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

TIM LUCAS. JANUARY 6, 2016

1. Abstract

1.0.1. *One or two sentences providing a basic introduction to the field.* It is still unclear what factors determine the number of pathogens a wild species carries. But once understood, these factors could provide a way to prioritise surveillance of wild populations for zoonotic disease and make predictions as to how pathogen richness will respond to global change.

1.0.2. *Two to three sentences of more detailed background.* The pattern of contacts between individuals (i.e. population structure) has long been known to strongly affect epidemic processes. Theory suggests that population structure can promote pathogen richness while the ecological literature generally assumes it will decrease richness.

1.0.3. *One sentence clearly stating the general problem (the gap).* Previous studies in wild bat populations have had contradictory results and the different measures of population structure have different shortcomings. There is therefore a need for more robust comparative studies testing for a relationship between pathogen richness and population structure.

1.0.4. *One sentence summarising the main result.* Here I used comparative data to test whether population structure influences pathogen richness using bats as a case study as they have been associated with a number of important, recent zoonotic outbreaks. Unlike previous studies I used two measures of population structure: a novel measure, number of subspecies, and a more careful application of genetic measures which have been used previously. I find that both of these measures are positively associated with pathogen richness and are probably in the best supported model.

1.0.5. *Two or three sentences explaining what the main result reveals in direct comparison to what was thoughts to be the case previously.* My results add more robust support to the hypothesis that pathogen structure promotes pathogen richness in bats and lends clarity to the contradictory results previously published. The results support predictions from theory. They contradict the assumption commonly made in the ecological literature that factors that increase R_0 should increase pathogen richness implying that competitive processes amongst pathogens are stronger than previously thought.

1.0.6. *One or two sentences to put the results into a more general context.* Although my analysis implies that population structure does promote pathogen richness in bats, the weakness of the relationship and the difficulty in obtaining some measurements means this is probably not a usefully predictive factor on it's own for optimising zoonotic surveillance. However, the relationship has implications for global change, implying that increased habitat fragmentation might promote greater viral richness in bats.

2. Introduction

2.0.7. *General Intro.* The number of pathogen species carried by a host species has important consequences for the ecology of the host and the probability that the host will be a reservoir of a zoonotic pathogen. However, the factors that affect pathogen richness are poorly understood.

2.0.8. *Specific Intro.*

2.0.9. *Theoretical background.* Single pathogen models show that increasing population structure simply slows disease spread and makes establishment less likely (Colizza and Vespignani 2007; Vespignani 2008). In the ecological literature this is often taken as predicting that increased population structure will decrease pathogen richness (Nunn et al. 2003; Morand 2000). However, models of competition between multiple pathogens show that in unstructured populations a competitive exclusion process occurs but that splitting the population into two patches allows coexistence (Qiu et al. 2013; Allen et al. 2004; Nunes et al. 2006).

2.0.10. *Previous Studies.* Three studies have used comparative data to test for an association between population structure and viral richness. A study on 15 African bats found a positive relationship between distribution fragmentation and viral richness (Maganga et al. 2014) while a study on 20 South-East Asian bats found the opposite relationship (Gay et al. 2014). A global study on 33 bats found a positive relationship between F_{ST} — a measure of genetic structure — and viral richness (Turmelle and Olival 2009). However, this study included measures using mtDNA which only measures female dispersal which may have biased the results many bat species show female philopatry (Kerth et al. 2002; Hulva et al. 2010). Furthermore, this study used measures of F_{ST} irrespective of the study scale with studies covering from tens (McCracken and Bradbury 1981) to thousands (Petit and Mayer 1999) of kilometers. As isolation by distance has been shown in a number of bat species (Burland et al. 1999; Hulva et al. 2010; O'Donnell et al. 2015; Vonhof et al. 2015) this could bias results further. Finally, when a global F_{ST} value is not given they used the mean of all pairwise F_{ST} between sites. This is not correct as from global F_{ST} we expect migration rates of $M = (1 - F_{ST})/8F_{ST}$ while from F_{ST} between pairs of populations we expect migration rates of $M = (1 - F_{ST})/16F_{ST}$ where M is the absolute number of diploid inviduals dispersing per generation (Slatkin 1995). To use studies that only present pairwise F_{ST} values the raw data would have to be gathered and global F_{ST} calculated from those. As it is in fact the movement of individuals that is epidemiologically relavent, using these studies is probably not correct without attempting to correct for these difference. Studies on single pathogens have also shown that space can allow persistence where a well mixed population ould experience a single, large epidemic followed by pathogen extinction (Blackwood et al. 2013; Pons-Salort et al. 2014; Plowright et al. 2011).

2.0.11. *Choice of measure of population structure.* A number of measurments of population structure have been used in the literature and each has it's own shortcomings. In particular, the better, more direct measurements tend to be very work intensive which consequently means data is available for few species.

2.0.12. *Direct dispersal measurements.* The ideal metric of population structure is direct measurement of dispersal rates and distance. These are incredibly difficult to obtain, especially over large scales. Due to white nose syndrom, some very large mark-recapture studies have been conducted, but recapture rates are low (). Further, these large studies have been in species that live in a few large colonies, so recapture rates should be higher than in less social species. In practise, direct measurements are not practical for comparative analysis due to the lack of data and inconsistency in data collection methods.

2.0.13. *Genetic measures.* As direct measurement of dispersal are difficult, genetic data is often used. Measurement such as F_{ST} are used to calculate migration. There are strong model assumptions under the conversion from F_{ST} to migration. However, the main issue with this measure is the effort required for each study and the subsequent low number of measurements. Further, there are differences in the scales of the studies and the genetic regions being sequenced. This differences should not be ignored.

2.0.14. *Number of Subspecies.* For a population to evolve distinct phenotypic or genetic traits, such that they can be classed as a subspecies, there must be limited migration between populations. The number of subspecies a species has therefore reflects the level of population structure in that species. The value of this measurement is available for every bat species. However, it is likely biased, with well studied species being likely to have more recognised subspecies. Further, this is a very course measure and it is important to consider whether it is measuring migration at a timescale and rate that is epidemiologically relevant.

For both measures from F_{ST} and the number of subspecies it is useful to consider the rates of animal movement that are being measured. Rates of migration estimated from F_{ST} tend to be between 1 and 100 individuals per generation dispersing across all subpopulations.

2.0.15. *Measures from range.* The final measurement that has been used is derived from the shape of the species' range, typically from IUCN (IUCN 2010) maps. The ratio between the perimeter of the range and the area (or similar values) are calculated. Range maps are very course for many species. Furthermore there is a potential bias with island living species being given sea based edges where continental species might be assumed to live everywhere in between locations where it is known to live, without considering the different terrestrial habitats in these areas.

2.0.16. *The gap.* There is a lack of studies using multiple measures of population structure and larger datasets to robustly estimate the importance of population structure.

2.0.17. *What I did/found.* Here I have used two measures of population structure — the number of subspecies and gene flow — to robustly test for an association between population structure and pathogen richness in bats. Furthermore, I used a much larger dataset for one of these analyses, further promoting robustness of results. I found that both measures of population structure are positively associated with viral richness and are likely to be in the best models for describing viral richness. Further, I found that the role of phylogeny is very weak in the models and in the distribution of viral richness amongst taxa.

3. Methods

To test for an association between pathogen richness I have performed multivariate regression using a model selection framework to establish whether or not two measures of pathogen richness are likely to be in a 'best model' and therefore important. As species cannot be considered independant due to shared evolutionary history, phylogeny was controlled for in all regressions. A number of other factors that have previously been found to be important were included as additional independant variables: body mass (Kamiya et al. 2014; Turmelle and Olival 2009; Gay et al. 2014; Maganga et al. 2014), range size (Kamiya et al. 2014; Turmelle and Olival

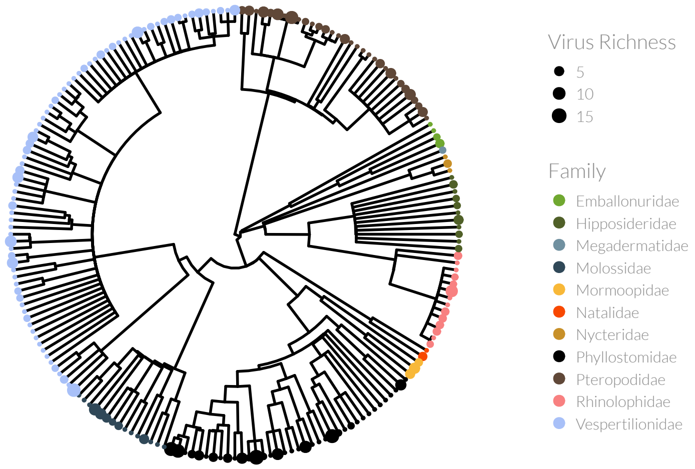

Figure 1. Pruned phylogeny with dot size showing number of pathogens and colour showing family.

2009; Maganga et al. 2014) and study effort (Turmelle and Olival 2009; Gay et al. 2014; Maganga et al. 2014). This was to attempt to avoid spurious positive results occuring simply due to correlation between pathogen richness and a different, causal factor. Despite commonly bein associated with pathogen richness (Arneberg 2002; Kamiya et al. 2014; Nunn et al. 2003), population density is not included in the analysis as there is very little data for bat densities — however Chapter 4 examines the relationship between density and population structure and Chapter 5 presents a method that allows the estimation of density from acoustic surveys. I used both the number of subspecies a bat species has and estimates of gene flow (analysed separately) as measures of population structure. All analyses were run in R (R Development Core Team 2010)

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. Rows with host species that were not identified to species level were removed. Many viruses were not identified to species level or their specified species names were not in the ICTV virus taxonomy (ICTV 2014). I counted a virus if it was the only virus, for that host species, in the lowest taxonomic level (present in the ICTV taxonomy) identified. 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. Gene flow and the number of subspecies. The number of subspecies was counted using the Wilson and Reeder taxonomy (Wilson and Reeder 2005). Gene flow is calculated from estimates of F_{ST} collated from the literature. Studies are from a wide range of spatial scales, from local ($\sim$ 10 km) to continental. As F_{ST} often increases with spatial scale (Burland et al. 1999; Hulva et al. 2010; O'Donnell et al. 2015; Vonhof et al. 2015) 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 (measured with `http://www.distancefromto.net/` if not stated) to the width of the IUCN species range (IUCN 2010) and only used studies if this ratio was greater than 0.2. This is an arbitrary value that was a comprimise between retaining a reasonable number of data points and controlling for the bias in spatial scale. I converted all F_{ST} value to migration using $M = \frac{1-F_{ST}}{8F_{ST}}$. This removes the $(0,1)$ bounds of F_{ST} and is more easily interpretable though the results are unaffected. 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 22 bat species. For the subspecies analysis all bat species in Luis et al. (2013) were used (i.e. all species with at least one known virus species). However, for the gene flow analysis, all bat species with suitable F_{ST} estimates were used. As this included some species not present in Luis et al. (2013) this includes some bat species with zero known virus species.

To control for study bias I collected the number of Pubmed and Google Scholar citations for each bat species name 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 = 19.32, df = 194, p = 0). As this correlation is strong, the results here are for analyses using only Google Scholar citations.

Measures of body mass are taken from Pantheria (Jones, Bielby, et al. 2009) and primary literature (Canals et al. 2005; Arita 1993; López-Baucells et al. 2014; Orr and Zuk 2013; Lim and Engstrom 2001; Aldridge 1987; Ma et al. 2003; Owen et al. 2003; Henderson and Broders 2008; Heaney et al. 2012; Oleksy et al. 2015; Zhang et al. 2009). *Pipistrellus pygmaeus* was assigned the same mass as *P. pipistrellus* as they indistinguishable by mass. Body mass measurements were log transformed as they were strongly right skewed. Distribution size was estimated by downloading range maps for all species from IUCN (IUCN 2010) and were also logged due to right skew.

To control for phylogenetic nonindependance I used the best-supported phylogeny from Fritz et al. (2009) (shown in Figure 1) 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). The importance of the phylogeny on each variable separately (the λ parameter of the variable regressed against an intercept) was estimated and tested against the null of $\lambda = 0$ with log-likelihood ratio tests using caper (D. Orme et al. 2012). I also performed the analysis using the tree from (Jones, Bininda-Emonds, et al. 2005) as this has some broad changes with families in different places. However the phylogeny did not affect the analysis.

3.1. **Statistical analysis.** Statistical analysis for both dependant variables was conducted using a information theoretical/model averaging approach (Burnham and Anderson 2002) specifically following (Whittingham, Swetnam, et al. 2005; Whittingham, Stephens, et al. 2006). I chose a credible set of models including all combinations of independent variables and a model with just an intercept. In the analysis using the number of subspecies dependant variable I also included an interaction term between study effort and number of subspecies. This interaction was included as I believed *a priori* that this interaction may be present as subspecies in well studied species are more likely to be identified. The interaction was only included in models with both study effort and number of subspecies as individual terms.

I fitted phylogenetic regressions of all models using nlme (Pinheiro et al. 2015). The independant variables were centered and scaled to allow direct comparison of the coefficients (Schielzeth 2010). In each case I simultaneously fitted the λ

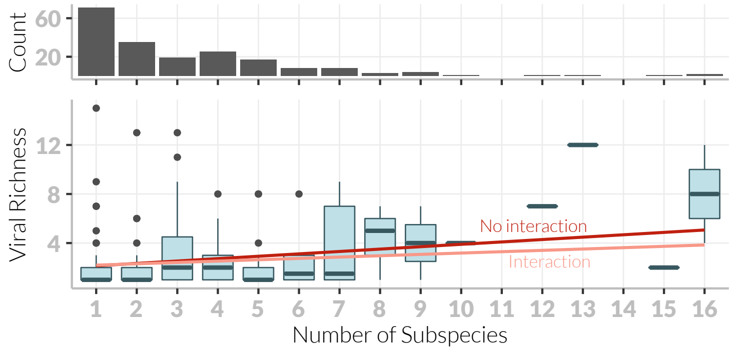

Figure 2. Number of virus species against number of subspecies. The top panel shows the distribution of the data, with most species having few subspecies. Data within a number of subspecies are plotted as boxplots with the dark bar showing the median, the box showing the interquartile range, vertical lines showing the range and outliers shown as seperate points. Regression lines are from multivariate phylogenetic models with all other independant variables set at their median value. The models shown are those with (pink) and without (red) an interaction between study effort and number of subspecies.

parameter as this avoids mispecifying the model (Revell 2010). κ and δ parameters were constrained to one as they are more concerned with when along a branch evolution occurs and because fitting multiple parameters makes interpretation difficult.

To establish the importance of variables I calculated the probability, Pr, that each variable would be in the best model if the data were recollected. For each variable the mean of it's coefficient, β, in all models that contained that variable was also calculated to determine the direction and strength of the variables. In the subspecies analysis, this mean of β was calculated for all models, only models with a interaction term and only models without an interaction. As the interaction term greatly affects the estimated value of β, considering these value seperately aids interpretation. Following (Whittingham, Swetnam, et al. 2005) I included a uniformally random variable as a null variable as even unimportant variables can have Akaiki weights notably greater than zero. The whole analysis was run 50 times, resampling the random variable each time. I calculated AICc for each model. I calculated the average AICc, $\bar{\text{AICc}}$, by averaging AICc scores within models. ΔAICc was calculated as $\min(\bar{\text{AICc}}) - \bar{\text{AICc}}$, not the mean of the individual ΔAICc scores, to guarantee that the best model has $\Delta\text{AICc} = 0$. From these ΔAICc I calculated Akaiki weights, w. This value can be interpreted as the probability that a model would be the best model if the data were recollected. For each variable, the sum of the Akaiki weights of models containing that variable are summed to give Pr. This value can be interpreted as the probability that the given variable is in the best model.

4. Results

4.1. **Number of Subspecies.**

4.1.1. *More descriptive.* After data cleaning there was data for 196 bat species in 11 families. There appears to be a positive relationship between the number of subspecies and viral richness (Figure 2) though few species have more than four subspecies. The number of described virus species for a bat host ranged up to 15 viruses in *Carollia perspicillata*. Figure 1 shows the phylogeny used and the

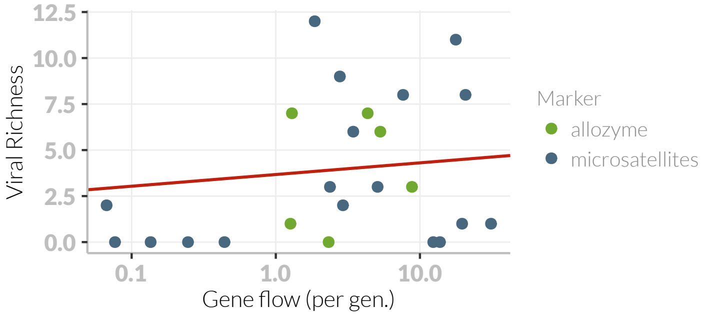

Figure 3. Gene flow per generation (on a log scale) against viral richness. The genetic marker used is shown with colour. The red line shows the univariate phylogenetic model.

number of viruses for each species. The mean number of viruses across families is fairly constant with a lower range of 1.67 for Nycteridae. The highest mean is Mormoopidae with 5 virus species per bat species, but this is based on a sample size of 3. The Phyllostomidae have the second highest mean of 3.49 (n = 37).

The small change in mean pathogen richness across families and the lack of clear pattern in Figure 1 implies that viral richness is not strongly phylogenetic. This is corroborated by the small estimated size of λ (λ = 0.04, p = 0.21). This fact implies that other factors must control pathogen richness. It also implies that pathogens are not directly inherited down the phylogeny, although this is to be expected by the fast evolution of viruses.

Of the explanatory variables, the number of subspecies has no phylogenetic autocorrelation ($\lambda = 10^{-6}$, p = 1), study effort and distribution size have weak but significant autocorrelation (Study Effort: λ = 0.1, $p = 9.12 \times 10^{-3}$, Distribution size: λ = 0.46, $p = 2.82 \times 10^{-9}$) and mass is strongly phylogenetic (λ = 0.93, p = 0). The parameter λ is fitted and governs how important the phylogeny is in the model. Across all models the mean value of λ is 0.08 implying the residuals from the model are weakly phylogenetic. A small number of models (0.4%) had negatively phylogenetically distributed residuals.

4.1.2. *Model results.* The top seven models all have ΔAICc < 4 meaning there is no clear best model (Table 2). These top seven models have a combined weight of 0.96 meaning that there is a 0.96% chance that one of these models would be the best model if the data was recollected. However these top seven models all contain study effort, number of subspecies and the interaction between these two variables. log(Mass) and log(Range Size) and the random variable are all in three of the top seven models.

Summing the Akaiki weights of all models that contain a given variable gives a probability, Pr, that the variable would be in the best model (Figure 4A) if the data were recollected (Whittingham, Stephens, et al. 2006). The number of subspecies is very likely in the best model ($Pr > 0.99$) as is the interaction between number of species and study effort ($Pr = 0.96$) compared to the benchmark random variable which has $Pr = 0.21$ (see Figure 4A and Table 1). When models with the interaction term are removed, on average (mean weighted by Akaiki weights) there is a positive relationship between the number of subspecies and viral richness ($\beta = 0.63$, variance = 0.02). Models with an interaction between number of subspecies and study effort have a positive interaction term ($\beta = 0.5$, variance = 5.11×10^{-5}) and linear term ($\beta = 0.31$, variance = 2.13×10^{-4}). This supports the hypothesis that

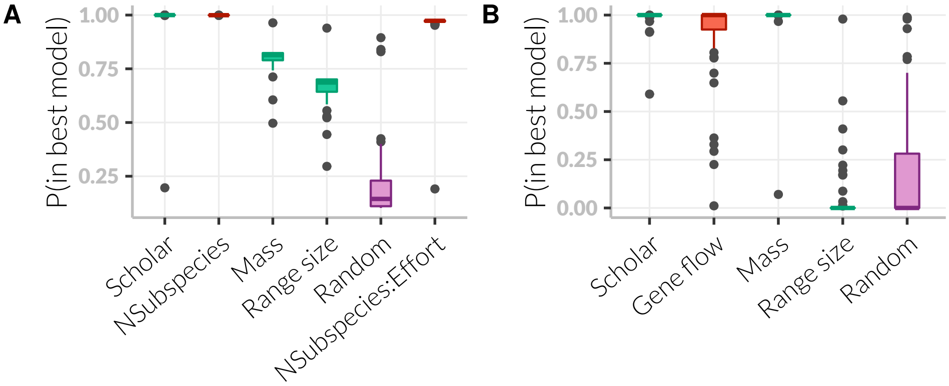

Figure 4. Akaika variable weights for both analyses. The probability that each variable will be in the best model if the data were recollected is shown for each of the bootstrap analyses. The purple "Random" box is a uniform random variable used as a null. Population structure (Number of subspecies and Gene flow), shown in red, is likely to be in the best model in both analyses.

population structure promotes pathogen richness. The strong support for a positive interaction term implies that population structure has a stronger relationship with known pathogen richness in the presence of study effort. only predicts high known pathogen richness in the presence of high study effort. One interpretation of this is that population structure alone does not predict high known richness; reasonable study effort is also needed to turn the expected high richness into known and recorded viral richness. Another interpretation is that having few subspecies does not predict low viral richness unless the species has been suitable sampley as the low number of subspecies is probably due to a lack of study rather than an accurate measurement.

As seen in Figure 4A, study effort is very likely in the best model ($\beta = 0.99$, $Pr > 0.98$). Body mass and range size are also probably in the best model ($\beta = 0.48$, $Pr = 0.8$ and $\beta = 0.35$, $Pr = 0.66$) with positive relationships of slightly lower strength than the number of subspecies in models without an interaction term ($\beta = 0.63$, variance = 0.02).

4.2. **Gene Flow.**

4.2.1. *More Descriptive.* Due to the low number of studies and the restrictive requirements imposed on study design, there are only data for 24 bat species in 7 families. The number of described virus species for a bat host ranged up to 12 viruses in *Miniopterus schreibersii*.

As with the Number of Subspecies dataset, there is no phylogenetic signal in the number of virus species (λ = 10^{-6}, p = 1) Gene flow also has no phylogenetic autocorrelation (λ = 10^{-6}, p = 1). Due to the low sample size, significance tests are unlikely to have much power. However, study effort has some phylogenetic autocorrelation (λ = 0.15, p = 0.56) while distribution size and mass seem to show phylogenetic signal (Distribution size: λ = 0.67, p = 0.53, Mass: λ = 0.79, p = 2.69×10^{-3}).

4.2.2. *Model results.* Only the model with study effort, gene flow and mass is well supported with the second model having an ΔAICc of 34 (Table 2). While less strongly supported than the number of subspecies, gene flow is likely in the best model ($Pr = 0.89$) compared to the benchmark random variable which has $Pr = 0.18$ (Figure 4B and Table 1). On average (mean weighted by Akaiki weights)

Table 1. Estimated variable weights (probability that a variable is in the best model) and their estimated coefficients for both number of subspecies and gene flow analyses. The coefficients for number of subspecies are also seperated for models with and without the interaction term because this term strongly changes the coefficient. However, there are no weights for these seperated terms as they are not directly compared in the model selection framework.

	Number of Subspecies		*Gene flow*	
Variable	*Pr*	Coefficient	*Pr*	Coefficient
Number of subspecies				
Total	1.00	0.32		
Models without interaction term		0.63		
Models with interaction term		0.31		
Number of subspecies:log(Scholar)	0.96	0.50		
Gene flow			0.89	-0.67
log(Scholar)	0.98	0.99	0.99	2.49
log(Mass)	0.80	0.48	0.98	-0.35
log(Range size)	0.66	0.35	0.06	1.57
Random	0.21	0.05	0.18	0.23

there is a negative relationship between gene flow and viral richness ($\beta = -0.67$, variance = 5.48×10^{-3}) despite the apparent positive relationship (see Figure 3) weakly suggested by the bivariate model (pgls: β = 0.63, t = 1.16, df = 13, p = 0.27). This supports the hypothesis that population structure promotes viral richness. Possibly due to the smaller sample size, or a weaker relationship, this coefficient is much more varied than the number of subspecies coeficient with 21.75% of models estimating a positive relationship.

As in the number of subspecies analysis, study effort is very likely in the best model ($Pr = 0.99$) as is body mass ($Pr = 0.98$). However, body mass has a negative average coefficient ($\beta = -0.35$, variance = 0.04) which is in contrast to number of subspecies analysis, many studies in the literature (Kamiya et al. 2014; Turmelle and Olival 2009; Gay et al. 2014; Maganga et al. 2014) and the bivariate model (pgls: β = 0.6, t = 0.38, df = 13, p = 0.71). In contrast to the number of subspecies analysis, range size is almost certainly not in the best model with $Pr = 0.06$ being much less than the random variable. This variable being less supported than the random variable is probably because range size is closely correlated with study effort (pgls: β = 0.6, t = 4.45, df = 13, p = 6.58×10^{-4}).

Across all models the mean value of λ is -1.64 and a large number of individual models (58%) had negatively phylogenetically distributed residuals implying the residuals from the model are strongly negatively phylogenetic. Due to the small sample size this is probably due to a small number of data points with large residuals being distant on the tree.

5. Discussion

5.0.3. *Restate results.* I have tested the hypothesis that population structure promotes pathogen richness in bats. By analysing data on two measures of population structure, and using larger datasets than previous studies, it is hoped that any conclusions may be more robust than the conflicting results in the literature (Gay et al. 2014; Turmelle and Olival 2009; Maganga et al. 2014). I have found that a positive affect of population structure (a positive effect of the number of subspecies

Table 2. Model selection results for number of subspecies and gene flow analysis. $\bar{\text{AICc}}$ is the mean AICc score across 50 resamplings of the null random variable. ΔAICc is the $\bar{\text{AICc}}$ score minus the lowest score. w is the Akaike weight and can be interpreted as the probability that the model is the best model (of those in the plausible set). $\sum w$ is the cumulative sum of the Akaike weights. log(Scholar)*NSubspecies implies the interaction term between study effort and number of subspecies as well as both of the individual linear terms. In the number of subspecies analysis there are many models with low ΔAICc scores suggesting there there is no single 'best model'. In the gene flow analysis, only the top model is supported.

Model	$\bar{\text{AICc}}$	ΔAICc	w	$\sum w$
Number of Subspecies				
log(Scholar)*NSubspecies + log(Mass) + log(RangeSize)	882	0.00	0.38	0.38
log(Scholar)*NSubspecies + log(Mass)	884	1.39	0.19	0.57
log(Scholar)*NSubspecies + rand + log(Mass)	885	2.24	0.12	0.70
log(Scholar)*NSubspecies	885	3.14	0.08	0.78
log(Scholar)*NSubspecies + log(RangeSize)	886	3.18	0.08	0.86
log(Scholar)*NSubspecies + rand + log(RangeSize)	886	3.94	0.05	0.91
log(Scholar)*NSubspecies + rand	886	3.95	0.05	0.96
log(Scholar) + NSubspecies log(Mass) + rand	889	6.93	0.01	0.97
log(Scholar) + NSubspecies + log(Mass) + log(RangeSize) + rand	890	7.80	0.01	0.98
Gene flow				
log(Scholar) + log(Gene flow) + log(Mass)	71	0.00	1.00	1.00
log(Range size)	105	34.09	0.00	1.00
log(Mass)	106	35.06	0.00	1.00

and a negative effect of gene flow) are likely to be in the best models for explaining viral richness. Study effort is also clearly supported confirming the expectation that additional study of a bat species yields more known viruses infecting that species and highlighting again that this bias cannot be ignored in studies using known pathogen richness as a proxy for total pathogen richness.

5.0.4. *Weaknesses and limitations.* Although I have used measures of study effort to try to control for biases in the viral richness data, this bias could still make the results here unreliable — this is especially true as study effort is by far the strongest predictor of viral richness in both datasets. It is hoped that as untargeted sequencing of viral genetic material (e.g. Anthony et al. (2013)) becomes cheaper and more common this bias can be reduced. The strength of the relationship between study effort and known viral richness also highlights the number of virus species and bat-virus host-pathogen relationships yet to be discovered.

I have used two measures of pathogen richness and the number of subspecies dataset is larger than those used in previous studies. However it is clear that the gene flow dataset is small (n = 24). Furthermore, while the model averaging approach has given a negative model averaged coefficient, the univariate model of gene flow against viral richness gave a positive coefficient. It is not easy to interpret these contradictions but it is clear that the results from the gene flow analysis alone should not be considered strong evidence for a relationship between pathogen richness and population structure. The sensitivity of this analysis reiterates the need to use large datasets where possible and use mutliple measures of population structure to promote robust conclusions.

5.0.5. *Broader context of results.* The results here suggest that there is a positive relationship between population structure and pathogen richness in bats. This is in agreement with (Maganga et al. 2014; Turmelle and Olival 2009) but in disagreement with (Gay et al. 2014). Furthermore it contradicts the assumption that factors that promote high R_0 will automatically promote high pathogen richness (Nunn et al. 2003; Morand 2000).

This relationship implies that direct or indirect competitive mechanisms are acting such that population structure is needed in order to allow escape from competition.

The relationship between population structure and pathogen richness suggests that population structure has a least some potential as being predictive of high pathogen richness and therefore of a species likelihood of being a reservoir of a potentially zoonotic pathogen. However given that it is difficult to measure population structure and given that the relationship appears to be weak at best, this trait on it's own is unlikely to be useful in predicting zoonotic risk. However, as a number of other factors are also associated with pathogen richness (body mass and to a lesser extent range size here but also other traits elsewhere), using a combination traits in a predictive (i.e. machine learning) framework has potential to be used in prioritising zoonotic disease surveillance. The main hurdle in this approach is finding a way to validate models — due to the study effort bias in current data, predictive models will also be biased.

The relationship between pathogen richness and population structure also has implications for habitat fragmentation and range shifts due to global change. In short habitat fragmentation and range shifts that reduce movement between populations would be predicted to increase pathogen richness. However, depending on the mechanisms by which population structure increase pathogen richness this may not be a cause for concern. I the main mechanism is one that reduces pathogen extinction rates, a newly fragmented population is unlikely to increase it's pathogen richness over any appreciable timescale. If however population structure actively promotes the evolution of new pathogen strains or allows the persistence of more virulent strains this could have important public health implications. Therefore further study on the exact mechanisms by which population structure affects pathogen richness is needed.

5.0.6. *Conclusions.* In conclusion, this study adds to the evidence that population structure may promote pathogen richness. It does not support the view that factors that increase R_0 will increase pathogen richness. Using larger datasets and multiple measurements makes the weight of the evidence here stronger than in previous studies. However, caution must still be taken in interpreting these results as the data is biased and sparse and the analyses show that small changes in model and data choice can give opposite results.

6. Appendix

References

Colizza, V. and A. Vespignani (2007). "Invasion threshold in heterogeneous metapopulation networks". In: *Physical review letters* 99.14, p. 148701.

Vespignani, A. (2008). "Reaction-diffusion processes and epidemic metapopulation models in complex networks". In: *The European Physical Journal B* 64.3-4, pp. 349–353.

Nunn, C. L. et al. (2003). "Comparative tests of parasite species richness in primates". In: *The American Naturalist* 162.5, pp. 597–614.

Morand, S. (2000). “Wormy world: comparative tests of theoretical hypotheses on parasite species richness”. In: *Evolutionary Biology of Host-Parasite Relationships*. Ed. by S. M. R. Poulin and A. Skorping. Amsterdam: Elsevier, pp. 63–79.

Qiu, Z. et al. (2013). “The vector–host epidemic model with multiple strains in a patchy environment”. In: *Journal of Mathematical Analysis and Applications* 405.1, pp. 12–36. doi: `doi:10.1016/j.jmaa.2013.03.042`.

Allen, L. J., N. Kirupaharan, and S. M. Wilson (2004). “SIS epidemic models with multiple pathogen strains”. In: *Journal of Difference Equations and Applications* 10.1, pp. 53–75. doi: `10.1080/10236190310001603680`.

Nunes, A., M. T. da Gama, and M. Gomes (2006). “Localized contacts between hosts reduce pathogen diversity”. In: *Journal of theoretical biology* 241.3, pp. 477–487.

Maganga, G. D. et al. (2014). “Bat distribution size or shape as determinant of viral richness in african bats”. In: *PloS one* 9.6, e100172.

Gay, N. et al. (2014). “Parasite and viral species richness of Southeast Asian bats: Fragmentation of area distribution matters”. In: *International Journal for Parasitology: Parasites and Wildlife* 3.2, pp. 161–170. doi: `doi:10.1016/j.ijppaw.2014.06.003`.

Turmelle, A. S. and K. J. Olival (2009). “Correlates of viral richness in bats (order Chiroptera)”. In: *EcoHealth* 6.4, pp. 522–539.

Kerth, G., F. Mayer, and E. Petit (2002). “Extreme sex-biased dispersal in the communally breeding, nonmigratory Bechstein’s bat (*Myotis bechsteinii*)”. In: *Molecular Ecology* 11.8, pp. 1491–1498.

Hulva, P. et al. (2010). “Mechanisms of radiation in a bat group from the genus Pipistrellus inferred by phylogeography, demography and population genetics”. In: *Molecular ecology* 19.24, pp. 5417–5431.

McCracken, G. F. and J. W. Bradbury (1981). “Social organization and kinship in the polygynous bat *Phyllostomus hastatus*”. In: *Behavioral Ecology and Sociobiology* 8.1, pp. 11–34.

Petit, E. and F. Mayer (1999). “Male dispersal in the noctule bat (Nyctalus noctula): where are the limits?” In: *Proceedings of the Royal Society of London B: Biological Sciences* 266.1430, pp. 1717–1722.

Burland, T. et al. (1999). “Population genetic structure and gene flow in a gleaning bat, Plecotus auritus”. In: *Proceedings of the Royal Society of London B: Biological Sciences* 266.1422, pp. 975–980.

O’Donnell, C. F. et al. (2015). “Genetic diversity is maintained in the endangered New Zealand long-tailed bat (Chalinolobus tuberculatus) despite a closed social structure and regular population crashes”. In: *Conservation Genetics*, pp. 1–12.

Vonhof, M. J., A. L. Russell, and C. M. Miller-Butterworth (2015). “Range-Wide Genetic Analysis of Little Brown Bat (Myotis lucifugus) Populations: Estimating the Risk of Spread of White-Nose Syndrome”. In: *PloS one* 10.7.

Slatkin, M. (1995). “A measure of population subdivision based on microsatellite allele frequencies.” In: *Genetics* 139.1, pp. 457–462.

Blackwood, J. C. et al. (2013). “Resolving the roles of immunity, pathogenesis, and immigration for rabies persistence in vampire bats”. In: *Proceedings of the National Academy of Sciences* 110.51, pp. 20837–20842.

Pons-Salort, M. et al. (2014). “Insights into persistence mechanisms of a zoonotic virus in bat colonies using a multispecies metapopulation model”. In: *PloS one* 9.4, e95610.

Plowright, R. K. et al. (2011). “Urban habituation, ecological connectivity and epidemic dampening: the emergence of Hendra virus from flying foxes (Pteropus spp.)” In: *Proceedings of the Royal Society B: Biological Sciences* 278.1725, pp. 3703–3712.

IUCN (2010). *Red List Of Threatened Species. Version 2010.1*. www.iucnredlist.org.

Kamiya, T. et al. (2014). "What determines species richness of parasitic organisms? A meta-analysis across animal, plant and fungal hosts". In: *Biological Reviews* 89.1, pp. 123–134.

Arneberg, P. (2002). "Host population density and body mass as determinants of species richness in parasite communities: comparative analyses of directly transmitted nematodes of mammals". In: *Ecography* 1, pp. 88–94.

R Development Core Team (2010). *R: A Language And Environment For Statistical Computing*. ISBN 3-900051-07-0. R Foundation For Statistical Computing. Vienna, Austria. url: `Http://Www.R-Project.Org`.

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.

ICTV (2014). *International Committee on Taxonomy of Viruses Master Species List*. url: `http://talk.ictvonline.org/files/ictv_documents/m/msl/5208/download.aspx`.

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., J. Bielby, 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.

Canals, M. et al. (2005). "Relative size of hearts and lungs of small bats". In: *Acta Chiropterologica* 7.1, pp. 65–72.

Arita, H. T. (1993). "Rarity in Neotropical bats: correlations with phylogeny, diet, and body mass". In: *Ecological Applications*, pp. 506–517.

López-Baucells, A. et al. (2014). "Echolocation of the big red bat *Lasiurus egregius* (Chiroptera: Vespertilionidae) and first record from the Central Brazilian Amazon". In: *Studies on Neotropical Fauna and Environment* 49.1, pp. 18–25.

Orr, T. J. and M. Zuk (2013). "Does delayed fertilization facilitate sperm competition in bats?" In: *Behavioral Ecology and Sociobiology* 67.12, pp. 1903–1913.

Lim, B. K. and M. D. Engstrom (2001). "Bat community structure at Iwokrama forest, Guyana". In: *Journal of Tropical Ecology* 17.05, pp. 647–665.

Aldridge, H. (1987). "Turning flight of bats". In: *Journal of Experimental Biology* 128.1, pp. 419–425.

Ma, J. et al. (2003). "Dietary analysis confirms that Rickett's big-footed bat (*Myotis ricketti*) is a piscivore". In: *Journal of Zoology* 261.03, pp. 245–248.

Owen, S. F. et al. (2003). "Home-range size and habitat used by the northern myotis (*Myotis septentrionalis*)". In: *The American midland naturalist* 150.2, pp. 352–359.

Henderson, L. E. and H. G. Broders (2008). "Movements and resource selection of the Northern Long-Eared Myotis (*Myotis septentrionalis*) in a forest–agriculture landscape". In: *Journal of Mammalogy* 89.4, pp. 952–963.

Heaney, L. R. et al. (2012). "*Nyctalus plancyi* and *Falsistrellus petersi* (Chiroptera: Vespertilionidae) from northern Luzon, Philippines: ecology, phylogeny, and biogeographic implications". In: *Acta Chiropterologica* 14.2, pp. 265–278.

Oleksy, R., P. A. Racey, and G. Jones (2015). “High-resolution GPS tracking reveals habitat selection and the potential for long-distance seed dispersal by Madagascan flying foxes *Pteropus rufus*”. In: *Global Ecology and Conservation* 3, pp. 678–692.

Zhang, L. et al. (2009). “Recent surveys of bats (Mammalia: Chiroptera) from China. I. Rhinolophidae and Hipposideridae”. In: *Acta Chiropterologica* 11.1, pp. 71–88.

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`.

Jones, K. E., O. R. Bininda-Emonds, and J. L. Gittleman (2005). “Bats, clocks, and rocks: diversification patterns in Chiroptera”. In: *Evolution* 59.10, pp. 2243–2255.

Burnham, K. P. and D. R. Anderson (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media.

Whittingham, M. J., R. D. Swetnam, et al. (2005). “Habitat selection by yellowhammers *Emberiza citrinella* on lowland farmland at two spatial scales: implications for conservation management”. In: *Journal of applied ecology* 42.2, pp. 270–280.

Whittingham, M. J., P. A. Stephens, et al. (2006). “Why do we still use stepwise modelling in ecology and behaviour?” In: *Journal of animal ecology* 75.5, pp. 1182–1189.

Pinheiro, J. et al. (2015). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-122. url: `http://CRAN.R-project.org/package=nlme`.

Schielzeth, H. (2010). “Simple means to improve the interpretability of regression coefficients”. In: *Methods in Ecology and Evolution* 1.2, pp. 103–113.

Revell, L. J. (2010). “Phylogenetic signal and linear regression on species data”. In: *Methods in Ecology and Evolution* 1.4, pp. 319–329.

Anthony, S. J. et al. (2013). “A strategy to estimate unknown viral diversity in mammals”. In: *MBio* 4.5, e00598–13.