Phylogenetic and geographic patterns of bartonella host shifts among bat species

Clifton D. McKee[a,b,*], **David T. S. Hayman**[c], **Michael Y. Kosoy**[b], and **Colleen T. Webb**[a]

[a]Department of Biology, Colorado State University, Fort Collins, CO 80523, USA [b]Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, Fort Collins, CO 80521, USA [c]Molecular Epidemiology and Public Health Laboratory, Infectious Disease Research Centre, Massey University, Palmerston North 4442, New Zealand

Abstract

The influence of factors contributing to parasite diversity in individual hosts and communities are increasingly studied, but there has been less focus on the dominant processes leading to parasite diversification. Using bartonella infections in bats as a model system, we explored the influence of three processes that can contribute to bartonella diversification and lineage formation: (1) spatial correlation in the invasion and transmission of bartonella among bats (phylogeography); (2) divergent adaptation of bartonellae to bat hosts and arthropod vectors; and (3) evolutionary codivergence between bats and bartonellae. Using a combination of global fit techniques and ancestral state reconstruction, we found that codivergence appears to be the dominant process leading to diversification of bartonella in bats, with lineages of bartonellae corresponding to separate bat suborders, superfamilies, and families. Furthermore, we estimated the rates at which bartonellae shift bat hosts across taxonomic scales (suborders, superfamilies, and families) and found that transition rates decrease with increasing taxonomic distance, providing support for a mechanism that can contribute to the observed evolutionary congruence between bats and their associated bartonellae. While bartonella diversification is associated with host sympatry, the influence of this factor is minor compared to the influence of codivergence and there is a clear indication that some bartonella lineages span multiple regions, particularly between Africa and Southeast Asia. Divergent adaptation of bartonellae to bat hosts and arthropod vectors is apparent and can dilute the overall pattern of codivergence, however its importance in the formation of *Bartonella* lineages in bats is small relative to codivergence. We argue that exploring all three of these processes yields a more complete understanding of bat-bartonella relationships and the evolution of the genus *Bartonella*, generally. Application of these methods to other infectious bacteria and viruses could uncover common processes that lead to parasite diversification and the formation of host-parasite relationships.

*Corresponding author. clifton.mckee@colostate.edu (C. McKee).

1. Introduction

Parasites are astoundingly diverse and community ecology provides some conceptual foundations for the processes influencing parasite diversity (Seabloom et al., 2015). Specifically, the dominant processes can be broken down into four classes: selection, drift, dispersal, and speciation (Vellend, 2010). Previous work on zoonotic parasites has explored numerous factors that can influence the abundance and diversity of parasites in an individual host or community, including host behavior (Nunn and Dokey, 2006), the shape and fragmentation of a host's geographic range (Gay et al., 2014; Maganga et al., 2014), host group size (Ezenwa et al., 2006), host density (Lindenfors et al., 2007), latitudinal gradients (Bordes et al., 2011), climatic factors and host species richness (Brierley et al., 2016). Most of these factors relate to selection, drift, and dispersal, but not directly to speciation. Therefore, our aim with this study was to explore the influence of processes that may lead directly to parasite diversification and lineage formation.

We focused on three main processes, based on their preponderance in the literature, which may contribute to the diversification of parasites and the formation of phylogenetic structure (clades or lineages). First, spatial correlation in the invasion and transmission of parasites may result in the formation of parasite lineages based on the geographic origin or overlap of host species. This has been used in the past to track how parasites evolve and move across a landscape, such as with feline immunodeficiency virus in mountain lions (Biek et al., 2006), avian influenza in North American waterfowl (Lu et al., 2014), and brucellosis in cows, elk, and bison (Kamath et al., 2016). Second, lineages may form due to adaptation to a definitive host (the host in which a parasite reaches maturity) with only weak adaptation to intermediate hosts. This process could be especially important for vector-borne parasites, wherein the diversification of the parasite is more influenced by adaptation to vector species than to any individual animal host. This process has been demonstrated for parasites including *Borrelia burgdorferi* in *Ixodes* spp. ticks (Joy et al., 2008) and *Plasmodium vivax* in *Anopheles* spp. mosquitoes (Pal and Fikrig, 2003). Third, lineages of parasites may form as populations undergo radiating adaptation to a number of host species, as for *Spinturnix* mites and bats (Bruyndonckx et al., 2009) and between chewing lice and pocket gophers (Hafner and Page, 1995). This adaptation may occur through a process of strict cospeciation resulting in host and parasite phylogenies that are perfectly congruent, although this scenario is rare and also difficult to demonstrate (de Vienne et al., 2013). Alternatively, cross-species transmission and successful host shift speciation can become constrained by host species relatedness, such that parasites adapted to a particular host species will not successfully persist in a phylogenetically distant host species, as seen with rabies in North American bats (Faria et al., 2013; Streicker et al., 2010). This process can also generate congruence in host and parasite phylogenies. Overall, these three process all contribute to the isolation of

parasites by restricting dispersal and gene flow among populations, thereby encouraging the development of separate parasite lineages or species (Seabloom et al., 2015).

In this study, we tested the influence of these phylogenetic and geographic processes that potentially contribute to the diversification of bartonellae in bats. We used this as a model system because of the extreme diversity of bartonellae in bats and previous research demonstrating patterns of phylogeography (Berglund et al., 2010; Hayman et al., 2013), divergent adaptation to hosts and vectors (Chomel et al., 2009; Harms and Dehio, 2012; Tsai et al., 2011), and evolutionary codivergence with mammalian hosts (Lei and Olival, 2014) among *Bartonella* species. *Bartonella* is a genus of facultative intracellular bacteria found in a wide variety of mammals worldwide (Kosoy, 2010). Of the more than 30 described *Bartonella* species, around half have been identified as human pathogens causing a range of illnesses from mild fever to potentially fatal endocarditis (Breitschwerdt et al., 2010; Chomel and Kasten, 2010). Numerous studies have demonstrated that bats and their ectoparasites show a high prevalence and genetic diversity of bartonella bacteria (Anh et al., 2015; Bai et al., 2015, 2012, 2011; Brook et al., 2015; Concannon et al., 2005; Judson et al., 2015; Kamani et al., 2014; Kosoy et al., 2010; Lin et al., 2012; Olival et al., 2015; Reeves et al., 2007, 2005; Veikkolainen et al., 2014). Recently, bats have been implicated in potential spillover of bartonellae into dogs (Bai et al., 2010; Lin et al., 2012) and a single human case (Lin et al., 2010; Veikkolainen et al., 2014), although the role of bats as sources of zoonotic bartonellosis is still unclear (Mannerings et al., 2016).

Given that bats are evolutionarily ancient mammals (O'Leary et al., 2013) that are globally distributed, the accumulation of bartonella diversity may not be surprising. However, we seek to test the influence of processes leading to the diversification of bartonellae in bats and the formation of lineages that reflect parasite phylogeography, divergent adaptation to bat hosts or to arthropod vectors, or codivergence with bat hosts. Previous studies have only explored the influence of these processes in *Bartonella* singly whereas here we attempt to use an integrative approach that tests multiple hypotheses, which we believe will provide a better understanding of the evolution of these widespread and complex bacteria. We hypothesize that significant patterns of evolutionary codivergence with hosts, phylogeography, and divergent adaptation to hosts and vectors can all be found in bat bartonellae, although they may have varying degrees of importance in the formation of distinct *Bartonella* lineages. We first use global fit methods (Balbuena et al., 2013; Legendre et al., 2002) to look for significant patterns of phylogeography, evolutionary codivergence with bat hosts, and divergent adaptation to bats and vectors. We add to this analysis a linear model that can test the relative influence of codivergence and host sympatry on the evolution of *Bartonella* in bats. Finally, we use ancestral state reconstruction to reveal possible mechanisms that generate these significant patterns and test the importance of phylogeography, evolutionary codivergence with bat hosts, and divergent adaptation to bats and vectors in the formation of *Bartonella* lineages in bats using model selection.

2. Materials and methods

2.1. Compiled sequence data

We first compiled sequence data for this study from a previous analysis of bat-bartonella codivergence by Lei and Olival (2014) and then expanded by searching Web of Science and Google Scholar using the terms "bat* bartonella". This compiled data from the literature included partial citrate synthase gene sequences (*gltA*) for *Bartonella* genotypes from bats and ectoparasites (bat flies and fleas) from the UK, Guatemala, Peru, Taiwan, Finland, Puerto Rico, multiple countries in Africa, Costa Rica, and Vietnam (Anh et al., 2015; Bai et al., 2015, 2012, 2011; Billeter et al., 2012; Brook et al., 2015; Concannon et al., 2005; Judson et al., 2015; Kamani et al., 2014; Kosoy et al., 2010; Lin et al., 2012; Morse et al., 2012; Olival et al., 2015; Veikkolainen et al., 2014). The *gltA* gene has been shown to provide good phylogenetic resolution among known *Bartonella* species and subspecies (Norman et al., 1995) and is widely used for detection of bartonella infections. In addition, we searched for additional unpublished sequences on GenBank using the search terms "bat* bartonella" and found *gltA* sequences from bartonellae in bats and ectoparasites from Peru and Poland. We also included *gltA* sequences in the CDC database from bartonellae cultured from bats in Thailand (M. Kosoy, unpublished data). From each unique *Bartonella gltA* genotype found on GenBank, we extracted data on the genus and species of the bat host (Table S1). For *gltA* genotypes detected in ectoparasites, we extracted the genus and species of the ectoparasite and the bat host from GenBank and the associated published articles (Table S2).

We then collected cytochrome b (*cytb*) gene sequences (Table S3) from GenBank for each bat host species; this mitochondrial gene provides good phylogenetic resolution among mammalian species (Agnarsson et al., 2011; Bradley and Baker, 2001; Kocher et al., 1989). For bats identified only to the genus level or in cases where a suitable *cytb* sequence could not be found, representative or substitute species were chosen, as in Lei and Olival (2014). The criteria for representative and replacement species are discussed in detail in Appendix A. A sensitivity analysis using alternative suitable replacement bat species suggests that these host substitutions do not alter the observed phylogenetic or geographic patterns (see Appendix A). This is likely due to our constraint that substitute host species must be from the same genus, which have similar phylogenetic distances from other hosts in the tree and hence limits the effect of species substitution on the analyses. Host bat family, superfamily, and suborder were recorded based on IUCN Red List of Threatened Species (IUCN, 2014), the Mammal Species of the World 3rd Edition (Wilson and Reeder, 2005), and published articles (Agnarsson et al., 2011; Teeling et al., 2002) (Table S3).

In total, this dataset includes 173 unique *Bartonella* genotypes from 66 bat species, 41 genera, 11 families, and both recognized suborders, Yinpterochiroptera and Yangochiroptera (Teeling et al., 2002; Agnarsson et al., 2011). To check for evidence of sampling bias in measured diversity of *Bartonella* genotypes from each bat species, we counted the number of sampled bats of each species from the research studies included in the dataset and counted the number of articles published on each species by searching the binomial species name in

Web of Science (Table S3). Log-transformed host-parasite links were tested for correlation with log-transformed values of sampling effort (see Appendix B).

2.2. Compiled geographic range data

We downloaded shape files for geographic ranges of each bat species from the International Union for Conservation of Nature (IUCN) Red List website (http://www.iucnredlist.org/technical-documents/spatial-data) (IUCN, 2014). Using the command "over" from the R package "sp" and the commands "gIntersection", "gArea", and "gUnion" from the package "rgeos" for each species in the dataset, we calculated a) if each pair of bats' ranges overlapped, and if they overlapped, b) the area of the intersection between the two ranges and the geographic union of the two ranges (Bivand and Rundel, 2014; Bivand et al., 2013; Luis et al., 2013; Pebesma and Bivand, 2005; R Core Team, 2015). Percent overlap of species ranges was calculated by dividing the area of intersection of each pair of species by the geographic union of the total area covered by both species. This creates a symmetrical matrix that scales the percent overlap among species relative to the size of their respective ranges.

2.3. Phylogenetic analysis of sequence data

Lengths of *gltA* sequences varied considerably in the bartonella dataset, so sequence lengths were trimmed to 333 base pairs covered by all of the genotypes. The total length of *cytb* sequences in the bat species dataset was 1140 base pairs. *Brucella melitensis*, *Rhizobium leguminosarum*, and *Ochrobactrum anthropi* were chosen as outgroups for the bartonella phylogeny and *Ornithorynchus anatinus*, *Rattus rattus*, and *Equus caballus* were chosen as outgroups for the bat phylogeny. We aligned sequences with MAFFT using the accurate, local L-INS-I method (Katoh and Standley, 2013). Based on evolutionary model selection using jModelTest2 (Darriba et al., 2012), we chose the generalized time reversible substitution model (Nei and Kumar, 2000) with four gamma categories and a proportion of invariant sites (GTR+Γ+I) as the most appropriate model according to Akaike's information criterion. Maximum likelihood (ML) phylogenetic trees were generated with MEGA6 (Tamura et al., 2013); support for nodes in the tree was estimated from 1000 bootstrap replicates.

2.4. Calculation of host phylogeny and sympatry matrices

We calculated phylogenetic distances from branch lengths of the ML tree (patristic distances) of bat species using the "cophenetic" function in the "ape" package in R (Paradis et al., 2004; R Core Team, 2015). We then standardized these phylogenetic matrices by dividing the longest branch length in the matrix to constrain the matrix values between zero and one. Second, we transformed the geographic range overlap matrix into a distance by subtracting the percent overlap from one. Thus, like phylogenetic distances where closely related species have low distance values, species with highly overlapping ranges have low geographic distance values.

2.5. Overview of hypothesis testing approaches

Our analysis focused both on detecting significant patterns of evolutionary codivergence among hosts, parasite phylogeography, and divergent adaptation to hosts in vectors within bat *Bartonella* genotypes, but also on elucidating some of the underlying mechanisms that contribute to these patterns and the importance of these processes in the formation of *Bartonella* lineages. Thus, we combined several complementary analytical approaches. First, global fit methods were used to test for congruence between parasite phylogeny and host phylogeny or host range overlap. Global fit methods account for two confounding factors: some bat species host multiple *Bartonella* genotypes and some *Bartonella* genotypes are linked with multiple bat species. Measures of individual host-parasite linkage importance from these global fits were then used to test the contribution of potential adaptation to bat hosts and insect vectors. After evaluating the correlation between host phylogeny and sympatry, these factors were combined into a linear model to assess the relative contribution of codivergence and host sympatry to the topography of the bartonella tree. All of the above analyses were performed using both maximum likelihood and Bayesian trees to confirm results.

We used Bayesian ancestral state reconstruction of host bat taxonomy, geographic regions, and vector/host associations to assess how well the processes of codivergence with hosts, parasite phylogeography, and divergent adaptation to hosts and vectors correspond to the arrangement of observed bartonella lineages. These analyses simultaneously reconstruct the phylogeny, count the number of discrete state changes, and estimate the likelihood of the tree given the separate state models. Model selection was then performed in order to assess the predictive fit of the models (host taxonomy, geographic regions, or vector/host associations) to the arrangement of bartonella lineages. The patterns of discrete state changes were used to propose potential mechanisms that may give rise to the observed patterns of codivergence, phylogeography, and host/vector associations.

2.5.1. Global fit tests—Global fit analyses were performed first on the ML trees of bat species and *Bartonella* genotypes. We calculated two patristic distance matrices from bat and bartonella trees using the "cophenetic" command in the "ape" package in R (Paradis et al., 2004; R Core Team, 2015). We generated a third matrix for host-parasite links, which allows for multiple linkages among bat species and *Bartonella* genotypes. Two methods were used to measure the fit between bat and bartonella tree topologies through the matrix of host-parasite linkages, ParaFit (Legendre et al., 2002) and the Procrustean Approach to Cophylogeny (PACo) (Balbuena et al., 2013). Both methods decompose distance matrices of host and parasite phylogenies into principal components and combine those principal components with the matrix of host-parasite associations. However, the two methods have different null hypotheses and analytical approaches.

ParaFit tests the null hypothesis that host and parasite phylogenies are random with respect to one another. A global fit statistic for the cophylogeny is calculated based on the sum of squares of the combined matrices of the host and parasite phylogenetic principal components and the host-parasite association matrix. If a parasite and its associated parasite reside at corresponding branches of their respective trees, this results in a small contribution

to the fit statistic and is evidence of phylogenetic congruence. To test the significance of the global fit, the algorithm randomly changes the host-parasite association matrix by randomly assigning parasites to hosts and recalculates a global fit statistic. By performing many permutations, the algorithm can calculate a p-value for the observed global fit relative to the distribution of randomly permuted fits. A small p-value indicates that the host and parasite trees are not independent, providing evidence of evolutionary codivergence. Measures of host-parasite linkage importance are assessed by removing individual links during permutations and the resulting change in the residual sum of squares, called the ParaFitLink1 or F1 statistic.

In contrast, PACo takes the principal components of the host and parasite phylogenies and projects them in multivariate space, then the parasite matrix is scaled and rotated to fit the host matrix. A global fit from this scaling is calculated as the sum of squared residual distances between hosts and their associated parasites in the ordination. The significance of the fit statistic is tested using a similar permutation as in ParaFit, except that hosts are randomly assigned to parasites. In this way, PACo explicitly tests the degree to which parasite phylogeny depends on the host phylogeny (Balbuena et al., 2013). Individual host-parasite linkages are assessed based on their squared residuals alone. Individual residual values from PACo and linkage test results from ParaFit (F1 statistic, p-values) were saved to quantify the number of significant linkages among bats and *Bartonella* genotypes. Both tests were implemented using the "ape" and "vegan" packages in R (Oksanen et al., 2015; Paradis et al., 2004; R Core Team, 2015) with 10000 permutations.

These tests were repeated using the patristic distance matrix of *Bartonella* genotypes, the distance matrix of geographic overlaps, and the matrix of host-parasite linkages. The null hypotheses for the ParaFit analysis posits that the bartonella phylogeny is independent of host geographic range overlap and PACo tests the dependence of the bartonella phylogeny on host range overlap. Again, individual PACo residuals and ParaFit F1 statistics were saved. We used these individual linkage fit values to test the influence of divergent adaptation to bat hosts and ectoparasites on bartonella diversification using Wilcoxon rank sum tests.

2.5.2. Correlation and relative influence of host phylogeny and sympatry—A Mantel test (Mantel, 1967) was used to find the correlation between the two matrices, bat phylogenetic distance and bat sympatry. We used the "mantel" command in the "vegan" package in R (Oksanen et al., 2015; R Core Team, 2015) to calculate the correlation between the matrices using 10000 permutations. Based on our observation that host phylogeny and sympatry are both significantly associated with the bartonella phylogeny and that these factors are not highly correlated (discussed below), we assess the relative contribution of bat phylogeny and bat geographic range overlap to bartonella phylogeny. We combined the two matrices (M_{phy}, M_{geo}) in a linear model with individual weights (ω_{phy}, ω_{geo}) assigned to each matrix: $M_{combo} = \omega_{phy}M_{phy} + \omega_{geo}M_{geo}$, where $\omega_{phy} + \omega_{geo} = 1$. We assigned weights iteratively to each matrix in steps of 0.0001 and global fits to the bartonella phylogeny matrix were assessed at each step using ParaFit and PACo. We chose the optimum combination of weights based on the lowest global fit values for each algorithm.

2.5.3. Bayesian phylogenetic analysis and reconstruction of host shifts—

Following a previous study reconstructing host shifting events among *Bartonella* genotypes in rats (Hayman et al., 2013), we performed Bayesian Markov chain Monte Carlo (MCMC) analysis of bartonella sequence data from bats using BEAST 1.8.2 (Drummond and Rambaut, 2007; Drummond et al., 2012). The GTR+Γ+I substitution model with four gamma categories (Nei and Kumar, 2000) was used for the MAFFT alignments (Katoh and Standley, 2013) of bartonella *gltA* sequences. Base frequencies were estimated from the data and nodes of the tree were estimated using substitutions per site. We assumed the population sizes of *Bartonella* genotypes were constant for the coalescent model. We assigned discrete traits to sequences based on the family, superfamily, and suborder of the host bat, the geographic region in which the host bat was captured, and the origin of the genotype (directly from a bat or from an ectoparasite). BEAST independently estimates the rate of these discrete state transitions across the topology of trees generated from bartonella sequence data. Starting with a prior value of one, the clock rate for each discrete state was estimated from the average number of state transitions across all nodes in the phylogenetic tree. Due to lack of prior information regarding state transitions in bat bartonella, individual host family, superfamily, suborder, geographic region, and host/vector association transition rates were estimated, starting with a gamma prior distribution with shape and scale parameters set to one. This assumes initially that all state-to-state transitions will occur in the phylogeny at least once and estimated rates are then compared with this assumption. Choosing another initial value, such as zero, for any transitions may hinder convergence away from such a strong prior in the MCMC, and we have no other prior information about these transition rates. We kept all other priors for nucleotide frequencies and substitution rates at the default, diffuse settings. The choice of these priors appears to be justified based on the convergence of parameters away from the prior distribution in all MCMC chains.

We ran four separate MCMC chains with lengths of 2.5E8 iterations each in the analysis, sampling every 50000 iterations. Parameter log files and tree files were combined from these separate chains after discarding the first 10% of samples as burn-in using LogCombiner in BEAST. Tracer 1.6 (Drummond and Rambaut, 2007) was used to assess the mixing and convergence of parameters. We performed a second, identical Bayesian MCMC analysis on bat *cytb* sequences, using the discrete traits (bat host family, superfamily, suborder, geographic region, and vector/host association) to compare how the states transitioned across the bat tree topology versus the bartonella tree topology.

Gamma-distributed discrete state transition rates were estimated from the posterior of the MCMC chains. We inspected the median and 95% highest probability density (HPD) interval of each rate to find families, superfamilies, suborders, and geographic regions that had a number of exchanges over the topology of the phylogeny above one. Clock rates, or the mean number of transitions across all nodes, for each state were also inspected to quantify the overall trend in exchanges among bat families, superfamillies, suborders, and regions. We extracted all posterior state transition rates, tree likelihoods, and Akaike's information criterion using MCMC (AICM) (Raftery et al., 2007) values using the program Tracer 1.6. Finally, since ML and Bayesian methods can draw slightly different trees, we repeated global fit analyses and linear modeling (from sections 2.5.1 and 2.5.2) using the Bayesian phylogenies of bat species and *Bartonella* genotypes, as well as the bat species

sympatry matrix and the Bayesian tree of *Bartonella* genotypes to confirm the results obtained from ML trees.

3. Results

3.1. Phylogenetic analysis

The maximum likelihood (ML) tree of bat *cytb* sequences (Fig. S1) matches well with previous phylogenies of bats (Agnarsson et al., 2011; Teeling et al., 2002), but with strong support (>90%) only for closely related genera and for some families. The ML tree of bartonella *gltA* sequences (Fig. S2) also had the best support (>90%) for closely related genotypes and around the putative species level, i.e., >4% sequence divergence for *gltA* following La Scola et al. (2003). However, deeper nodes linking clusters of genotypes generally had lower support.

3.2. Global fit tests

ParaFit and PACo analyses provided strong support for an evolutionary association between bartonella and bats in the dataset as well as a significant relationship between bat sympatry and the bartonella phylogeny (Table S4). Specifically, the ParaFit tests found significant evidence to reject the null hypotheses that the bat phylogeny and the bartonella phylogeny are independent, and that bat sympatry and the bartonella phylogeny are independent. The PACo tests found significant evidence to conclude that the bartonella phylogeny is strongly predicted by both bat phylogeny and bat sympatry.

3.3. Correlation and relative influence of host phylogeny and sympatry

The map of species distributions (Fig. 1) indicates that there is a high level of range overlap (indicated by darker shading) among bats in the dataset, particularly within Central America and the Caribbean, South America, Europe, Africa, Central and East Asia, and Southeast Asia. The Mantel test (Mantel, 1967) shows that matrices of bat phylogenetic distances and geographic range overlaps have a significant positive correlation (Pearson correlation coefficient, $r = 0.28$, $P = 1E-5$). However, this is a weak correlation (Hinkle et al., 2003) and thus these factors would not have high collinearity in our linear model (outlined in section 2.5.2) that tests the relative influence of both covariates in explaining the topology of the bartonella tree.

Using the linear combination of bat phylogeny and bat geographic overlap matrices, the optimal combination using ParaFit was $\omega_{phy} = 1$ and $\omega_{geo} = 0$, ParaFitGlobal = 12.07. The optimal combination using PACo was $\omega_{phy} = 1$ and $\omega_{geo} = 0$, m^2 global value = 8.11. In both cases, the assignment of all of the weight to the bat phylogeny matrix indicates that bat geographic overlap plays an insignificant role in the structuring of bartonella phylogeny relative to host phylogeny, despite its apparent significance in a single factor model.

3.4. Analysis of vector-associated versus bat-associated genotypes

Comparison of global fit statistics for bat-bartonella associations for *Bartonella* genotypes detected from bat hosts or from ectoparasites suggests that bartonellae from ectoparasites have weaker fits than bartonellae from bats hosts, based on Wilcoxon rank sum tests. In

particular, the medians of the residuals from the PACo tests based on ML and Bayesian trees for *Bartonella* genotypes detected in ectoparasites were greater than the medians of the residuals from genotypes detected in bats (Fig. S9; ML trees: Wilcoxon rank sum, W = 3110, P = 0.05; Bayesian trees: W = 3066.5, P = 0.04). The median of F1 statistics from ParaFit tests for bartonellae from ectoparasites were significantly lower than bartonellae from bats using ML trees (Figure S9; Wilcoxon rank sum, W = 4494, P = 0.04) but were not significantly different using the Bayesian trees (Wilcoxon rank sum, W = 3612.5, P = 0.63). The squared residuals of the PACo global fit quantify the error in the prediction of the bartonella phylogeny using the host phylogeny. ParaFit F1 statistics measure the effect of each individual link on the global fit when the link is randomly removed. Based on these interpretations, bartonellae from ectoparasites are more likely to have large residuals compared with bartonellae from bat hosts and thus weaken the predictive relationship between host phylogeny and parasite phylogeny as assessed by PACo. And related to the F1 statistics from ParaFit, removing an ectoparasite linkage is more likely to improve the global fit than removing a host linkage, but this is only true for ML trees.

3.5. Bayesian phylogenetic analysis and reconstruction of host shifts

The Bayesian phylogenetic analyses of *Bartonella gltA* genotypes and bat *cytb* sequences yielded trees (Fig. 2–4; Fig. S3–S8) with good convergence and large effective sample sizes (ESS >200) for all parameters. There was strong posterior probability (PP >90%) for nodes linking closely related genotypes as with the ML trees, however there was also equally strong support for some deeper nodes. In support of the phylogeography hypothesis, there does appear to be some division between Old World and New World *Bartonella* genotypes (Fig. 3) that is mirrored in the bat phylogeny (Fig. S6), but this applies mainly to bat families that are restricted to particular continents. For example, bats in the families Pteropodidae, Rhinolophidae, and Hipposideridae are restricted to the Old World whereas bats in the families Phyllostomidae, Noctolionidae, and Mormoopidae are restricted to the New World. For more cosmopolitan families, such as Vespertilionidae, there is some exchange between *Bartonella* genotypes associated with these bats across Europe, North America, and Central/South America. Related to the vector/host association hypothesis, there appears to be no strict separation between bat-associated *Bartonella* genotypes and vector-associated genotypes (Fig. 4), with many clades consisting of both bat-associated and vector-associated bartonellae. In general, the trees provide the most support for the codivergence hypothesis. The *Bartonella* genotypes cluster according to bat suborder (Fig. 2), with defined clades for *Bartonella* associated with bats from Yinpterochiroptera and Yangochiroptera (compare to Fig. S5). These phylogenetic patterns extend to lower taxa, with *Bartonella* clades separated into groups corresponding to related bat families and superfamilies (Fig. S7–S8; compare to Fig. S3–S4).

The relative importance of the three processes in predicting the structure of bat bartonella lineages was inferred by model selection using AICM values. The trees supporting the codivergence hypothesis (suborder AICM = 53.46; superfamily AICM = 144.41; family AICM = 296.82) have generally lower AICM values than trees supporting the phylogeography hypothesis (region AICM = 227.98) or the divergent vector/host association hypothesis (mean vector tree likelihood = 304.1). These results suggest that the most likely

process leading to the formation of bartonella lineages is evolutionary codivergence between bat hosts and their bartonellae.

Extracted posterior estimates of state transitions among families, superfamilies, suborders, and geographic regions were generally low; Table 1 shows only the transition rates with a median value greater than one. All but one (Emballonuridae – Pteropodidae) of the family transition rates listed in Table 1 are between pairs of families within the same suborder and six of the ten were in the same superfamily (Fig. S8). Four pairs of families had particularly strong connections, Hipposideridae – Megadermatidae, Miniopteridae – Vespertilionidae, Mormoopidae – Phyllostomidae, and Pteropodidae – Rhinolophiae, with median transition rates greater than two. With the exception of Pteropodidae – Rhinolophidae, these families with higher transition rates are in the same superfamily. Like family transitions, superfamily transition rates were generally low, with only three pairs of superfamilies with median rates greater than one, and only one pair, Pteropoidea – Rhinolophoidea, with a transition rate greater than two (Table 1). Exchange between the two suborders occurs only once in the tree of *Bartonella* genotypes after the split between Yinpterochiroptera and Yangochiroptera, as can be seen in Fig. 2, with a single transition between Pteropodidae and Emballonuridae. In general, bartonella transition rates among bat hosts as measured by clock rates appear to decline with increasing phylogenetic distance. The family clock rate shows that on average 1.16 cross-family transitions occur across the bartonella tree. Superfamily and suborder clock rates are lower than the prior expectation of one, estimating 0.49 cross-superfamily transitions and 0.16 cross-suborder transitions across the tree.

There is a significant amount of exchange between several geographic regions, particularly between Africa – Southeast Asia, with a median number of transitions (5.47) significantly greater than one. These geographical exchanges, especially between African and Southeast Asian fruit bats, are also seen in the bat phylogeny (Fig. S6) and have been observed in previous studies of Old World fruit bat phylogeography (Juste et al., 1999; Almeida et al., 2016). Additionally, it appears that transitions from bat hosts to arthropod vectors occur frequently, with a median of 2.73 transitions in the tree, and Fig. 4 shows that these exchanges can be bidirectional.

Repeated global fit analyses using Bayesian trees yielded similar results to the tests using ML trees, with strong support for evolutionary codivergence between bats and bartonella and a relationship between host sympatry and the bartonella tree (Table S4). Similar to the analyses using ML trees, there is fairly weak correlation between bat phylogeny and geographic range overlap (Mantel test, Pearson correlation coefficient, r = 0.4, P = 1E-5) and the optimal linear combination of the bat phylogeny and sympatry matrices from the Bayesian trees assigned all of the weight to the phylogeny matrices.

The outliers in the global fit analyses using both ML and Bayesian trees reflect either cross-family transitions that appear to be incidental or associations between bats and *Bartonella* genotypes that are very distant from other clades. *Myotis keaysi* carries two genotypes (KJ816676 and KJ816669) that are more closely related to bartonellae hosted by phyllostomid bats than vespertilionid bats. *Pteronotus davyi* shares a genotype (HM597202) with *Glossophaga soricina* and appears to have acquired other *Bartonella* genotypes from

phyllostomid bats. The five other outliers (JN17066, KP100358, KP100353, KP100346, and KP100343) are basal lineages associated with *Eidolon helvum* (more specifically, the bat fly *Cyclopodiae greefi* found on this host) and *Rhinolophus* spp.

4. Discussion

In this study, we investigated the influence of three processes that may contribute to the diversification of bartonellae in bats: phylogeography, divergent adaptation to bat hosts and vectors, and evolutionary codivergence. Due to the wide diversity of bartonellae found in bats worldwide, we believe this is a good system to investigate these patterns and may serve as a model for understanding diversification processes in other vector-borne zoonotic diseases.

4.1. Evolutionary codivergence

Our analyses largely support the primacy of adaptation to bat hosts in the evolution of bat-associated bartonellae, which confirms previous work by Lei and Olival (2014). The dataset we used included a larger number of bartonella *gltA* sequences from bats and bat ectoparasites representing a greater number of families and from more regions than this previous study; bartonella sequences from ectoparasitic bat flies, fleas, and mites are now additionally included. Using these additional data, we could have found support for alternative hypotheses regarding sympatry and vectors that were previously untested in this context. Moreover, the inclusion of more sequences could have diluted the congruence observed previously, especially if many of the host-parasite associations arose from apparent host-shift events over large phylogenetic distances. Yet our analysis shows that this overall congruence between bats and bartonella is robust to the new sequences and perhaps even enhanced. The addition of the new sequences now allows us to better understand how *Bartonella* genotypes are constrained hierarchically at multiple taxonomic levels, a pattern that was not explored in this previous study. Fig. 2 shows a clear visual congruence between bat hosts and *Bartonella* genotypes when the branches are colored by the host suborders, providing another line of evidence supporting the division between Yinpterochiroptera and Yangochiroptera (Agnarsson et al., 2011; Teeling et al., 2002). Furthermore, the taxonomic pattern extends to the level of bat superfamilies and families (Fig. S7–S8). The formation of distinct clades of *Bartonella* genotypes linking families, superfamilies, and suborders of bats suggest that bartonellae have been codiverging with bats over significant evolutionary time. These patterns were not uncovered in Lei and Olival's study and could have implications for future studies of bat taxonomy.

Our Bayesian trees also clearly show that the vast majority of host transitions and duplication events occur within the same bat family and that transitions between families, superfamilies, and suborders happen infrequently. Specifically, our analysis estimates that only 1.16 cross-family transitions, 0.49 cross-superfamily transitions, and 0.16 cross-order transitions occur on average across all nodes of the tree. These results support the expectation that transitions of bartonellae between bat host species would be constrained by host relatedness, as has been demonstrated for bat rabies (Faria et al., 2013; Streicker et al., 2010). This association was not examined previously (Lei and Olival, 2014), but may be a

general mechanism, together with cospeciation, that generates the significant signal of evolutionary codivergence between bartonellae in bats. Nevertheless, the dataset of bartonella sequences is still small, so our estimations of transition rates across phylogenetic scales may be limited in their accuracy.

4.2. Divergent adaptation to bats and vectors

It appears from the Bayesian reconstruction of bartonella associations with bat hosts and ectoparasite vectors (Fig. 4) that there is not a strict separation of bat-associated and vector-associated *Bartonella* genotypes. Instead, the model indicates that there is frequent and bidirectional exchange of bartonellae between bats and ectoparasites (Table 1). Nevertheless, our analyses suggest that vectors can still have some influence on the patterns of bartonella divergence in bats. We found that the most significant outliers from the global fit analyses were produced by links between bats and *Bartonella* genotypes associated with a different family of bats than the apparent host species, or by *Bartonella* genotypes at the base of the phylogenetic tree with uncertain relationships with other sequence types. One possibility is that these deeply divergent genotypes are symbiotic *Bartonella* genotypes primarily adapted to an arthropod reservoir (Morse et al., 2012; Zhu et al., 2014), and the presence of the bacteria in the bats is accidental and/or transient. Additionally, comparison of the residuals and test statistics from the global fit tests for *Bartonella* genotypes detected in bats versus ectoparasites showed that *Bartonella* genotypes from ectoparasites had weaker fits to host phylogeny than those genotypes detected in bats. This may be due to adaptation of these genotypes to a symbiotic lifestyle in the arthropods or a lack of host specificity in the ectoparasite feeding behavior, such that a *Bartonella* genotype acquired from one bat may have been retained in the vector when it moved to feed on another bat host species. For instance, the bat fly *Cyclopodia greefi* can be found feeding on *Eidolon helvum*, *Micropteropus* spp., and *Epomophorus* spp. bats (Kamani et al., 2014) and *Trichobius adamsi* bat flies have been picked from *Macrotus waterhousii* and *Phyllonycteris poeyi* (Morse et al., 2012). Though it was not possible with this study with our current dataset, in the future it would be very informative to include the host range and phylogeny of ectoparasites to further explore bartonella associations with bats and vectors and the history of host transitions.

4.3. Phylogeography

An important confounding factor in the study of host-parasite relationships is the influence of geography, specifically the correlation between host species relatedness and sympatry. If two host species are closely related and also have a high degree of range overlap, it might be expected that these two species would share similar parasites. High amounts of sympatry and interaction in common habitats may be able to facilitate cross-species transitions despite phylogenetic distance. If there was a high correlation between host phylogeny and sympatry, then one would not be able to distinguish which factor explains the structure of the parasite tree. This issue was not addressed by Lei and Olival (2014) in their previous study of bat-bartonella codivergence.

Our analysis found significant congruence between bat sympatry and our bartonella phylogeny. When we tested the correlation between host phylogeny and sympatry, we found

this association to be fairly weak for both ML and Bayesian trees. Our linear model then found that between these two processes, host phylogeny appears to be the primary factor explaining the topology of the bartonella trees. It is likely that the significant association between sympatry and bartonella phylogeny is due to range restrictions of some bat families (e.g., Pteropodidae, Rhinolophidae, and Hipposideridae in the Old World and Phyllostomidae, Noctolionidae, and Mormoopidae in the New World). Future work should explore the association between host sympatry and bartonella diversification at a more restricted geographic scale. Furthermore, other methods are available for examining relationships between sympatry and phylogeny, such as Faith's measure of phylogenetic diversity (Faith, 1992) and distance-based Moran's eigenvector maps (Legendre et al., 2015), and will be investigated in future analyses.

Transitions between geographic regions are infrequent, with a median of 1.05 region transitions across all nodes of the tree (Table 1). The continental regions contributing most to this rate are exchanges between Africa and Southeast Asia, but the apparent exchanges of bartonellae across these species are difficult to explain geographically. There are no obvious bridge species that would connect these two regions in the dataset (Figure 1), however it is possible that a bridge species exists and has not yet been sampled. We intentionally limited the number of regions in this analysis to reduce the number of transition rates that needed to be estimated, so this may obscure some more subtle patterns of exchange at smaller spatial scales, especially within East/Southeast Asia and Central/South America and the Caribbean. There is a high level of sympatry among phyllostomid bats represented in the dataset from the Americas, with numerous species having ranges that span the entire region. Hence, the interaction between many closely related host species in sympatry would be expected to facilitate transmission of bartonellae across geographic boundaries.

4.4. Limitations and future work

Our analyses captured some very general trends in the evolution of bats and bartonella, but there are still substantial gaps in our understanding of the mechanisms that contribute to this pattern. These gaps may begin to be closed with the acquisition of new bartonella sequences from other bat species and other regions. The 66 species used in this study represent less than 5% of the ~1240 species of bats worldwide, with sampling from only 26 (13%) of 196 countries. Figure 1 highlights some of these geographic deficiencies, particularly Australia and Oceania, Central and Western Asia, and North America. Our test of bias in research effort indicates that we have probably only scratched the surface of bartonella diversity in bats, even within individual species (Appendix B; Fig. S10). In the future, it would be useful to account for this sampling bias in models of bartonella diversification in bats. However, we are not aware of a formal framework for including sampling effort in the global fit tests nor in our current ancestral state reconstruction approach, which is aggregated at the family level and above, and thus we present these results separately.

Another important gap in the study of bat-bartonella relationships is the limited amount of information contained within the citrate synthase gene (*gltA*), the most popular marker used for the detection of bartonella. The short sequence length prevents us from resolving the position of many branches across the bartonella phylogenetic tree or measuring mutation

rates for the estimation of divergence times (Hayman et al., 2013). Estimated divergence times for clades of *Bartonella* genotypes would have been especially useful in our analysis to compare with published bat phylogenies, to see if host species and parasite genotypes began to radiate at the same time. However, to estimate divergence times we would need more sequence information, perhaps in the form of multi-locus sequence typing (MLST) or whole-genome sequencing (WGS) datasets. MLST or WGS datasets could also measure the frequency of lateral gene transfer (LGT) and recombination events that could confound patterns of cophylogeny. For example, some of the apparent host shifting events may not represent invasion by an entirely separate species of *Bartonella*, but rather just the *gltA* gene that has undergone homologous recombination into a separate genome after coinfection of two species within an individual mammalian or arthropod host. Recent studies have shown that rates of LGT and recombination in bartonellae are higher than previously expected given its intracellular lifestyle (Bai et al., 2015; Berglund et al., 2010; Buffet et al., 2013; Paziewska et al., 2012, 2011; Vos and Didelot, 2008). Therefore, sequencing of multiple genomic regions or genes related to the host cell invasion process may be more informative for showing fine-scale differences among bartonellae that better reflect their transmission history.

We also recognize that our analyses of bat sympatry depend on the quality of the geographic range data used. For some species, the range database on IUCN may be incomplete due to lack of sampling, and this has the potential to affect our calculation of range overlaps between species. Currently, IUCN ranges estimates are the best data available for all of the bat species in our dataset and previous studies have highlighted important patterns in bat viruses using IUCN data (Luis et al., 2013, 2015). Other approaches, such as species distribution modeling, may provide better range estimates, yet this would require reliable environmental estimates for all regions represented in our dataset and is beyond the scope of our study.

Finally, other statistical methods are available for testing the relative influence of evolutionary codivergence, phylogeography, and vector specificity in the evolution of bartonella in bats. For example, generalized linear modeling (GLM) has been used previously to investigate the determinants of cross-species rabies transmission in bats, including host distance, range overlap, and bat morphological traits (Faria et al., 2013). However, further improvements to the dataset are needed for this approach. Particularly, sufficient data on the host range of the bat ectoparasites are needed to include this as a factor in a GLM. Additionally, there was a large number of bat species in our dataset (66), which would have required 2145 species-to-species transition rate estimations, a number that was computationally prohibitive. Increased genotypic diversity of bartonellae in a restricted number of species would improve parameter estimation of species-to-species transition rates in the ancestral state reconstruction. Given these limitations, linear modeling of sympatry and bat phylogeny in the global fit analyses and model selection after ancestral state reconstruction were the best available methods for testing the relative importance of our three hypothesized processes in determining the observed phylogeny. Future work that combines denser sampling of *Bartonella* genotypes in a limited number of bat species, information on the host range of ectoparasites, and factors relating to sampling effort (as

discussed above) in a GLM would further elucidate the relative influence of evolutionary codivergence, host sympatry, and vector specificity on the evolution of bartonellae in bats.

5. Conclusion

The study of bat-bartonella evolutionary relationships, and by extension host-parasite relationships generally, is not only interesting from a biological perspective, but can also aid in the identification of zoonoses in humans and domestic animals. For instance, Lin et al. (2012) saw that bartonella *gltA* sequences from *Miniopterus schreibersii* bats were 96% similar to bartonellae found in stray dogs in Thailand by Bai et al. (2010), suggesting potential spillover. Veikkolainen et al. (2014) found sequences in vespertilionid bats that were very similar to *Bartonella mayotimonensis*, a novel agent of endocarditis in a human patient from the United States (Lin et al., 2010). Numerous other cases of human and domestic animal bartonellosis have been ultimately attributed to zoonotic origin. Studying how these bartonellae evolve and persist in their reservoir species may help to understand the mechanisms that facilitate emergence in novel host species and cause disease. The specific methods used in this study are particularly useful for diverse and rapidly evolving microparasites like bacteria and viruses. Application to other systems could reveal general mechanisms of host-parasite evolution and diversification and discover deep relationships at the root of some of our most pervasive infectious diseases.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Agnarsson I, Zambrana-Torrelio CM, Flores-Saldana NP, May-Collado LJ. A time-calibrated species-level phylogeny of bats (Chiroptera, Mammalia). PLoS Curr. 2011; 3:RRN1212.doi: 10.1371/currents.RRN1212 [PubMed: 21327164]

Almeida FC, Giannini NP, Simmons NB. The evolutionary history of the African fruit bats (Chiroptera: Pteropodidae). Acta Chiropterologica. 2016; 18:73–90. DOI: 10.3161/15081109ACC2016.18.1.003

Anh PH, Van Cuong N, Son NT, Tue NT, Kosoy MY, Woolhouse ME, Baker S, Bryant JE, Thwaites G, Carrique-Mas JJ, Rabaa MA. Diversity of *Bartonella* spp. in bats, southern. Vietnam Emerg Infect Dis. 2015; 21:1266–1267. DOI: 10.3201/eid2107.141760 [PubMed: 26079810]

Bai Y, Hayman DT, McKee CD, Kosoy MY. Classification of *Bartonella* strains associated with straw-colored fruit bats (*Eidolon helvum*) across Africa using a multi-locus sequence typing platform. PLoS Negl Trop Dis. 2015; 9:e0003478.doi: 10.1371/journal.pntd.0003478 [PubMed: 25635826]

Bai Y, Kosoy MY, Boonmar S, Sawatwong P, Sangmaneedet S, Peruski LF. Enrichment culture and molecular identification of diverse *Bartonella* species in stray dogs. Vet Microbiol. 2010; 146:314–319. DOI: 10.1016/j.vetmic.2010.05.017 [PubMed: 20570065]

Bai Y, Kosoy MY, Recuenco S, Alvarez Castillo D, Moran D, Turmelle AS, Ellison J, Garcia DL, Estevez A, Lindblade K, Rupprecht CE. *Bartonella* spp. in bats, Guatemala. Emerg Infect Dis. 2011; 17:1269–1272. DOI: 10.3201/eid1707.101867 [PubMed: 21762584]

Bai Y, Recuenco S, Gilbert AT, Osikowicz LM, Gomez J, Rupprecht CE, Kosoy MY. Prevalence and diversity of *Bartonella* spp. in bats in Peru. Am J Trop Med Hyg. 2012; 87:518–523. DOI: 10.4269/ajtmh.2012.12-0097 [PubMed: 22826480]

Balbuena JA, Míguez-Lozano R, Blasco-Costa I. PACo: a novel procrustes application to cophylogenetic analysis. PLoS One. 2013; 8:e61048.doi: 10.1371/journal.pone.0061048 [PubMed: 23580325]

Berglund EC, Ellegaard K, Granberg F, Xie Z, Maruyama S, Kosoy MY, Birtles RJ, Andersson SG. Rapid diversification by recombination in *Bartonella grahamii* from wild rodents in Asia contrasts with low levels of genomic divergence in Northern Europe and America. Mol Ecol. 2010; 19:2241–2255. DOI: 10.1111/j.1365-294X.2010.04646.x [PubMed: 20465583]

Biek R, Drummond AJ, Poss M. A virus reveals population structure and recent demographic history of its carnivore host. Science. 2006; 311:538–541. DOI: 10.1126/science.1121360 [PubMed: 16439664]

Billeter SA, Hayman DT, Peel AJ, Baker KS, Wood JL, Cunningham AA, Suu-Ire RD, Dittmar K, Kosoy MY. *Bartonella* species in bat flies (Diptera: Nycteribiidae) from western Africa. Parasitology. 2012; 139:324–329. DOI: 10.1017/S0031182011002113 [PubMed: 22309510]

Bivand, R.; Pebesma, E.; Gomez-Rubio, V. Applied spatial data analysis with R. 2. Springer; New York: 2013.

Bivand, R.; Rundel, C. rgeos: interface to geometry engine - open source (GEOS) [WWW Document]. 2014. http://cran.r-project.org/package=rgeos/

Bordes F, Guégan JF, Morand S. Microparasite species richness in rodents is higher at lower latitudes and is associated with reduced litter size. Oikos. 2011; 120:1889–1896. DOI: 10.1111/j.1600-0706.2011.19314.x

Bradley RD, Baker RJ. A test of the genetic species concept: cytochrome-b sequences and mammals. J Mammal. 2001; 82:960–973. DOI: 10.1644/1545-1542(2001)082<0960:ATOTGS>2.0.CO;2

Breitschwerdt EB, Maggi RG, Chomel BB, Lappin MR. Bartonellosis: an emerging infectious disease of zoonotic importance to animals and human beings. J Vet Emerg Crit Care. 2010; 20:8–30. DOI: 10.1111/j.1476-4431.2009.00496.x

Brierley L, Vonhof MJ, Olival KJ, Daszak P, Jones KE. Quantifying global drivers of zoonotic bat viruses: a process-based perspective. Am Nat. 2016; 187:E53–E64. DOI: 10.1086/684391 [PubMed: 26807755]

Brook CE, Bai Y, Dobson AP, Osikowicz LM, Ranaivoson HC, Zhu Q, Kosoy MY, Dittmar K. *Bartonella* spp. in fruit bats and blood-feeding ectoparasites in Madagascar. PLoS Negl Trop Dis. 2015; 9:e0003532.doi: 10.1371/journal.pntd.0003532 [PubMed: 25706653]

Bruyndonckx N, Dubey S, Ruedi M, Christe P. Molecular cophylogenetic relationships between European bats and their ectoparasitic mites (Acari, Spinturnicidae). Mol Phylogenet Evol. 2009; 51:227–237. DOI: 10.1016/j.ympev.2009.02.005 [PubMed: 19236931]

Buffet JP, Pisanu B, Brisse S, Roussel S, Félix B, Halos L, Chapuis JL, Vayssier-Taussat M. Deciphering *Bartonella* diversity, recombination, and host specificity in a rodent community. PLoS One. 2013; 8:e68956.doi: 10.1371/journal.pone.0068956 [PubMed: 23894381]

Chomel BB, Boulouis HJ, Breitschwerdt EB, Kasten RW, Vayssier-Taussat M, Birtles RJ, Koehler JE, Dehio C. Ecological fitness and strategies of adaptation of *Bartonella* species to their hosts and vectors. Vet Res. 2009; 40:29.doi: 10.1051/vetres/2009011 [PubMed: 19284965]

Chomel BB, Kasten RW. Bartonellosis, an increasingly recognized zoonosis. J Appl Microbiol. 2010; 109:743–750. DOI: 10.1111/j.1365-2672.2010.04679.x [PubMed: 20148999]

Concannon R, Wynn-Owen K, Simpson V, Birtles RJ. Molecular characterization of haemoparasites infecting bats (Microchiroptera) in Cornwall, UK. Parasitology. 2005; 131:489–496. DOI: 10.1017/S0031182005008097 [PubMed: 16174413]

Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. 2012; 9:772–772. DOI: 10.1038/nmeth.2109 [PubMed: 22847109]

de Vienne DM, Refrégier G, López-Villavicencio M, Tellier A, Hood M, Giraud T. Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. New Phytol. 2013; 198:347–385. DOI: 10.1111/nph.12150 [PubMed: 23437795]

Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007; 7:214.doi: 10.1186/1471-2148-7-214 [PubMed: 17996036]

Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012; 29:1969–1973. DOI: 10.1093/molbev/mss075 [PubMed: 22367748]

Ezenwa VO, Price SA, Altizer SM, Vitone ND, Cook KC. Host traits and parasite species richness in even and odd-toed hoofed mammals, Artiodactyla and Perissodactyla. Oikos. 2006; 115:526–536. DOI: 10.1111/j.2006.0030-1299.15186.x

Faith DP. Conservation evaluation and phylogenetic diversity. Biol Conserv. 1992; 61:1–10. DOI: 10.1016/0006-3207(92)91201-3

Faria NR, Suchard Ma, Rambaut A, Streicker DG, Lemey P. Simultaneously reconstructing viral cross-species transmission history and identifying the underlying constraints. Philos Trans R Soc Lond B Biol Sci. 2013; 368:20120196–20120196. DOI: 10.1098/rstb.2012.0196 [PubMed: 23382420]

Gay N, Olival KJ, Bumrungsri S, Siriaroonrat B, Bourgarel M, Morand S. Parasite and viral species richness of Southeast Asian bats: fragmentation of area distribution matters. Int J Parasitol Parasites Wildl. 2014; 3:161–170. DOI: 10.1016/j.ijppaw.2014.06.003 [PubMed: 25161915]

Hafner MS, Page RD. Molecular phylogenies and host-parasite cospeciation: gophers and lice as a model system. Philos Trans R Soc Lond B Biol Sci. 1995; 349:77–83. [PubMed: 8748020]

Harms A, Dehio C. Intruders below the radar: molecular pathogenesis of *Bartonella* spp. Clin Microbiol Rev. 2012; 25:42–78. DOI: 10.1128/CMR.05009-11 [PubMed: 22232371]

Hayman DT, McDonald KD, Kosoy MY. Evolutionary history of rat-borne *Bartonella*: the importance of commensal rats in the dissemination of bacterial infections globally. Ecol Evol. 2013; 3:3195–3203. DOI: 10.1002/ece3.702 [PubMed: 24223261]

Hinkle, D.; Wiersma, W.; Jurs, S. Applied statistics for the behavioral sciences. 5. Houghton Mifflin; Boston: 2003.

IUCN. The IUCN Red List of threatened species [WWW Document]. 2014. http://www.iucnredlist.org/

Joy DA, Gonzalez-Ceron L, Carlton JM, Gueye A, Fay M, McCutchan TF, Su XZ. Local adaptation and vector-mediated population structure in *Plasmodium vivax* malaria. Mol Biol Evol. 2008; 25:1245–1252. DOI: 10.1093/molbev/msn073 [PubMed: 18385220]

Judson S, Frank H, Hadly E. Bartonellae are prevalent and diverse in Costa Rican bats and bat flies. Zoonoses Public Health. 2015; 62:609–617. DOI: 10.1111/zph.12188 [PubMed: 25810119]

Juste J, Alvarez Y, Tabarés E, Garrido-Pertierra A, Ibáñez C, Bautista JM. Phylogeography of African fruitbats (Megachiroptera). Mol Phylogenet Evol. 1999; 13:596–604. DOI: 10.1006/mpev.1999.0669 [PubMed: 10620416]

Kamani J, Baneth G, Mitchell M, Mumcuoglu KY, Gutiérrez R, Harrus S. *Bartonella* species in bats (Chiroptera) and bat flies (Nycteribiidae) from Nigeria, West Africa. Vector-Borne Zoonotic Dis. 2014; 14:625–632. DOI: 10.1089/vbz.2013.1541 [PubMed: 25229701]

Kamath PL, Foster JT, Drees KP, Luikart G, Quance C, Anderson NJ, Clarke PR, Cole EK, Drew ML, Edwards WH, Rhyan JC, Treanor JJ, Wallen RL, White PJ, Robbe-Austerman S, Cross PC. Genomics reveals historic and contemporary transmission dynamics of a bacterial disease among wildlife and livestock. Nat Commun. 2016; 7:11448.doi: 10.1038/ncomms11448 [PubMed: 27165544]

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30:772–780. DOI: 10.1093/molbev/mst010 [PubMed: 23329690]

Kocher T, Thomas W, Meyer A, Edwards S, Pääbo S, Villablanca F, Wilson A. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc Natl Acad Sci. 1989; 86:6196–6200. DOI: 10.1073/pnas.86.16.6196 [PubMed: 2762322]

Kosoy MY. Ecological associations between bacteria of the genus *Bartonella* and mammals. Biol Bull. 2010; 37:716–724. DOI: 10.1134/S1062359010070071

Kosoy MY, Bai Y, Lynch T, Kuzmin IV, Niezgoda M, Franka R, Agwanda B, Breiman RF, Rupprecht CE. *Bartonella* spp. in bats, Kenya. Emerg Infect Dis. 2010; 16:1875–1881. DOI: 10.3201/eid1612.100601 [PubMed: 21122216]

La Scola B, Zeaiter Z, Khamis A, Raoult D. Gene-sequence-based criteria for species definition in bacteriology: the *Bartonella* paradigm. Trends Microbiol. 2003; 11:318–321. DOI: 10.1016/S0966-842X(03)00143-4 [PubMed: 12875815]

Legendre P, Desdevises Y, Bazin E. A statistical test for host-parasite coevolution. Syst Biol. 2002; 51:217–234. DOI: 10.1080/10635150252899734 [PubMed: 12028729]

Legendre P, Fortin MJ, Borcard D. Should the Mantel test be used in spatial analysis? Methods Ecol Evol. 2015; 6:1239–1247. DOI: 10.1111/2041-210X.12425

Lei BR, Olival KJ. Contrasting patterns in mammal-bacteria coevolution: *Bartonella* and *Leptospira* in bats and rodents. PLoS Negl Trop Dis. 2014; 8:e2738.doi: 10.1371/journal.pntd.0002738 [PubMed: 24651646]

Lin EY, Tsigrelis C, Baddour LM, Lepidi H, Rolain JM, Patel R, Raoult D. *Candidatus Bartonella mayotimonensis* and endocarditis. Emerg Infect Dis. 2010; 16:500–503. DOI: 10.3201/eid1603.081673 [PubMed: 20202430]

Lin JW, Hsu YM, Chomel BB, Lin LK, Pei JC, Wu SH, Chang CC. Identification of novel *Bartonella* spp. in bats and evidence of Asian gray shrew as a new potential reservoir of *Bartonella*. Vet Microbiol. 2012; 156:119–126. DOI: 10.1016/j.vetmic.2011.09.031 [PubMed: 22005177]

Lindenfors P, Nunn CL, Jones KE, Cunningham AA, Sechrest W, Gittleman JL. Parasite species richness in carnivores: effects of host body mass, latitude, geographical range and population density. Glob Ecol Biogeogr. 2007; 16:496–509. DOI: 10.1111/j.1466-8238.2006.00301.x

Lu L, Lycett SJ, Leigh Brown AJ. Determining the phylogenetic and phylogeographic origin of highly pathogenic avian influenza (H7N3) in Mexico. PLoS One. 2014; 9:e107330.doi: 10.1371/journal.pone.0107330 [PubMed: 25226523]

Luis AD, Hayman DT, O'Shea TJ, Cryan PM, Gilbert AT, Pulliam JR, Mills JN, Timonin ME, Willis CK, Cunningham AA, Fooks AR, Rupprecht CE, Wood JL, Webb CT. A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? Proc R Soc London Ser B, Biol Sci. 2013; 280:e20122753.

Luis AD, O'Shea TJ, Hayman DT, Wood JL, Cunningham AA, Gilbert AT, Mills JN, Webb CT. Network analysis of host-virus communities in bats and rodents reveals determinants of cross-species transmission. Ecol Lett. 2015; 18:1153–1162. DOI: 10.1111/ele.12491

Maganga GD, Bourgarel M, Vallo P, Dallo TD, Ngoagouni C, Drexler JF, Drosten C, Nakoune ER, Leroy EM, Morand S. Bat distribution size or shape as determinant of viral richness in African bats. PLoS One. 2014; 9:e100172.doi: 10.1371/journal.pone.0100172 [PubMed: 24959855]

Mannerings AO, Osikowicz LM, Restif O, Nyarko E, Suu-Ire R, Cunningham AA, Wood JL, Kosoy MY. Exposure to bat-associated *Bartonella* spp. among humans and other animals, Ghana. Emerg Infect Dis. 2016; 22:922–924. DOI: 10.3201/eid2205.151908 [PubMed: 27088812]

Mantel N. The detection of disease clustering and a generalized regression approach. Cancer Res. 1967; 27:209–220. DOI: 10.1038/214637b0 [PubMed: 6018555]

Morse SF, Olival KJ, Kosoy MY, Billeter SA, Patterson BD, Dick CW, Dittmar K. Global distribution and genetic diversity of *Bartonella* in bat flies (Hippoboscoidea, Streblidae, Nycteribiidae). Infect Genet Evol. 2012; 12:1717–23. DOI: 10.1016/j.meegid.2012.06.009 [PubMed: 22771358]

Nei, M.; Kumar, S. Molecular evolution and phylogenetics. Oxford University Press; New York: 2000.

Norman A, Regnery R, Jameson P, Greene C, Krause D. Differentiation of *Bartonella*-like isolates at the species level by PCR-restriction fragment length polymorphism in the citrate synthase gene. J Clin Microbiol. 1995; 33:1797–1803. [PubMed: 7545181]

Nunn CL, Dokey ATW. Ranging patterns and parasitism in primates. Biol Lett. 2006; 2:351–354. DOI: 10.1098/rsbl.2006.0485 [PubMed: 17148401]

O'Leary MA, Bloch JI, Flynn JJ, Gaudin TJ, Giallombardo A, Giannini NP, Goldberg SL, Kraatz BP, Luo ZX, Meng J, Ni X, Novacek MJ, Perini FA, Randall ZS, Rougier GW, Sargis EJ, Silcox MT, Simmons NB, Spaulding M, Velazco PM, Weksler M, Wible JR, Cirranello AL. The placental mammal ancestor and the post-K-Pg radiation of placentals. Science. 2013; 339:662–667. DOI: 10.1126/science.1229237 [PubMed: 23393258]

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara R, Simpson GL, Solymos P, Stevens MHH, Wagner H. vegan: community ecology package. 2015; doi: 10.4135/9781412971874.n145

Olival KJ, Dittmar K, Bai Y, Rostal MK, Lei BR, Daszak P, Kosoy MY. *Bartonella* spp. in a Puerto Rican bat community. J Wildl Dis. 2015; 51doi: 10.7589/2014-04-113

Pal U, Fikrig E. Adaptation of *Borrelia burgdorferi* in the vector and vertebrate host. Microbes Infect. 2003; 5:659–666. DOI: 10.1016/S1286-4579(03)00097-2 [PubMed: 12787742]

Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 2004; 20:289–290. DOI: 10.1093/bioinformatics/btg412 [PubMed: 14734327]

Paziewska A, Harris PD, Zwoli ska L, Bajer A, Si ski E. Recombination within and between species of the alpha proteobacterium *Bartonella* infecting rodents. Microb Ecol. 2011; 61:134–45. DOI: 10.1007/s00248-010-9735-1 [PubMed: 20740281]

Paziewska A, Si ski E, Harris PD. Recombination, diversity and allele sharing of infectivity proteins between *Bartonella* species from rodents. Microb Ecol. 2012; 64:525–36. DOI: 10.1007/s00248-012-0033-y [PubMed: 22419104]

Pebesma E, Bivand R. Classes and methods for spatial data in R. R News. 2005:5.

R Core Team. R: a language and environment for statistical computing. R Found. Statistical Comput; Vienna, Austria: 2015.

Raftery, A.; Newton, M.; Satagopan, J.; Krivitsky, P. Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. In: Bernardo, J.; Bayarri, M.; Berger, J., editors. Bayesian Statistics. Oxford University Press; Oxford: 2007. p. 1-45.

Reeves WK, Loftis AD, Gore JA, Dasch GA. Molecular evidence for novel bartonella species in *Trichobius major* (Diptera: Streblidae) and *Cimex adjunctus* (Hemiptera: Cimicidae) from two southeastern bat caves, U.S.A. J Vector Ecol. 2005; 30:339–341. [PubMed: 16599175]

Reeves WK, Rogers TE, Durden LA, Dasch GA. Association of *Bartonella* with the fleas (Siphonaptera) of rodents and bats using molecular techniques. J Vector Ecol. 2007; 32:118–122. [PubMed: 17633432]

Seabloom EW, Borer ET, Gross K, Kendig AE, Lacroix C, Mitchell CE, Mordecai EA, Power AG. The community ecology of pathogens: coinfection, coexistence and community composition. Ecol Lett. 2015; :e12418.doi: 10.1111/ele.12418

Streicker DG, Turmelle AS, Vonhof MJ, Kuzmin IV, McCracken GF, Rupprecht CE. Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. Science. 2010; 329:676–679. DOI: 10.1126/science.1188836 [PubMed: 20689015]

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013; 30:2725–2729. DOI: 10.1093/molbev/mst197 [PubMed: 24132122]

Teeling EC, Madsen O, Van den Bussche RA, de Jong WW, Stanhope MJ, Springer MS. Microbat paraphyly and the convergent evolution of a key innovation in Old World rhinolophoid microbats. Proc Natl Acad Sci. 2002; 99:1431–1436. DOI: 10.1073/pnas.022477199 [PubMed: 11805285]

Tsai YL, Chang CC, Chuang ST, Chomel BB. *Bartonella* species and their ectoparasites: selective host adaptation or strain selection between the vector and the mammalian host? Comp Immunol Microbiol Infect Dis. 2011; 34:299–314. DOI: 10.1016/j.cimid.2011.04.005 [PubMed: 21616536]

Veikkolainen V, Vesterinen EJ, Lilley TM, Pulliainen AT. Bats as reservoir hosts of human bacterial pathogen, *Bartonella mayotimonensis*. Emerg Infect Dis. 2014; 20:960–967. DOI: 10.3201/eid2006.130956 [PubMed: 24856523]

Vellend M. Conceptual synthesis in community ecology. Q Rev Biol. 2010; 85:183–206. DOI: 10.1086/652373 [PubMed: 20565040]

Vos M, Didelot X. A comparison of homologous recombination rates in bacteria and archaea. ISME J. 2008; 3:199–208. DOI: 10.1038/ismej.2008.93 [PubMed: 18830278]

Wilson, D.; Reeder, DM. Mammal species of the world: a taxonomic and geographic reference. Smithsonian Institution Press; Washington, D.C: 2005.

Zhu Q, Kosoy MY, Olival KJ, Dittmar K. Horizontal transfers and gene losses in the phospholipid pathway of *Bartonella* reveal clues about early ecological niches. Genome Biol Evol. 2014; doi: 10.1093/gbe/evu16

Highlights

- Patterns of diversification and lineage formation in *Bartonella* genotypes from bats are analyzed.

- The dominant process contributing to bartonella diversification is adaptation to bat hosts.

- Rates of host-switching in bartonella are constrained by bat phylogeny.

- Geographic overlap among bat hosts and divergent adaptation to arthropod vectors contribute less to bartonella diversification.
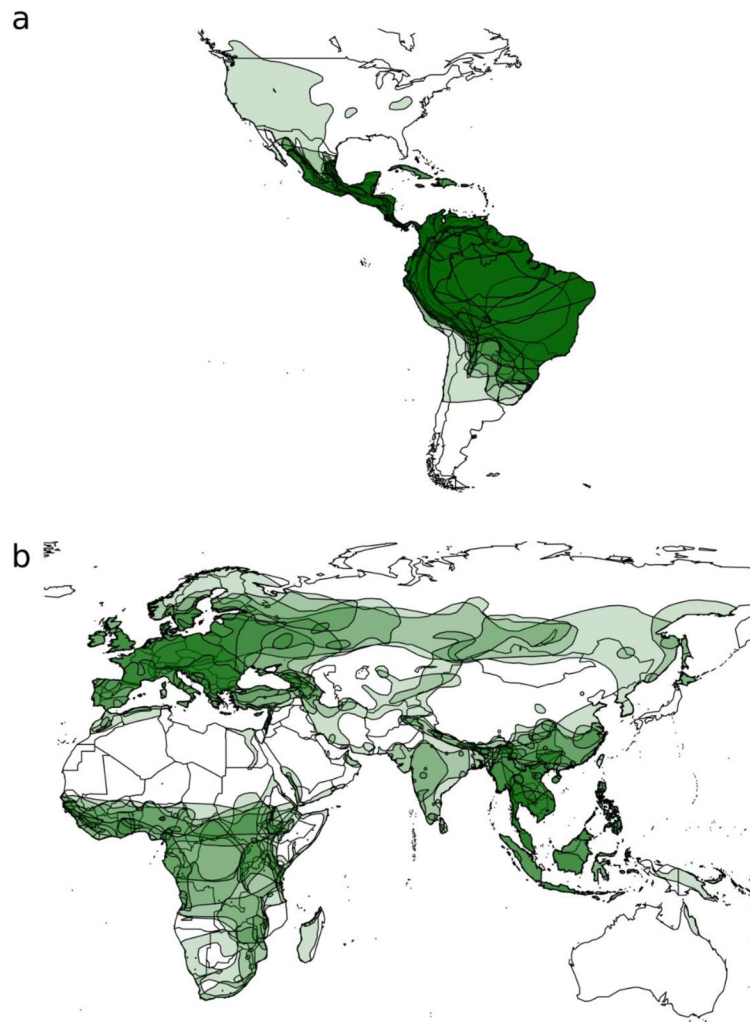
Fig. 1.
Geographic distributions of bat species represented in the study. Darker green areas show high levels of range overlap among sampled species, particularly within Central America, South America, and the Caribbean (a), and Europe, Africa, Central and East Asia, and Southeast Asia (b).

Fig. 2.
Bayesian phylogeny of *Bartonella* genotypes reconstructing bat host suborders, shown by colored branches. Posterior probabilities for nodes are shown as circles (●) scaled by size from 0 to 1 (Posterior.prob) and colored by the support for the bat host suborder at that node (Suborder.prob). Mean tree likelihood = −9981.92, ESS = 2713; mean suborder tree likelihood = −18.95, ESS = 1785. Details on tip labels for *Bartonella* genotypes and associated host species are listed in Table S1 and S2.
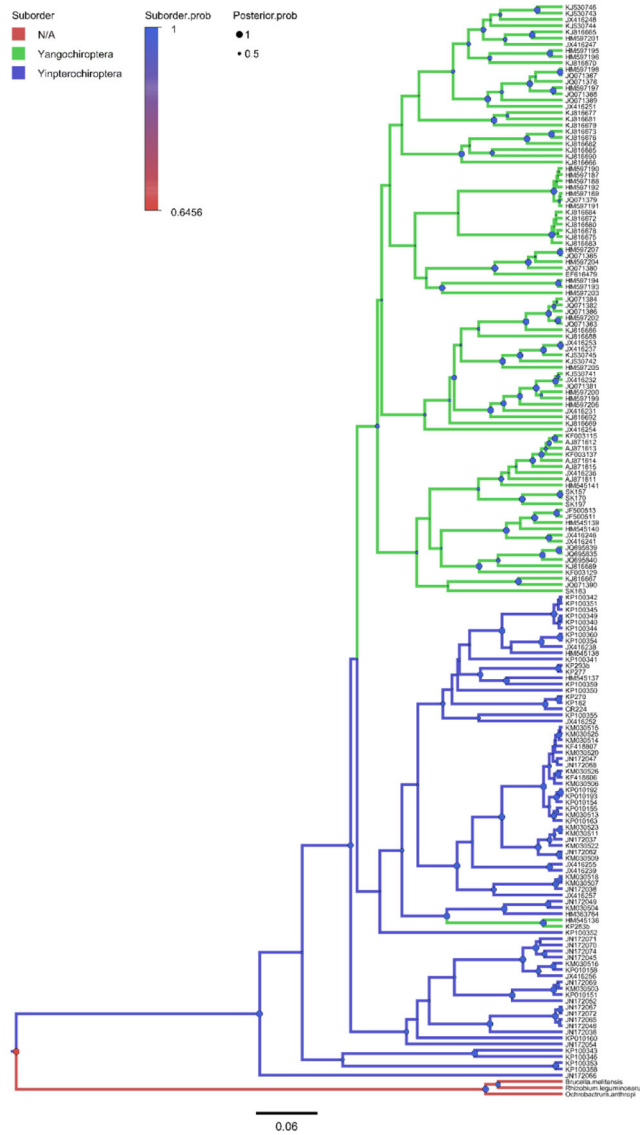
Fig. 3.
Bayesian phylogeny of *Bartonella* genotypes reconstructing bat host geographic regions, shown by colored branches. Posterior probabilities for nodes are shown as circles (●) scaled by size from 0 to 1 (Posterior.prob) and colored by the support for the bat host geographic region at that node (Region.prob). Mean tree likelihood = −9981.92, ESS = 2713; mean region tree likelihood = −95.84, ESS = 660. Details on tip labels for *Bartonella* genotypes and associated host species are listed in Table S1 and S2.

Fig. 4.
Bayesian phylogeny of *Bartonella* genotypes reconstructing bat host or arthropod vector associations, shown by colored branches. Posterior probabilities for nodes are shown as circles (●) scaled by size from 0 to 1 (Posterior.prob) and colored by the support for the bat host geographic region at that node (Vector.prob). Mean tree likelihood = −9981.92, ESS = 2713; mean vector tree likelihood = −131.17, ESS = 5741. Details on tip labels for *Bartonella* genotypes and associated host species are listed in Table S1 and S2.

Table 1

Posterior state transition rate estimates from the Bayesian analysis of bartonella *gltA* sequences, with data partitions for bat host family, superfamily, suborder, and geographic region. Only transition rates with a median rate greater than one are shown. Probability estimates indicate the likelihood of the median number of transition occurring since the time of the common ancestor of the 173 genotypes, as tested against a null gamma distribution (shape = 1, scale = 1). Bolded probability values are marginally significant (0.05 < P 0.1), underlined values are statistically significant (P < 0.05). Clock rates reflect the mean number of state transitions occurring across all nodes of the tree.

States	Median rate	95% HPD interval	Probability
Family transitions			
Emballonuridae – Pteropodidae	1.13	[0.01, 3.32]	0.32
Hipposideridae – Megadermatidae	2.13	[0.17, 5.13]	0.12
Hipposideridae – Pteropodidae	1.14	[0.01, 3.3]	0.32
Hipposideridae – Rhinolophidae	2.53	[0.31, 5.88]	**0.08**
Miniopteridae – Vespertilionidae	2.82	[0.46, 6.3]	**0.06**
Molossidae – Vespertilionidae	1.88	[0.06, 4.79]	0.15
Mormoopidae – Phyllostomidae	2.58	[0.4, 5.53]	**0.07**
Noctolionidae – Phyllostomidae	1.08	[6E-3, 3.16]	0.34
Phyllostomidae – Vespertilionidae	1.72	[0.14, 4.34]	0.18
Pteropodidae – Rhinolophidae	2.32	[0.26, 5.48]	**0.1**
Family clock rate	1.16	[0.69, 1.72]	
Superfamily transitions			
Emballonuroidea – Pteropoidea	1.11	[2E-4, 3.3]	0.33
Noctilionoidea – Vespertilionoidea	1.63	[0.11, 4.19]	0.2
Pteropoidea – Rhinolophoidea	2.77	[0.48, 5.84]	**0.06**
Superfamily clock rate	0.49	[0.21, 0.85]	
Suborder transitions			
Yangochiroptera – Yinpterochiroptera	1.41	[0.03, 3.44]	0.24
Suborder clock rate	0.16	[0.02, 0.36]	
Region transitions			
Africa – Asia S/E/SE	5.47	[2.16, 9.32]	<u>**0.004**</u>
Africa – Europe-Asia C/E	1.05	[5E-4, 3.16]	0.35
America C/S/Carib – Europe-Asia C/E	1.07	[5E-4, 2.91]	0.34
Asia S/E/SE – Europe-Asia C/E	1.14	[8E-5, 3.38]	0.32
Region clock rate	1.03	[0.61, 1.55]	
Host-vector transitions			
Bat - Ectoparasite	2.73	[0.29, 5.87]	**0.06**
Host-vector clock rate	2.03	[1.34, 2.76]	