

## Study

### The historical and contemporary processes driving global phylogenetic structure – the case of bat communities

#### Supplementary Information 3

#### Appendix S3. Assessing and correcting for the sampling frame size bias on indices of community phylogenetic relatedness

##### Sampling frame size-dependence in indices for phylogenetic community structure

The magnitude of standardized effect sizes of indices for phylogenetic community structure may increase with the size of the phylogeny used to compute the null models pertinent to these metrics (Sandel 2018). This bias requires that studies comparing standardized effect sizes of phylogenetic relatedness between communities to assess whether these comparisons hold after correcting for this problem.

One possible approach to correct for the sampling frame size bias is to repeatedly randomly subsample (*i.e.* rarefy) the community matrix to smaller sampling frame sizes, then to calculate the average indices for phylogenetic community structure [here, the net relatedness index (NRI) and the nearest taxon index (NTI); see Methods] across the rarefied subset (see Sandel 2018). The resulting index from this procedure should show a negligible relationship with the size of the sampling frame (or size or richness of the phylogeny), but at the expense of losing power intrinsic to the rarefaction procedure (Sandel 2018). The information from sampling frames that have less species than the size chosen for the subsampling procedure is also lost.

##### Implementation

To assess whether our results still held after correcting for the influence of the size of sampling frames in the standardized effect sizes index values for phylogenetic community structure, we recalculated the net relatedness index (NRI) and nearest taxon index (NTI) of bat communities applying two rarefaction procedures that differed in the number of species subsampled in each computation.

For the first procedure, we computed  $NRI_{raref_{LOW}}$  and  $NTI_{raref_{LOW}}$  by rarefying the community presence-absence matrix to the lowest species pool possible for that sampling frame extent, and then computing NRI and NTI on this matrix (which is the same approach used in Sandel, 2018).

In the second procedure, we calculated  $NRI_{raref_{REL}}$  and  $NTI_{raref_{REL}}$  by rarefying the community presence-absence matrix to one third of the number of species in the sampling frame of that region. This procedure allowed for a lower influence of the rarefaction procedure on the power, and allowed for less communities to be removed.

In both approaches, we repeated the rarefaction procedure 1000 times and reproduced the figures addressing the hypothesis testing from our manuscript using the separate averages of both rarefaction indices across their replicates. The interpretation of the rarefied indices follow the same ones for NRI and NTI (see Methods; Webb, 2000).

To demonstrate the influence of sampling frame size on the phylogenetic community structure indices and the robustness of both bias-correction approaches, we perform the same calculations on a simple set of simulations. We generated 10000 communities and a random pure-birth tree, both containing 1000 species, which represents the broad-scale sampling frame. We then subset 5000 communities and 250 species from the broad-scale sampling frame. We iteratively calculated NRI, NTI  $NRI_{raref_{FIX}}$ ,  $NTI_{raref_{FIX}}$ ,  $NRI_{raref_{REL}}$  and  $NTI_{raref_{REL}}$  for sampling frame sizes increasing from 250 species to 1000 species, in increments of 50 species.

This process was also performed in R and RStudio (R Core Team 2019, RStudio Team 2020), with the help of functions from the *ape* (Paradis and Schliep 2019), *picante* (Kembel et al. 2010), *PhyloMeasures* (Tsirogianis and Sandel 2017), *tidyverse* (Wickham et al. 2019) packages. See Appendix S1 for complementary information on tools used in this study.

## Results

As demonstrated in Sandel (2018), the traditional community phylogenetic relatedness indices (NRI and NTI) increased with the size of the phylogenetic sampling frame in our simulated communities [as shown by NRI (net relatedness index) and NTI (nearest taxon index); A and B]. However, the rarefaction-based calculation of phylogenetic relatedness effectively removes the sampling frame size-dependence in the simulations (see Figure SX) and in our empirical data set (see Figure SX).

The rarefied net relatedness index ( $NRI_{raref_{REL}}$ ) and nearest taxon index ( $NTI_{raref_{REL}}$ ) were expectedly closer to zero, but highly correlated with their respective unrarefied indices for phylogenetic relatedness. Moreover, all figures we reproduced with the rarefied NRI and NTI indices (Figures S3.2, S3.3, S3.4 and S3.5) had similar patterns to those we observed in the figures from the main manuscript (Figures 2, 3, 4

and 5). Hence, for the purposes of simplification and to avoid the loss of information introduced by the rarefaction procedure, we use the unrarefied values for NRI and NTI in the main manuscript.

## References for Supporting Information 4

Kembel, S. W. et al. 2010. Picante: R tools for integrating phylogenies and ecology. - *Bioinformatics* (Oxford, England) 26: 1463–1464.

Paradis, E. and Schliep, K. 2019. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. - *Bioinformatics* (Oxford, England) 35: 526–528.

R Core Team 2019. R: A Language and Environment for Statistical Computing. - R Foundation for Statistical Computing.

RStudio Team 2020. RStudio: Integrated development environment for r. - RStudio, PBC.

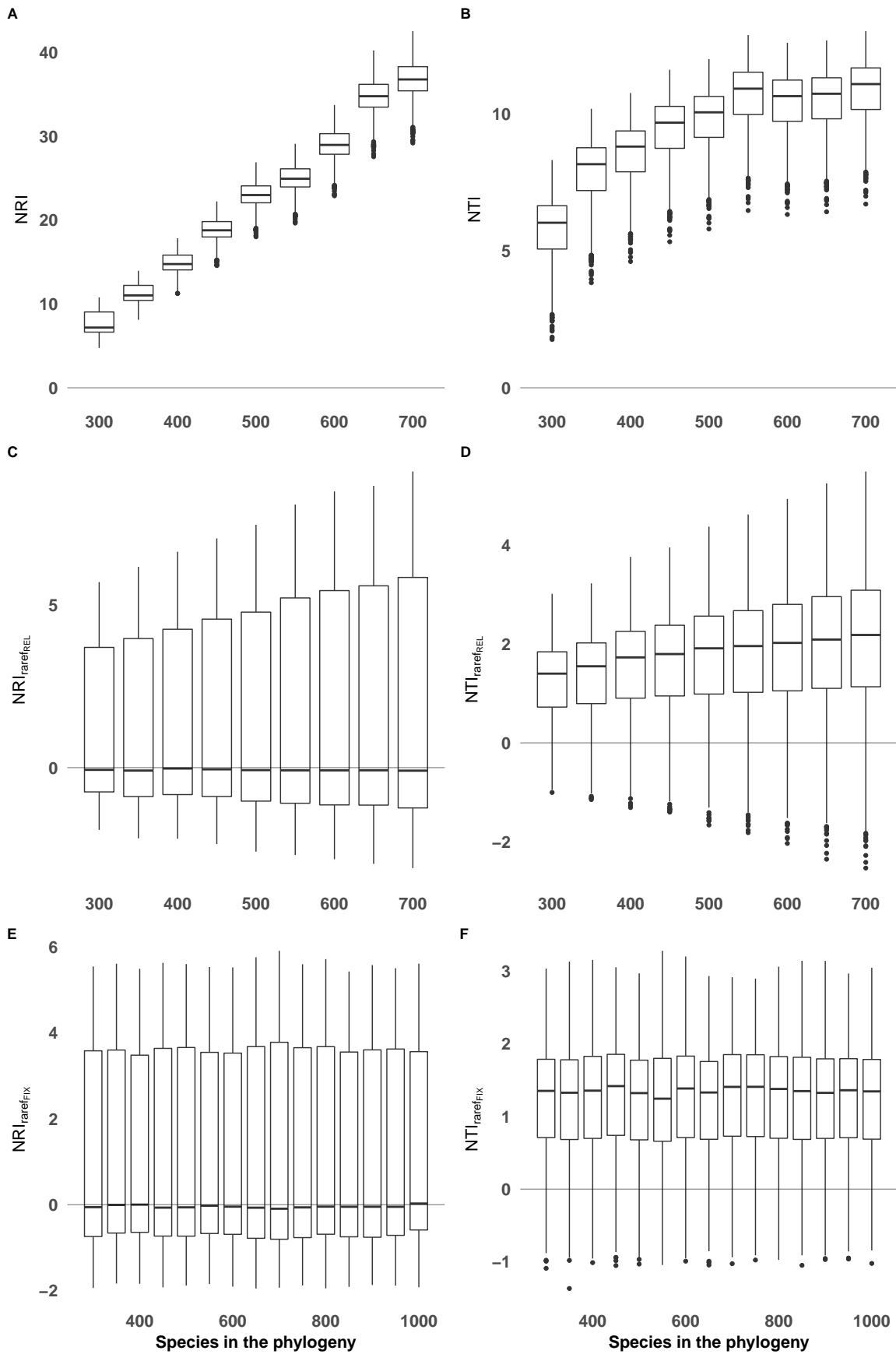
Sandel, B. 2018. Richness-dependence of phylogenetic diversity indices. - *Ecography* 41: 837–844.

Tsirogianis, C. and Sandel, B. 2017. PhyloMeasures: Fast and exact algorithms for computing phylogenetic biodiversity measures.

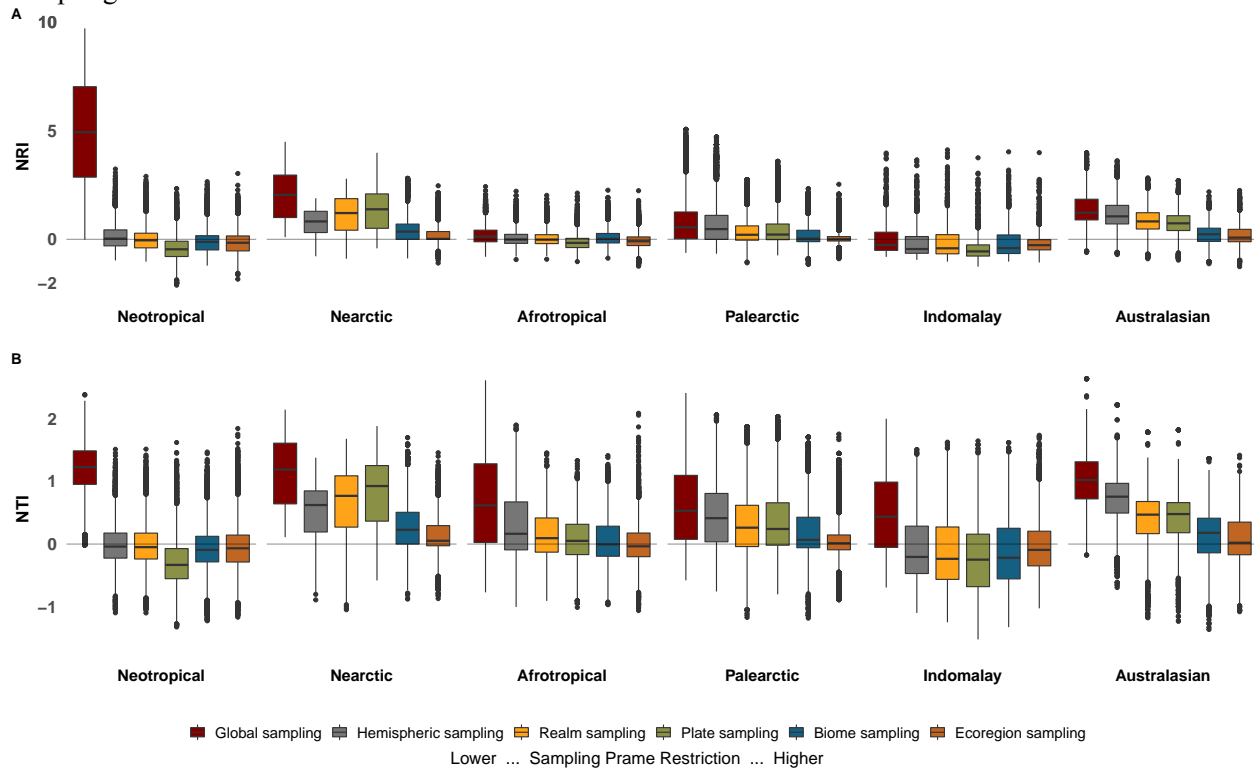
Wickham, H. et al. 2019. Welcome to the tidyverse. - *Journal of Open Source Software* 4: 1686.

## Figures

**Figure S3.1** The size of the phylogenetic sampling frame influences community phylogenetic relatedness index values of simulated communities [as shown by NRI (net relatedness index) and NTI (nearest taxon index); A and B]. Rarefaction with both fixed-size sub-sampling [ $NRI_{raref_{FIX}}$  and  $NTI_{raref_{FIX}}$ ; C and D] and relative-size sub-sampling [ $NRI_{raref_{REL}}$  and  $NTI_{raref_{REL}}$ ; E and F] removes the sampling frame size-dependence.

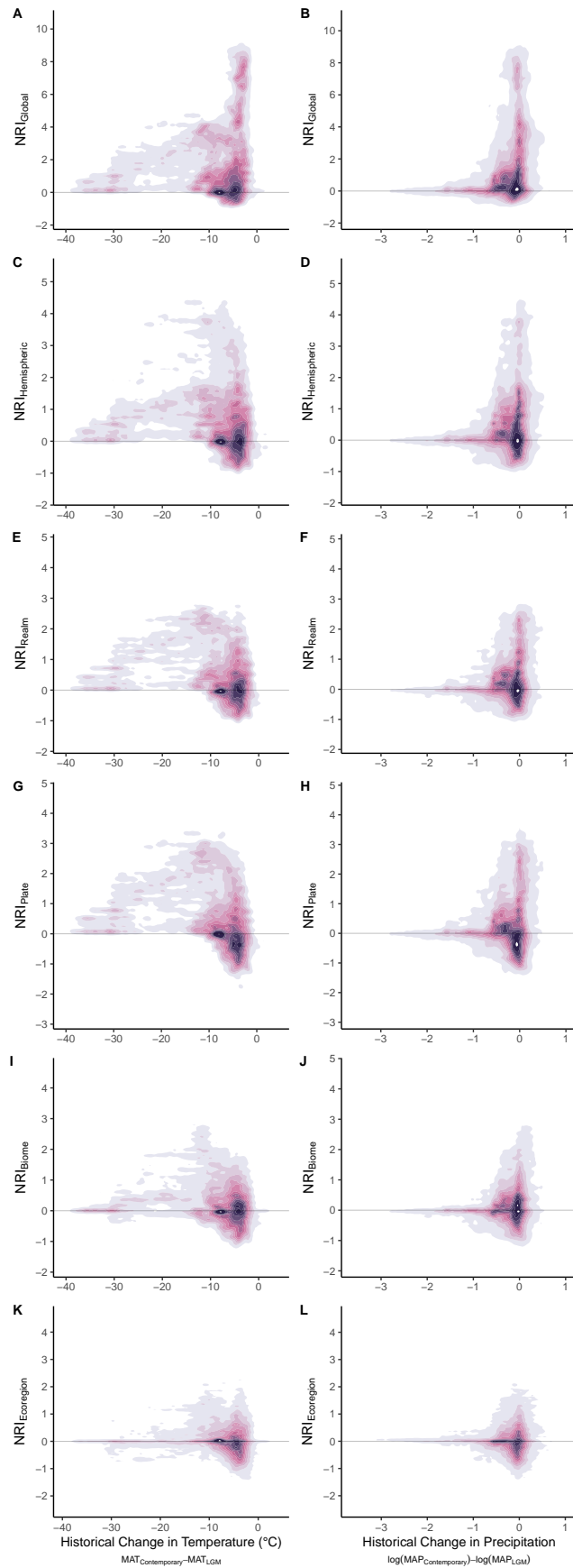


**Figure S3.2.** Realm-comparison of the sampling frame size bias-corrected phylogenetic structure of bat assemblages – measured through the (A) rarefied net relatedness index ( $NRI_{raref_{REL}}$ ) and (B) nearest taxon index ( $NTI_{raref_{REL}}$ ) – across a gradient of sampling frame spatial restrictions (see Methods). Sampling frames were restricted for the (i) global, (ii) east-west hemispheric (New World vs. Old World), (iii) biogeographical realm, (iv) tectonic plate, (v) within-realm biome, and (vi) ecoregional extents. Bat communities with positive values of  $NRI_{raref_{REL}}$  and  $NTI_{raref_{REL}}$  indicate that co-occurring species in these communities are phylogenetically related in relation to a given sampling frame extent. Conversely, negative values of NRI and NTI indicate that bat communities are mainly composed of distantly-related species in relation to that given sampling frame extent.



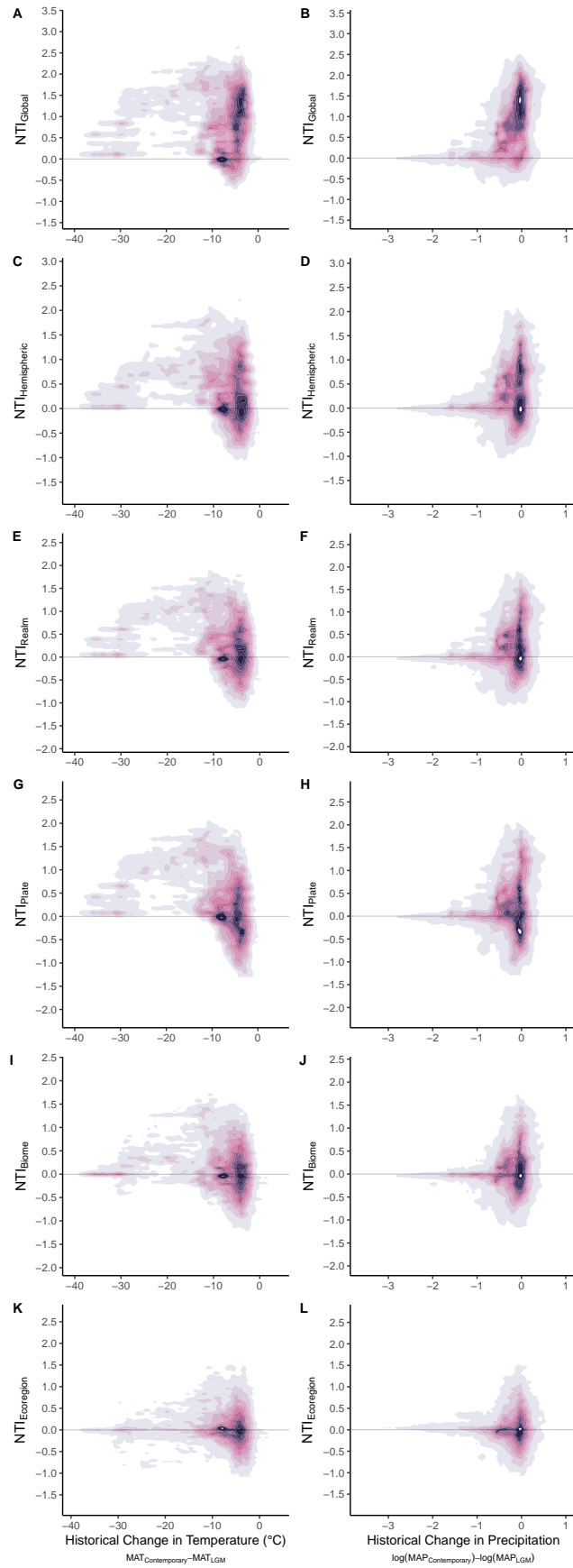
**Figure S3.3.** Probability contours for the kernel density estimation between paleoclimatic legacies and the sampling frame size bias-corrected rarefied net relatedness index ( $NRI_{raref}$ ) of worldwide bat communities across sampling frame extents. Paleoclimatic legacies for each bat community were obtained as the historical change in climate since the last glacial maximum (LGM); specifically, the mean annual temperature (MAT) or logarithm of the mean annual precipitation (MAP) from the contemporary period minus the estimated mean annual temperature or logarithm of the mean annual precipitation from 22,000 years ago (see Methods). Bivariate kernel density estimates are represented as the probability contours between the historical changes

94 in temperature (left-column) and precipitation (right-column and)  $NRI_{raref}$  was measured at the (A and B)  
95 global, (C and D) east-west hemispheric (New World vs. Old World), (E and F) biogeographical realm, (G  
96 and H) tectonic plate, (I and J) within-realm biome, and (K and L) ecoregional sampling frame extents.  
97 Lighter tones represent lower quantiles and less frequent combinations while darker tones represent higher  
98 quantiles of the probability distribution.



**Figure S3.4.** Probability contours for the kernel density estimation between paleoclimatic legacies and the sampling frame size bias-corrected rarefied nearest taxon index ( $NTI_{raref}$ ) of worldwide bat communities across sampling frame extents. Paleoclimatic legacies for each bat community were obtained as the historical change in climate since the last glacial maximum (LGM); specifically, the mean annual temperature (MAT) or logarithm of the mean annual precipitation (MAP) from the contemporary period minus the estimated mean annual temperature or logarithm of the mean annual precipitation from 22,000 years ago (see Methods). Bivariate kernel density estimates are represented as the probability contours between the historical changes in temperature (left-column) and precipitation (right-column) and  $NTI_{raref}$  was measured at the (A and B) global, (C and D) east-west hemispheric (New World vs. Old World), (E and F) biogeographical realm, (G and H) tectonic plate, (I and J) within-realm biome, and (K and L) ecoregional sampling frame extents. Lighter tones represent lower quantiles and less frequent combinations while darker tones represent higher quantiles of the probability distribution.





**Figure S3.5.** Relationships between community-weighted mean net diversification rates and the global phylogenetic structure of bat communities [represented by the (A) net relatedness index (NRI) and (B) nearest taxon index (NTI)]. Lines represent significant ( $P < 0.05$ ) quartic quantile regression estimates for the 5th (lower limits, in lighter tones), 15th, 25th, 35th, 45th, 55th, 65th, 75th, 85th and 95th (upper limits, in darker tones) quantile percentiles of net diversification rates (see Methods and Tables S3 and S4).

*[To be included]*