**Title: Fluctuating fortunes: genomes and habitat reconstructions reveal global climate mediated changes in bats’ genetic diversity**

**Authors:** Balaji Chattopadhyay1\*, Kritika M. Garg1, Rajasri Ray2,3 and Frank E. Rheindt1\*

1Department of Biological Sciences, National University of Singapore, Singapore.

2Center for Ecological Sciences, Indian Institute of Science, Bangalore - 560012, Karnataka, India.

3Centre for Studies in Ethnobiology, Biodiversity and Sustainability (CEiBa), B.G. Road, Mokdumpur, Malda-732103, West Bengal, India.

**SUPPORTING METHODS**

*Species occurrence points and climatic data*

Species occurrence records were collected from GBIF (accessed August 2017) as well as from the scientific literature (Table S2). Any occurrence records with issues as flagged by GBIF were removed. Available records were further manually checked for authenticity andduplicate values. Any occurrence records not within the known range of a species or which mapped into water bodies were removed. A total of 1718 occurrence records were used for the eleven species (Table S2): *Pteropus alecto* (211)*, Pteropus vampyrus* (75)*, Eidolon helvum* (398)*, Hipposideros armiger* (94)*, Rhinolophus ferrumequinum* (138)*, Rhinolophus sinicus* (12)*, Myotis brandtii* (286)*, Myotis davidii* (17)*, Eptesicus fuscus* (381)*, Megaderma lyra* (37) and *Pteronotus parnellii* (69).

We used ecological niche models to predict the suitable habitat at the present time, mid-Holocene (~6000 years ago), Last Glacial Maximum (LGM, ~ 20,000 years ago) and Last Interglacial period (LIG, ~110,000–130,000 years ago). Environmental variables were extracted from the WORLDCLIM database (version 1.4) [44]. We opted for the Community Climate System Model, version 4.0 (CCSM4), for the Holocene and the LGM. For the LIG, we used available data from Otto-Bliesner et al. [75], which have been frequently used in the literature [76, 77]. Further, we selected a subset of independent bioclimatic variables for ecological niche modelling. For this purpose, we extracted data for all 19 climatic variables from total occurrence points and log-transformed these values for further analysis. Next, we performed Pearson correlation analysis to group correlated variables (≥ 0.7) in R version 3.4.1 [78] (Table S3), from which we further selected variables through principal component analysis (PCA) as implemented in Past version 2.17 [79] with highest loading on principal components [80] (first five principal components explaining ~90% of variation) (Table S3).

In addition to using a global approach to bioclimatic variable selection, we used two other approaches for variable selection. The first of these two alternative approaches entailed a subgroup partitioning scheme according to continental occurrence: (1) America (2 species); (2) Africa (1 species); (3) Palearctic Region (2 species); (4) Australasia (6 species), followed by the reconstruction of ancient distributions for one representative species per subgroup (*Eptesicus fuscus, Eidolon helvum, Myotis brandtii, Pteropus alecto*). The second approach entailed a separation of all species individually (minus two species for which there were too few data points; *Myotis davidii, Rhinolophus sinicus*).

*Ecological niche modelling*

We performed ecological niche modelling in MaxEnt algorithm (version 3.3.3k) [52]. Feature selection was in accordance with the number of occurance points. For example, feature selection was linear when occurrence records were less than 10; linear and quadratic when occurrence records were between 10-14; linear, quadratic and hinge when occurrence records were between 15-79; and all features were selected when occurrence records were above 79 [53, 54]. We set the logistic output format based on probability of presence with the following colors: (0-0.1) pale yellow, (0.1-0.3) light green, (0.3-0.5) pale blue, (0.5-0.7) light blue, and (0.7-1) dark blue. We selected the mean representation of all 10 runs for further analysis of each species across each time period.

*Area estimation*

We calculated the area of suitable habitat for each time point considering the medium- to high- probability regions only (i.e. ~0.36 – 1, blue regions in Fig. 1). For area calculation, the grid area in tropical and subtropical regions was set at 1km2 (i.e. 30 arc second = ~1km2) for the present, Holocene and LIG periods. For the LGM, grid resolution was coarse (2.5 min.), so the area was calculated using a grid size of 20.25 km2 [44]. For temperate regions, we followed the available literature [81] in adjusting for latitudinal effects. For Europe, we used an average grid size of 0.536 km2 for the present, Holocene and LIG periods and 10.854 km2 for the LGM, for grids occupying ~350 - 700 N, whereas for North America and the Caribbean Islands, our grid size was averaged to 0.689 km2 for the present, Holocene and LIG periods and 13.95 km2 for the LGM for grids located between ~250 - 480 N.

**TABLES**

Table S1: Details of raw genome data, sampling, diet, forearm length and body mass for each study species (excel sheet).

Table S2: Location points used for niche modeling (Excel sheet).

Table S3: Principal components and their associated bioclimatic variables with highest loading values for niche modeling.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Principal component | Percentage of variance explained | Eigen value | Variable with highest loading value | Loading value for the selected variable |
| Principal component 1 | 53.285 | 10.1242 | Bio1 (annual mean temperature) | 0.3055 |
| Principal component 2 | 22.88 | 4.34724 | Bio17 (precipitation of driest quarter) | 0.4245 |
| Principal component 3 | 7.8221 | 1.48619 | Bio18 (precipitation of warmest quarter) | 0.5784 |
| Principal component 4 | 6.4322 | 1.22211 | Bio2 (mean diurnal range) | 0.4401 |
| Principal component 5 | 2.8553 | 0.542506 | Bio2 (mean diurnal range) | 0.5222 |
| Bio3 (isothermality) | 0.4068 |

Table S4: Areas under the curve at different time points estimated for both training and test datasets.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Species | Area under the curve | | | | | |
| Holocene | | Last Glacial Maximum | | Last Interglacial | |
| Training dataset | Test dataset | Training dataset | Test dataset | Training dataset | Test dataset |
| *Pteropus alecto* | 0.99 | 0.985 | 0.989 | 0.992 | 0.991 | 0.982 |
| *Pteropus vampyrus* | 0.983 | 0.98 | 0.982 | 0.986 | 0.985 | 0.979 |
| *Eidolon helvum* | 0.964 | 0.946 | 0.959 | 0.953 | 0.962 | 0.944 |
| *Pteronotus parnellii* | 0.987 | 0.988 | 0.989 | 0.985 | 0.992 | 0.979 |
| *Megaderma lyra* | 0.937 | 0.908 | 0.934 | 0.909 | 0.917 | 0.827 |
| *Hipposideros armiger* | 0.948 | 0.966 | 0.947 | 0.957 | 0.952 | 0.949 |
| *Eptesicus fuscus* | 0.948 | 0.944 | 0.952 | 0.929 | 0.949 | 0.936 |
| *Rhinolophus ferrumequinum* | 0.991 | 0.991 | 0.99 | 0.988 | 0.993 | 0.96 |
| *Rhinolophus sinicus* | 0.961 | 0.955 | 0.961 | 0.977 | 0.963 | 0.9 |
| *Myotis davidii* | 0.984 | 0.977 | 0.985 | 0.978 | 0.98 | 0.957 |
| *Myotis brandtii* | 0.983 | 0.971 | 0.982 | 0.988 | 0.983 | 0.982 |

Table S5: Permutation importance of each bioclimatic variable to the ecological niche models for each time point. Values in bold indicate the most important bioclimatic variable. Bio1: annual mean temperature; Bio2: mean diurnal range; Bio3: isothermality; Bio17: precipitation of driest quarter; Bio18: precipitation of warmest quarter.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Species and niche | Time period | Bio1 | Bio2 | Bio3 | Bio17 | Bio18 |
| *Pteropus alecto*, frugivore | Holocene | 6 | 1.1 | **58** | 10.2 | 24.7 |
| Last Glacial Maximum | 4.1 | 1.6 | **74.2** | 6.2 | 13.9 |
| Last Interglacial | 5.2 | 0.4 | **68.8** | 7.7 | 18 |
| *Pteropus vampyrus*, frugivore | Holocene | 0.9 | 1.4 | **96.3** | 0.4 | 1 |
| Last Glacial Maximum | 0.1 | 0.7 | **98** | 1.1 | 0.2 |
| Last Interglacial | 1.6 | 0.4 | **96.1** | 1.6 | 0.3 |
| *Eidolon helvum*, frugivore | Holocene | 4.2 | 15 | **61.6** | 4.3 | 14.9 |
| Last Glacial Maximum | 3.2 | 16.5 | **64** | 3 | 13.4 |
| Last Interglacial | 3.6 | 13.1 | **63.8** | 3.7 | 15.8 |
| *Pteronotus parnellii*; insectivore | Holocene | 15.4 | 18.7 | **51.7** | 10.5 | 3.8 |
| Last Glacial Maximum | 16.2 | 13.2 | **51.1** | 16.2 | 3.3 |
| Last Interglacial | 2.7 | 4.5 | **75.3** | 17.1 | 0.5 |
| *Megaderma lyra*; insectivore and carnivore | Holocene | **84.7** | 9.9 | 0.1 | 1.4 | 3.9 |
| Last Glacial Maximum | **69.5** | 14.2 | 2.4 | 12.5 | 1.4 |
| Last Interglacial | **84.6** | 8.2 | 0 | 5.2 | 2.1 |
| *Hipposideros armiger*; insectivore | Holocene | 23.7 | 16.5 | 16.5 | 9.7 | **33.7** |
| Last Glacial Maximum | 17.4 | 6 | 16.9 | 10.1 | **49.6** |
| Last Interglacial | 6.7 | 15 | 20.5 | 17.1 | **40.7** |
| *Eptesicus fuscus*; insectivore | Holocene | **81** | 6.8 | 2.6 | 3 | 6.6 |
| Last Glacial Maximum | **86.1** | 3.6 | 3.4 | 2.2 | 4.7 |
| Last Interglacial | **83.6** | 7.3 | 3.7 | 1.5 | 3.8 |
| *Rhinolophus ferrumequinum*; insectivore | Holocene | 23.5 | **49.5** | 13 | 6.5 | 7.5 |
| Last Glacial Maximum | 10.7 | **41.5** | 30.5 | 9.9 | 7.4 |
| Last Interglacial | 12.9 | 9.7 | 4.2 | **42.7** | 30.5 |
| *Rhinolophus sinicus*; insectivore | Holocene | 40.2 | **47.4** | 0 | 9.3 | 3.1 |
| Last Glacial Maximum | **51.1** | 39.6 | 0 | 9.2 | 0.1 |
| Last Interglacial | 23.1 | **56.5** | 0 | 20.4 | 0 |
| *Myotis davidii*; insectivore | Holocene | 30.2 | 0.2 | 0 | 3.3 | **66.2** |
| Last Glacial Maximum | **66.3** | 0 | 2.9 | 7.3 | 23.5 |
| Last Interglacial | **79.6** | 3.4 | 1.3 | 1.3 | 14.5 |
| *Myotis brandtii*; insectivore | Holocene | **50** | 33.9 | 4.2 | 5.4 | 6.5 |
| Last Glacial Maximum | **69.1** | 15.3 | 4 | 0.4 | 11.2 |
| Last Interglacial | **70.5** | 16.9 | 2 | 0.7 | 9.9 |

**REFERENCES**

1. Otto-Bliesner BL, Marshall SJ, Overpeck JT, Miller GH, Hu A. 2006 Simulating Arctic climate warmth and icefield retreat in the last interglaciation. *Science* **311**, 1751–1753. (doi: 10.1126/science.1120808)
2. Leipold M, Tausch S, Poschlod P, Reisch C. 2017 Species distribution modeling and molecular markers suggest longitudinal range shifts and cryptic northern refugia of the typical calcareous grassland species *Hippocrepis comosa* (horseshoe vetch). *Ecol. Evol*. **7**, 1919–1935. (doi: 10.1002/ece3.2811)
3. Raes N, Cannon CH, Hijmans RJ, Piessens T, Saw LG, van Welzen PC, Slik JF. 2014 Historical distribution of Sundaland’s Dipterocarp rainforests at Quaternary glacial maxima. *Proc. Natl. Acad. Sci. USA* **111**, 16790–16795. (doi: 10.1073/pnas.1403053111)
4. R Core Team. 2016 R: A language and environment for statistical computing.
5. Hammer Ø, Harper D, Ryan P. 2001 PAST-Palaeontological statistics. *www. uv. es/~ pardomv/pe/2001\_1/past/pastprog/past. pdf, acessado em*. **25**, 2009.
6. Snelder D, Weerd M, Zelfde M, Tamis W. 2013 Modelling the impact of climate and land use changes on forest bird species for adaptive management of the Northern Sierra Madre Natural Park (Philippines).
7. Daac L. 2004 Global 30 Arc-Second Elevation Data Set GTOPO30. Land Process Distributed Active Archive Center. URL: <http://edcdaac.usgs.gov/gtopo30/gtopo30.asp>
8. Herd RM. 1983 *Pteronotus* *parnellii*. *Mammalian Species*, 1–5.
9. You Y, Sun K, Xu L, Wang L, Jiang T, Liu S, Lu G, Berquist SW, Feng J. 2010 Pleistocene glacial cycle effects on the phylogeography of the Chinese endemic bat species, *Myotis* *davidii*. *BMC Evol. Biol.* **10**, 208 (doi:10.1186/1471-2148-10-208)

**SUPPORTING FIGURES**

Fig. S1: Temporal fluctuations in effective population size assuming a generation time of two years and a mutation rate of 2.2 X 10-9 per base pair per year. Estimates based on complete data and bootstraps are depicted in dark and light red color, respectively.

Fig. S2: Temporal fluctuations in effective population size assuming a generation time of one year and a mutation rate of 2.2 X 10-9 per base pair per year. Estimates based on complete data and bootstraps are depicted in dark and light red color, respectively.

Fig. S3: Temporal fluctuations in effective population size assuming a generation time of eight years and a mutation rate of 2.2 X 10-9 per base pair per year. Estimates based on complete data and bootstraps are depicted in dark and light red color, respectively.