enough as to detect significant patterns with acceptable levels of uncertainty.

Nineteen out of 183 individuals were captured in multiple years, but this variation could not be accounted for as a random term in GLMMs due to convergence failure. To overcome this difficulty and prevent pseudoreplication, the 19 individuals with multiple observations were randomly subsampled to only include one observation per individual in the dataset used to perform the GLMMs. Random sampling of multiple, while scarce, recaptures has been used in other studies facing similar problems of non-independence of data (e.g. Dunn *et al*. 2013, Jacobs *et al*. 2015, Lumpkin *et al*. 2014). Alternatively, other authors have used only the first observation when more than one sample is available for a few individuals (e.g. Knowles *et al*. 2014). In our case, we designed a resampling method that allows testing all possible combinations among resampled individuals. This procedure was repeated 1,000 times for each parasite genus. Hence, in each of the 1,000 GLMMs performed, a random combination of observations from these resampled individuals was used. Parameter estimates and standard errors from the resulting GLMMs were averaged and the range of *p*-values and percentage of models where each term was statistically significant calculated. We performed these analyses in R v3.0.2 (R Core Team 2015) with the package lme4 (Bates *et al*. 2015).

R code for resampling (Author: Duarte S. Viana)

```
# Create a new dataset with resampled individuals
datar<-data[duplicated(data$ID) | duplicated(data$ID, fromLast=TRUE),]
data0<-data[!(duplicated(data$ID) | duplicated(data$ID, fromLast=TRUE)),]
```

```
sp<-split(datar,datar$ID,drop=T)


# Random combinations

dfs<-list()
for(i in 1:1000){
 dfi<-datar[1,]
 for(j in 1:length(sp)){
        randn<-sample(1:nrow(sp[[j]]),1)
        dfi[j,]<-sp[[j]][randn,]
        }
        dfs[[i]]<-dfi
}

data.res<-list()
for(i in 1:length(dfs)) data.res[[i]]<-rbind(data0,dfs[[i]])

#-------------------------------------
#(Example) GLMMs for Plasmodium

#Notes

#pint = p-value of the interaction
#pmorph = p-value of the factor “morph type”
#psex = p-value of the factor “sex”
#pbc = p-value of the covariate “body condition index”
#cint = coefficients of the interaction, and so on for the rest of explanatory variables
 included in our models

mod.res<-
 data.frame(pint=numeric(length(data.res)),pmorph=numeric(length(data.res)),psex=nu
 meric(length(data.res)),pbc=numeric(length(data.res)),c0=numeric(length(data.res)),cm
 orph=numeric(length(data.res)),csex=numeric(length(data.res)),cBC=numeric(length(d
 ata.res)),cint=numeric(length(data.res)))

for(i in 1:length(data.res)){
 data.res.14<-data.res[[i]][!is.na(data.res[[i]]$BC),]
 plasmodium0i<- glmer(plasmodium~ morph*sex +BC+ +(1|year), family=binomial,
 data=data.res[[i]])
 plasmodium1i<- glmer(plasmodium~ morph+sex +BC+ +(1|year), family=binomial,
 data=data.res[[i]])
 plasmodium2i<- glmer(plasmodium~ sex +BC+ +(1|year), family=binomial,
 data=data.res[[i]])
 plasmodium3i<- glmer(plasmodium~ morph +BC+ +(1|year), family=binomial,
 data=data.res[[i]])
 plasmodium0i.14<- glmer(plasmodium~ morph*sex +BC+ +(1|year), family=binomial,
 data=data.res.14)
 plasmodium4i<- glmer(plasmodium~ morph*sex +(1|year), family=binomial,
 data=data.res.14)
```

```
t1i<-anova(plasmodium0i,plasmodium1i,test="Chisq")$"Pr(>Chisq)"[2]
t2i<-anova(plasmodium1i,plasmodium2i,test="Chisq")$"Pr(>Chisq)"[2]
t3i<-anova(plasmodium1i,plasmodium3i,test="Chisq")$"Pr(>Chisq)"[2]
t4i<-anova(plasmodium0i.14,plasmodium4i,test="Chisq")$"Pr(>Chisq)"[2]
coefs<-summary(plasmodium0i)$coefficients[,1]
mod.res[i,]<-c(t1i,t2i,t3i,t4i,coefs)
}

#Range of p-values

range(mod.res$pint)
range(mod.res$pmorph)
range(mod.res$psex)
range(mod.res$pbc)


# % of models where the term is statistically significant

length(mod.res$pint[mod.res$pint<0.05])/length(mod.res$pint)
length(mod.res$pmorph[mod.res$pmorph<0.05])/length(mod.res$pmorph)
length(mod.res$psex[mod.res$psex<0.05])/length(mod.res$psex)
length(mod.res$pbc[mod.res$pbc<0.05])/length(mod.res$pbc)
apply(mod.res[,5:9],2,mean)
apply(mod.res[,5:9],2,function(x) sd(x)/sqrt(nrow(mod.res)))
```

References

Bates D, Maechler M, Bolker B, Walker S. 2015 Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw*. **67**, 1–48.

Bensch S, Hellgren O, Pérez-Tris J. 2009 MalAvi: a public database of malaria parasites and related haemosporidians in avian hosts based on mitochondrial cytochrome b lineages. *Mol. Ecol. Resour*. **9**, 1353–1358.

Dunn JC, Goodman SJ, Benton TG, Hamer KC. 2013 Avian blood parasite infection during the non-breeding season: an overlooked issue in declining populations? *BMC Ecology*. **13**, 30.

Gangoso L, Grande JM, Ducrest AL, Figuerola J, Bortolotti GR, Andrés JA, Roulin A. 2011 MC1R-dependent, melanin-based colour polymorphism is associated with cell-mediated response in the Eleonora's falcon. *J. Evol. Biol* **24**, 2055–2063.

Gutiérrez-López R, Martínez-de la Puente J, Gangoso L, Soriguer RC, Figuerola J. 2015 Comparison of manual and semi-automatic DNA extraction protocols for the

barcoding characterization of hematophagous louse flies (Diptera: Hippoboscidae). *J. Vector Ecol*. **40**, 11–15.

Hellgren O, Waldenström J, Bensch S. 2004 A new PCR assay for simultaneous studies of *Leucocytozoon*, *Plasmodium*, and *Haemoproteus* from avian blood. *J. Parasitol*. **90**, 797–802.

Jacobs AC, Fair JM, Zuk M. 2015 Parasite infection, but not immune response, influences paternity in western bluebirds. *Behav. Ecol. Sociobiol*. **69**, 193.

Jovani R, Tella JL. 2006 Parasite prevalence and sample size: misconceptions and solutions. *Trends Parasitol.* **22**, 214-218.

Knowles SC, Wood MJ, Alves R, Sheldon BC. 2014 Dispersal in a patchy landscape reveals contrasting determinants of infection in a wild avian malaria system. *J. Anim. Ecol.* **83**, 429-439.

Lumpkin DC, Murphy TG, Tarvin KA. 2014 Blood parasite infection differentially relates to carotenoid-based plumage and bill color in the American goldfinch. *Ecol. Evol*. **4**, 3210–3217.

R Core Team. 2015 R: *A language and environment for statistical computing. R Foundation for Statistical Computing* Vienna, Austria: R Foundation for Statistical Computing.