

# THE ROLE OF POPULATION STRUCTURE IN PATHOGEN DIVERSITY IN WILD BAT POPULATIONS

TIM LUCAS. SEPTEMBER 23, 2015

## CONTENTS

1. Abstract	2
1.0.1. One or two sentences providing a basic introduction to the field	2
1.0.2. Two to three sentences of more detailed background	2
1.0.3. One sentence clearly stating the general problem (the gap)	2
1.0.4. One sentence summarising the main result	2
1.0.5. Two or three sentences explaining what the main result reveals in direct comparison to what was thought to be the case previously	2
1.0.6. One or two sentences to put the results into a more general context.	2
1.0.7. Two or three sentences to provide a broader perspective,	2
2. Introduction	3
3. Methods	4
4. Results	7
5. Discussion	8
References	8

## 1. ABSTRACT

- 1.0.1. *One or two sentences providing a basic introduction to the field. Write abstract here.*
- 1.0.2. *Two to three sentences of more detailed background.*
- 1.0.3. *One sentence clearly stating the general problem (the gap).*
- 1.0.4. *One sentence summarising the main result.*
- 1.0.5. *Two or three sentences explaining what the main result reveals in direct comparison to what was thought to be the case previously.*
- 1.0.6. *One or two sentences to put the results into a more general context.*
- 1.0.7. *Two or three sentences to provide a broader perspective,*

## 2. INTRODUCTION

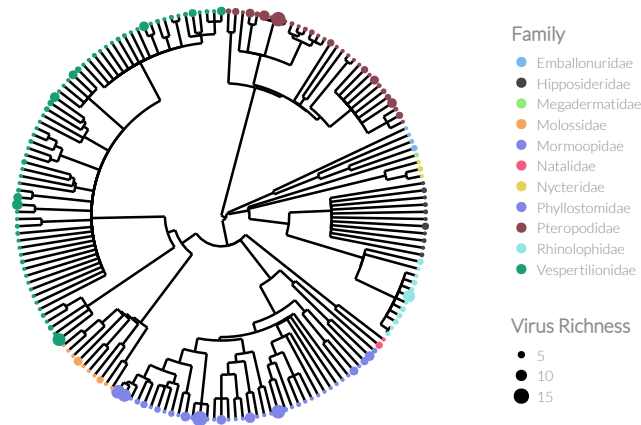


FIGURE 1. Pruned phylogeny with dot size showing number of pathogens and colour showing family.

### 3. METHODS

To measure pathogen richness I used data from [1]. These simply include known infections of a bat species with a pathogen species. Only species with at least one pathogen were included in the analysis. To control for study bias I collected the number of pubmed and scholar citations for each bat species including synonyms from ITIS [2] via the taxize package [3]. The counts were scraped using the rvest package [4].

I used two measures of population structure.  $F_{ST}$  and the number of subspecies. The number of subspecies was counted using the Wilson and Reeder taxonomy [5].

```
## Error in mget(words, envir, "any", NA, inherits = TRUE): cannot open file '/home/tim/Dropbox/phd/Documents/thesis/.Ch3Cache/w
No such file or directory
## Error in eval(expr, envir, enclos): cannot open file '/home/tim/Dropbox/phd/Documents/thesis/.Ch3Cache/w
No such file or directory
## Error in eval(expr, envir, enclos): cannot open file '/home/tim/Dropbox/phd/Documents/thesis/.Ch3Cache/w
No such file or directory
```

Measures of body mass are taken from Pantheria [6]. They are log transformed due to the strong right skew.

To control for phylogenetic nonindependence I used the best-supported phylogeny from [7] which is the supertree from [8] with names updated to match the Wilson & Reeder taxonomy [5]. Phylogenetic manipulation was performed using the ape package [9].

I wanted to run three models using the phylolm package testing the relationship between pathogen richness and log number of subspecies. I tried phylogenetically controlled, multivariate GLMs with poisson errors and identity links. This model was fitted both with and without an interaction term between number of subspecies and study effort. I also fitted a phylogenetically controlled, GLM with poisson errors and identity link to pathogen richness and study effort. The residuals from this model was then used as the response variable in a multivariate GLM. However, with these models the numerical optimisation failed to converge.

I ran three models using the caper package [10] testing the relationship between pathogen richness and log number of subspecies. All independent variables were log transformed — study effort was  $\log(\text{citations} + 1)$ . I ran phylogenetically controlled, multivariate linear models. This model was fitted both with and without an interaction term between number of subspecies and study effort.

We also fitted a phylogenetically controlled, GLM with poisson errors and identity link to pathogen richness and study effort. The residuals from this model was then used as the response variable in a multivariate GLM.

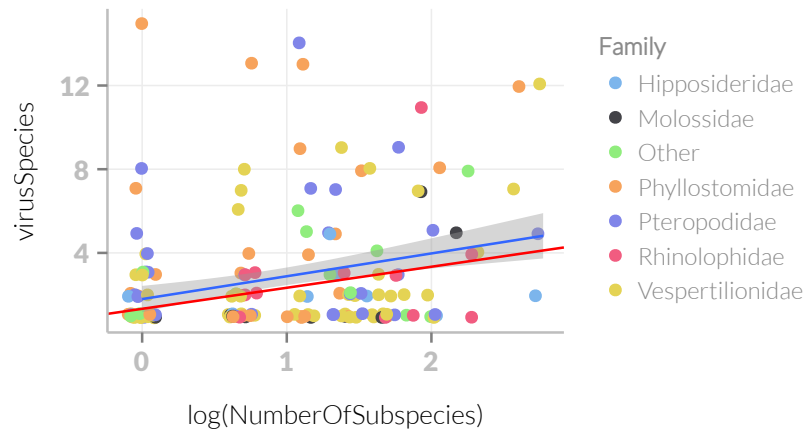


FIGURE 2. Number of virus species against log number of subspecies. Nonphylogenetic trend line in blue. Phylogenetic model (evaluated at mean body mass and mean study effort values) is shown in red.

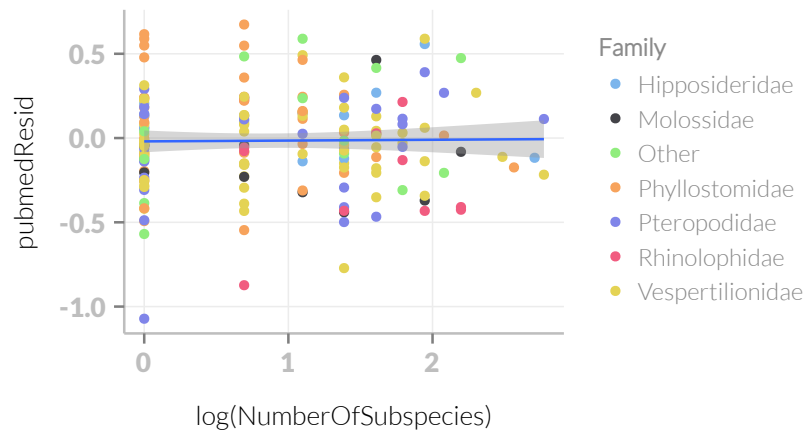


FIGURE 3. Plot using residuals from number of viruses against number of citations (study effort). Nonphylogenetic trend line added.

These are the fits from models fitted with the species with only 1 virus species removed. We can see that if number of subspecies is unlogged, it remains marginally significant. If subspecies is logged it is no longer significant.

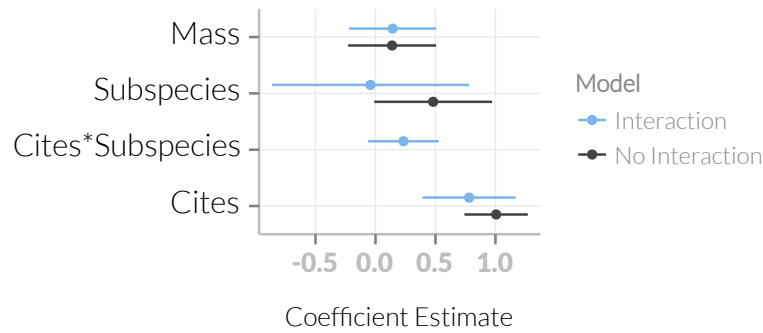


FIGURE 4. Plot of coefficient estimates and 95% confidence intervals for phylogenetic model with (inter) and without (joint) interactions between study effort and number of subspecies. Without interactions, number of subspecies is marginally significant.

#### 4. RESULTS

See Figure 4 for a display of estimated coefficients for the two models using number of viruses as the response variable. The main model with mass, study effort and number of subspecies as predictors found study effort to be highly significant ( $\beta = 1.01$ ,  $p = 3.9 \times 10^{-12}$ ). The number of subspecies was marginally significant ( $\beta = 0.48$ ,  $p = 0.06$ ). The effect of nonindependence due to phylogeny was very small ( $\lambda = 0.07$ ,  $p = 0.1$ ).

The interaction term between study effort and number of subspecies, when included, was not significant ( $\beta = 0.23$ ,  $p = 0.12$ ).

The model using the residuals from a regression between number of viruses and study effort as the response variable found no significant affect of number of subspecies. Mass was marginally significant.

## 5. DISCUSSION

## REFERENCES

- Luis, A. D. et al. (2013). “A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special?” In: *Proceedings of the Royal Society B: Biological Sciences* 280.1756, p. 20122753.
- Integrated Taxonomic Information System (ITIS)*. <http://www.itis.gov>.
- Chamberlain, S. A. and E. Szöcs (2013). “taxize: taxonomic search and retrieval in R”. In: *F1000Research* 2.
- Wickham, H. (2015). *rvest: Easily Harvest (Scrape) Web Pages*. R package version 0.2.0. URL: <http://CRAN.R-project.org/package=rvest>.
- Wilson, D. E. and D. M. Reeder (2005). *Mammal species of the world: a taxonomic and geographic reference*. Vol. 12. JHU Press.
- Jones, K. E. et al. (2009). “PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals: Ecological Archives E090-184”. In: *Ecology* 90.9, pp. 2648–2648.
- Fritz, S. A., O. R. Bininda-Emonds, and A. Purvis (2009). “Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics”. In: *Ecology letters* 12.6, pp. 538–549.
- Bininda-Emonds, O. R. et al. (2007). “The delayed rise of present-day mammals”. In: *Nature* 446.7135, pp. 507–512.
- Paradis, E., J. Claude, and K. Strimmer (2004). “APE: analyses of phylogenetics and evolution in R language”. In: *Bioinformatics* 20, pp. 289–290.
- Orme, D. et al. (2012). *caper: Comparative Analyses of Phylogenetics and Evolution in R*. R package version 0.5. URL: <http://CRAN.R-project.org/package=caper>.