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

TIM LUCAS. SEPTEMBER 29, 2015

1. Abstract

1.0.1. *One or two sentences providing a basic introduction to the field.* Bats are an important reservoire of zoonotic diseases. It is still unclear what factors determine the number of pathogens a wild bat species carries. But once understood, these factors could provide a way to priorities surveillance of bat populations.

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,*

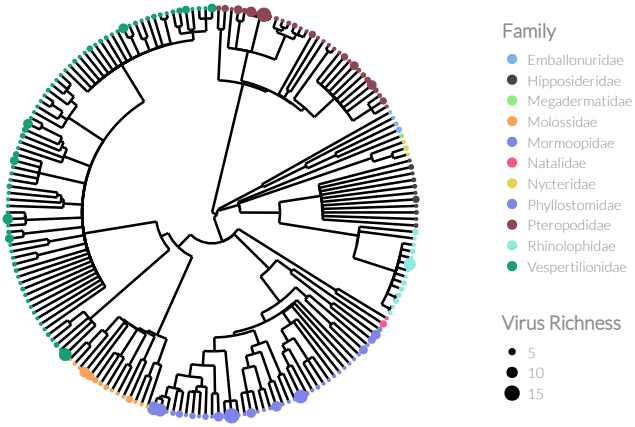

Figure 1. Pruned phylogeny with dot size showing number of pathogens and colour showing family.

2. Introduction

3. Methods

To measure pathogen richness I used data from (Luis et al. 2013). 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. As many viruses were not identified to species level, I counted a virus if it was the only virus, for that host species, in the lowest identified taxonomic level. That is, if a host carries an unknown Paramyxoviridae virus, then it must carry at least one Paramyxoviridae virus. If a host carries an unknown Paramyxoviridae virus and a known Paramyxoviridae virus, then it is hard to confirm that the unknown virus is not another record of the known virus. In this case, this would be counted as one virus species.

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 (Wilson and Reeder 2005). F_{ST} and other measures were collated from the literature. Studies are from a wide range of spatial scales, from local ($\sim 10\,\text{km}$) to continental. As F_{ST} inevitably increases with spatial scale I controlled for this by only using data from studies where a large proportion of the species range was studied. I used the ratio of the furthest distance between F_{ST} samples to the width of the species range and only used studies if this ratio was greater than 0.3. To allow comparison between different measures (F_{ST}, ϕ_{ST}) and data from different molecular regions I converted all data to diploid gene flow. WILL ADD EXTRA METHODS LATER. These two measures of population structure were analysed separately as the number of subspecies has 196 data points while there is only F_{ST} data for ~ 30 bat species.

To control for study bias I collected the number of Pubmed and Google Scholar citations for each bat species including synonyms from ITIS (*Integrated Taxonomic Information System (ITIS)* n.d.) via the taxize package (Chamberlain and Szöcs 2013). The counts were scraped using the rvest package (Wickham 2015). I log transformed these variables as they were strongly right skewed. The log number of citations on Pubmed and Google scholar were highly correlated (pgls: $t = 13.23$, df $= 194$, $p = 0$). The results here are for analyses using only Google Scholar citations. See the appendix for analyses run using Pubmed citations.

Measures of body mass are taken from Pantheria (Jones et al. 2009). They are log transformed due to the strong right skew.

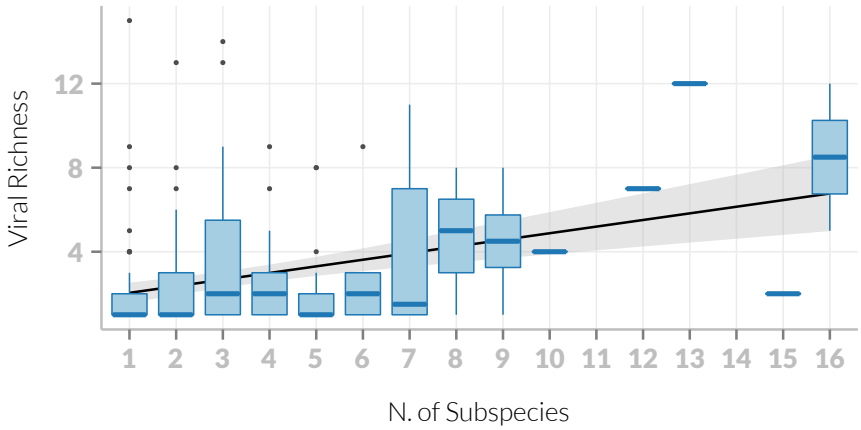

Figure 2. Number of virus species against number of subspecies. Data within a number of subspecies are plotted as boxplots with the dark bar showing the median, the box showing the interquartile range, vertical lines showing the range and outliers shown as seperate points. A non-phylogenetic linear model is shown in blue

To control for phylogenetic nonindependance I used the best-supported phylogeny from (Fritz et al. 2009) which is the supertree from (Bininda-Emonds et al. 2007) with names updated to match the Wilson & Reeder taxonomy (Wilson and Reeder 2005). Phylogenetic manipulation was performed using the ape package (Paradis et al. 2004).

I ran two models using the caper package (Orme et al. 2012) testing the relationship between pathogen richness and log number of subspecies. All independant variables were log transformed. 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.

Plots were created with a combination of (Wickham 2009; Lucas 2015; Solt and Hu 2015; Yu 2015)

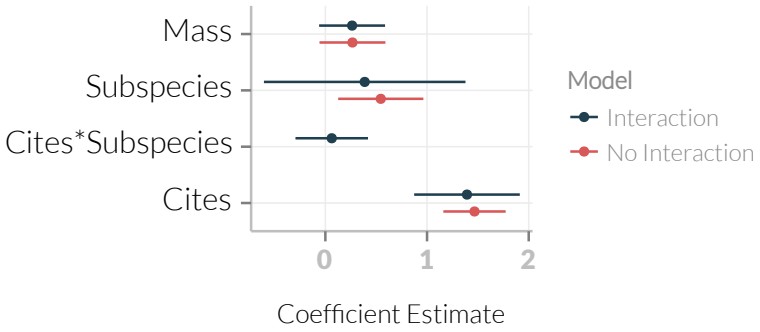

Figure 3. Plot of coefficient estimates and 95% confidence intervals for phylogenetic model with (inter) and without (joint) interactions between study effort and number of subspecies. Without interactions, number of subspecies is marginally significant.

4. Results

See Figure 3 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 (β =NA, p =0). The number of subspecies was marginally significant (β = 0.54, p =0.01). The effect of nonindependance due to phylogeny was very small (λ =0.07, p =0.1).

The interaction term between study effort and number of subspecies, when included, was not significant (β = NA, p =0.73).

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.

Wilson, D. E. and D. M. Reeder (2005). *Mammal species of the world: a taxonomic and geographic reference*. Vol. 12. JHU Press.

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`.

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`.

Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer New York. isbn: 978-0-387-98140-6. url: `http://had.co.nz/ggplot2/book`.

Lucas, T. (2015). *palettetown: Use Pokemon Inspired Colour Palettes*. R package version 0.1.0. url: `http://CRAN.R-project.org/package=palettetown`.

Solt, F. and Y. Hu (2015). *dotwhisker: Dot-and-Whisker Plots of Regression Coefficients from Tidy Data Frames*. Available at The Comprehensive R Archive Network (CRAN). url: `http://CRAN.R-project.org/package=dotwhisker`.

Yu, G. (2015). "ggtree: an R package for versatile annotation and visualization of phylogenetic tree". In: *in prepare*.