THE UNIVERSITY OF CHICAGO

REGIONAL INFLUENCES ON COMMUNITY STRUCTURE ACROSS THE TROPICAL-TEMPERATE DIVIDE

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF ECOLOGY AND EVOLUTION

BY
ALEXANDER EDWARDS WHITE

CHICAGO, ILLINOIS
AUGUST 2018

Contents

List of Figures iv

List of Tables vi

Acknowledgments vii

Introduction 1

1 Geographical barriers and dispersal propensity interact to limit range expansions of Himalayan birds 3

- 1.1 Abstract 3
- 1.2 Introduction 4
- 1.3 Methods 9
  - 1.3.1 Species 9
  - 1.3.2 Morphological measurements of dispersal 10
  - 1.3.3 Topographic complexity 11
  - 1.3.4 Climatic niche similarity 12
  - 1.3.5 Statistical analysis of range limits 13
  - 1.3.6 Indirect effects of dispersal on sister species pairs 14
- 1.4 Results 15
  - 1.4.1 Limits to range expansion – climate tracking 15
  - 1.4.2 Intrinsic limits to range expansion – wing morphology 16
  - 1.4.3 Extrinsic limits to range expansion – topographic complexity 18
  - 1.4.4 Interactions between intrinsic and extrinsic limits 18
  - 1.4.5 Effects of dispersal within habitats 21
  - 1.4.6 Superspecies are intermediate dispersers 25
- 1.5 Discussion 26

2 Regional influences on community structure across the tropical-temperate divide 32

- 2.1 Abstract 32
- 2.2 Introduction 33
- 2.3 Results and Discussion 34
- 2.4 Methods 44
  - 2.4.1 The Grade of Membership model 44
  - 2.4.2 Species motifs 45
  - 2.4.3 Species abundance motifs 46
  - 2.4.4 Phylogenetic motifs 48
  - 2.4.5 Geographical motifs 48
  - 2.4.6 Trait motifs 49
  - 2.4.7 Modeling considerations 50
  - 2.4.8 Model comparisons 51
  - 2.4.9 Comparison to other clustering methods 51
  - 2.4.10 Field methods 52
  - 2.4.11 Code availability 53

3 Quantifying geographic and phylogenetic structure in birds and mammals across Wallace's line 64
- 3.1 Abstract 64
- 3.2 Introduction 64
- 3.3 Methods 68
  - 3.3.1 Species presences 68
  - 3.3.2 Estimating biotas in the Indo-Pacific 69
  - 3.3.3 Phylogenetic motifs 70
- 3.4 Results 71
  - 3.4.1 Species motifs 71
  - 3.4.2 Phylogenetic comparison for $K = 2$ 75
  - 3.4.3 Phylogenetic motifs 76
- 3.5 Discussion 78

Data Accessibility 81
- Geographical barriers and dispersal propensity interact to limit range expansions of Himalayan birds 81
- Regional influences on community structure across the tropical-temperate divide 81
- Quantifying geographic and phylogenetic structure in birds and mammals across Wallace's line 81

References 82

Appendix 93
- Figure S1 93
- Figure S2 93
- Table S1 93
- Table S2 93
- Table S3 93

List of Figures

1 Map of India showing the topography of the regional study area .... pg. 7

2 Regional comparisons of climatic differences and landscape complexity between the southeast and northwest Himalaya. .... pg. 15

3 Boxplot of hand-wing index within foraging strata.. pg. 17

4 Interaction between wing morphology and least cost distance affects northern range limits. .... pg. 21

5 Recent allopatric divergence is predicted by both intermediate dispersal ability and complex topography .... pg. 25

6 Species motifs for breeding birds are organized by freezing temperatures .... pg. 35

7 Assignment of regional features to 38 local communities of Himalayan birds ($K = 2$ motifs) .... pg. 38

8 Motif contributions to local communities when $K = 4$. .... pg. 40

9 Graphical illustration of the method used to assign species in a localized area or community to geographical motifs .... pg. 54

10 Map cell contributions to motifs defined by species ranges .... pg. 55

11 Dispersion fields for four communities .... pg. 56

12 Motifs for $K = 2$ - 5. .... pg. 57

13 Climate across the Himalaya .... pg. 58

14 Screen-shot of interactive network of species co-occurrences. .... pg. 59

15 Species motifs generated using a presence-absence model for $K = 2$ .... pg. 60

16 Phylogenetic motifs at the 10 Ma and 20 Ma timelines .... pg. 61

17 Elevational and geographic similarities of communities .... pg. 62

18 Species richness and phylogenetic alpha diversity in communities .... pg. 63

19 Species motifs for birds and mammals in the Indo-Pacific, $K = 2$. .... pg. 72

20 Species motifs for birds and mammals in the Indo-Pacific, $K = 10$. .... pg. 74

21 Phylogenetic diversity of assemblages sampled from the membership proportions vector $\theta_{k.}$ for each of the two motifs.......... pg. 75

22 Phylogenetic motifs for birds and mammals, $K = 2$. .... pg. 76

22 Phylogenetic motifs for birds and mammals, $K = 10$. .... pg. 77

S1 Interactive network of species co-occurrences.......... pg. 93

S2 Comparisons with other methods to project similarities of local communities.......... pg. 93

List of Tables

1 Correlations among main effects and maximum northern range limit .... pg. 16

2 Phylogenetic generalized least-squares regression showing the effects of landscape and morphology on the northern range limit of southeastern species .... pg. 19

3 Phylogenetic generalized least-squares regression using niche availability instead of niche similarity .... pg. 20

4 Phylogenetic generalized least-squares regression showing the effects of landscape and morphology on northern range limit of southeastern species among foraging strata. pg. 23

5 Models to predict northern range limits of Himalayan birds .... pg. 24

S1 Site metadata for locations of Himalayan bird community surveys.......................... pg. 93

S2 Species contributions to Himalayan species abundance motifs .... pg. 93

S3 Species abundances in local Himalayan communities .... pg. 93

Acknowledgments

This work would not have been possible without the invaluable intellectual, financial, logistical, and personal contributions from a number of individuals, both at the University of Chicago and abroad. Foremost, I would like to thank my thesis advisor, Dr. Trevor Price, for his innumerable contributions to the research presented here – it is very safe to say that this work would have been impossible without his guidance and support. I also owe a great debt to members of the Price lab, both past and present, for their support and critical feedback while developing and executing this research. I would like to thank my thesis committee for their contributions, in particular for their willingness to allow me the latitude to pursue a highly collaborative and multidisciplinary thesis – Dr. Tim Wootton, Dr. John Bates, Dr. Matthias Steinrücken, and Dr. Richard Ree. This committee not only contributed to this research but also to my development as a scientist, and for their guidance I am very grateful. Three other faculty members at University of Chicago deserve notice for their contributions – Dr. David Jablonski, Dr. Catherine Pfister, and Dr. Mathew Stephens. Each of these individuals contributed greatly to the work presented here, particularly through their critical feedback as this research took shape. Finally, I must thank Dr. Sally Keith and Dr. Carsten Rahbek for their invitation to study at the Center for Macroecology Evolution and Climate at the University of Copenhagen – the time I spent in Denmark was invaluable to my development as a scientist.

I have also had the privilege of collaborating with a number of individuals, all of whom have been great collaborators and even better people – Dr. Sally Keith, Dr. Dhananjai Mohan, Dr. Suresh Rana, Dr. Jon Kennedy, Matthew Schumm, Stewart Edie, K. Supriya, and a number of others in the Jablonski Lab and at the Center for Macroecology Evolution and Climate in Denmark. I must single out the contributions my close collaborator Dr. Kushal Dey, with whom I

developed the methods for Chapters 2 and 3 and with whom I have excitedly pursued publication of these chapters. I have learned a great deal from Dr. Dey, foremost of which is the incredible value of collaboration. I am very grateful for his patience and his willingness to pursue these projects as a team.

I have been supported by a number of different institutions and organizations while at the University of Chicago and I must thank them for their financial support – the National Science Foundation, the National Geographic Society, the U. S. Department of Education, the Danish National Research Foundation, and the University of Chicago. This financial support allowed me to pursue my work in India, Denmark, and Chicago with immense freedom, and for that support I am very grateful. I must also acknowledge the logistical support from curators and collections managers for birds at the Field Museum of Natural History – Dr. Shannon Hackett, Dr. John Bates, Dr. Ben Marks, and Dr. Dave Willard.

Finally, there are many people who have had a great impact on me personally and professionally throughout my time pursuing this research: Dr. Joel Smith, Dr. Shane Dubay, Dr. Daniella Palmer, Dr. Natasha Bloch, Dr. Ben Winger, Dr. Daniel Hooper, Dr. Simon Lax, Dr. Victoria Flores, Dr. Evan Koch, Dr. Elizabeth Sander, Dallas Krentzel, Mo Siddiq, Roberto Marquez, and Natalia Piland. I am also very grateful for the support of my family, and most importantly, the unfailing support of my partner, Dr. Abby Brown.

Introduction

That organisms adhere to some characteristic spatial distribution is one of the fundamental natural history observations of ecology. Indeed, identifying the proximate mechanisms generating those distributions – either within a forest, on a mountain, or across a continent – has remained a primary goal for biologists since Humboldt (Humboldt and Bonpland 1807). Evolutionary biology, biogeography, and community ecology each offer unique explanations for the factors limiting the geographic distributions of species (see Chapter 1 for a discussion). However, disparate data sources and different temporal and spatial scales have long hindered a coherent synthesis (Ricklefs 2004; Harrison and Cornell 2007).

Parallel to the debate regarding the limits of species ranges, is one regarding the limits of groups of species or communities. Ricklefs (2008) suggested that community level influences are perhaps unimportant, that the coarse scale geographic influences on species ranges outweigh the local scale influences of community level interactions; others contend that local interactions are of greater consequence (Harrison and Cornell 2008). While this debate has largely focused on the *number* of species in a given locale, one can easily imagine how those numbers are directly related to the identities of the species in question. And given that the number of species in a given location is so well correlated with latitude, these debates generally beg the question of whether the temperate or tropical associations of individual species drive the organization of species into communities.

Much of biology has traditionally hinged on classifying communities or biotas into discrete groups, such as asking if communities are tropical *or* temperate, when in reality, assemblages may contain elements of both (species ranges themselves may me mixed as well).

These methods, from Wallace to the present, have all led to the application of hard boundaries, whereby biotas or communities can only belong to one pre- defined category even if they would be better explained by a mixing of two or more. Recent network-based advances (Vilhena and Antonelli 2015) can identify these “mixed” assemblages, but they are unsatisfyingly assigned to a grouping all their own, falling short of the quantification required to evaluate the underlying influences of continuous ecological gradients. In this dissertation, I confront this longstanding problem by developing a method that directly quantifies mixing, and thus the contributions of each fundamental grouping in the classification of interest.

Today, the rapid proliferation of data documenting species’ geographic distributions (BirdLife International and NatureServe 2014; IUCN 2017), their underlying phylogenetic relationships (Fritz et al. 2009; Jetz et al. 2012), and their ecological traits has begun to lift the veil obscuring the fundamental biological mechanisms organizing species geographically (Zanne et al. 2014). In the three chapters of this thesis, I combine these types of data with local field surveys to assess the composition of local communities that span the tropical-temperate boundary and assess their associations to the surrounding region. I develop a framework for understanding species’ geographic associations using a wide diversity of data-types by leveraging a novel machine learning method that has so far not been developed in the context of studying biogeography (although see Valle et al. (2014) for an early discussion of its potential applications in ecology). Most of this research is focused on birds of the Himalayas, the second most diverse location on earth for avian life. The final chapter focuses on composition of bird and mammal assemblages of the Indo-Pacific, in particular evaluating the composition of these groups in relation to the famous biogeographic partition designated by Wallace (Wallace 1863).

1 Geographical barriers and dispersal propensity interact to limit range expansions of Himalayan birds[1]

1.1 Abstract

Range expansions are limited by two key factors. These are (1) dispersal, which includes a species' intrinsic mobility, geographical barriers, and their interaction and (2) the ability of a species to persist beyond its current range. I evaluate the role of these in affecting bird species distributions across the Himalayas, under a hypothesis that many species have recently expanded their range out of an eastern Pleistocene refuge. I measured wing shape as a proxy for dispersal ability and topographic complexity across the Himalayas as a proxy for dispersal barriers. As a factor affecting the potential for persistence in novel locations, I compared similarity of a species climatic envelope in the east, the hypothesized historical refuge, and the west, the location of recent colonization. Climatic similarity, wing shape, and the interaction of topographic complexity with wing shape all contribute significantly to the range extent of a given species. The result highlights the important interaction between morphological and landscape factors in affecting successful range expansions. The two dispersal-related parameters together explain four times the variance explained by climate, but I present additional evidence that other factors besides climate, notably biotic interactions, affect the ability of a species to persist beyond its range.

---

[1] A version of this chapter has been published as: White, A.E. 2016. Geographical barriers and dispersal propensity interact to limit range expansions of Himalayan birds. American Naturalist. 188: 99-112.

1.2 Introduction

Climatic perturbations result in large shifts in species geographic ranges (Davis 1986). During periods of climatic amelioration two forces govern range expansions (Davis and Shaw 2001; Sandel et al. 2011). First, dispersal (defined here as movement of individuals across space; Ronce 2007) is critical for individuals to even reach a new locality. Second, successful range expansions require dispersing individuals to establish and persist in the new location. Various factors contribute to persistence, both abiotic (Parmesan and Yohe 2003) and biotic (Wethey 2002; Briers 2003; Garrick et al. 2013). Persistence factors are known to play a major role in driving range expansions, i.e., dispersal is not the only requirement (reviewed in Sexton et al. 2009). For example, niche models have been used to implicate the strong influence of climatic variables in setting range limits (Pounds et al. 1999; Pigot et al. 2010) and in driving historical range movements over timescales ranging from the Pleistocene (Assis et al. 2014) to the last few decades (Parmesan and Yohe 2003; Sandel et al. 2011). Climate has also been related to recolonization of a species' habitat (Menéndez et al. 2006) with the focal species tracking habitat. These are factors that affect persistence in a new location, but ease of dispersal across the landscape also affects range expansions (Pearson and Dawson 2003).

Assessing the role of dispersal limitation has been particularly informative for understanding the formation of island communities, which are a more mobile subset of mainland species (Kadmon 1995) and not necessarily the most suitable species based on local conditions. Remote islands, such as Hawaii (with no terrestrial mammals or tortoises), New Zealand (with no terrestrial mammals), or the Galápagos Islands (with no woodpeckers) have shown that more mobile forms evolve into niches that are otherwise occupied on continents (James and Burney 1997; Grant and Grant 2008; Wood et al. 2008). Thus, dispersal limits on range expansions influence the interactions between species, the process of community assembly, and the evolutionary dynamics of community level processes. Importantly, if dispersal limitation is strong, then recovery from recent disturbances such as glaciations takes time, and even continental communities may continue to be assembled by newly arriving species (Ghosh-Harihar and Price 2014; Moen and Morlon 2014). In this paper, I focus on the effects of dispersal limitation on range recolonization of birds since the last glacial maximum in a continental, mountainous landscape. Vagile groups such as birds are generally thought to more easily achieve climatic equilibrium (Sandel et al. 2011). However, barriers do limit bird dispersal, e.g., among tropical groups (Burney and Brumfield 2009; Smith et al. 2014), in aseasonal environments (Quintero et al. 2014), and between islands (Diamond 1981).

Dispersal itself is affected in two main ways – the intrinsic dispersal propensity of a species (e.g., the morphology of a species locomotor function) and landscape complexity, including both topographical barriers and habitat (Pearson and Dawson 2003). Intrinsic and landscape factors both affect dispersal, but they are not independent. For example, a small water barrier or forest break may be a great impediment to a flightless species, but not even recognized as a barrier by a more dispersive one. Normand et al. (2011) found not only that post-Pleistocene range recolonization of some European plants was limited by the accessibility of new habitats, but that the correlation between accessibility and range limits greatly increased for those species carrying few traits associated with long distance dispersal.

In this paper, I examine the role of dispersal in post-Pleistocene range recolonization of a species rich regional assemblage—Himalayan birds. The climate of the Himalayas is strongly influenced by the combined effects of the south Asian summer monsoon and the Eurasian westerlies (Benn and Owen 1998). Historical Quaternary fluctuations of these weather systems

(Overpeck et al. 1996; Herzschuh 2006), inferred from glaciological (Owen et al. 2002; Finkel et al. 2003), geochemical (Enzel et al. 1999; Jain and Tandon 2003), and palynological (Singh et al. 1990; Demske et al. 2009; Wünnemann et al. 2010) evidence, imply that the northwestern Himalayas were much cooler and drier during the last Pleistocene glacial maximum (Fig. 1, Wang et al. 2010). The reduction of northwestern forest habitat driven by these climatic changes (Trivedi et al. 2013), and the stability of southeastern forests through this period (although not immune to change, see Bhattacharyya et al. 2014), suggests that much of the Himalayan avifauna, at least that which was forest dependent, found refuge in the east Himalayas during periods of glaciation.

Present-day bird species richness is highest in the southeast and declines approximately twofold to the northwest (Fig. 1, Price et al. 2011). This pattern is consistent with dispersal limitation out of an eastern refuge. However, ability to establish in the west may instead be the main factor contributing to the decline in richness (Price et al. 2011; Ghosh-Harihar and Price 2014). Eastern species with ranges extending to the northwest are more often at higher elevations than those that are range restricted, and this has been attributed to the greater similarity of climatic conditions between the southeast and northwest above 3000 m rather than lower down (Price et al. 2011). At lower elevations, the northwest is much drier than the southeast, so the drop out of low elevation species towards the northwest could reflect deterioration in quantity and productivity of forest, i.e. persistence rather than dispersal as the limiting factor. Ghosh-Harihar and Price (2014) studied a particularly dispersive group (leaf warblers, Aves: Phylloscopidae) and found direct support for this proposition.

Despite this finding, a large fraction of the variance in range limits remains unexplained in previous analyses, and correlations with resources do not necessarily imply causality. First, in

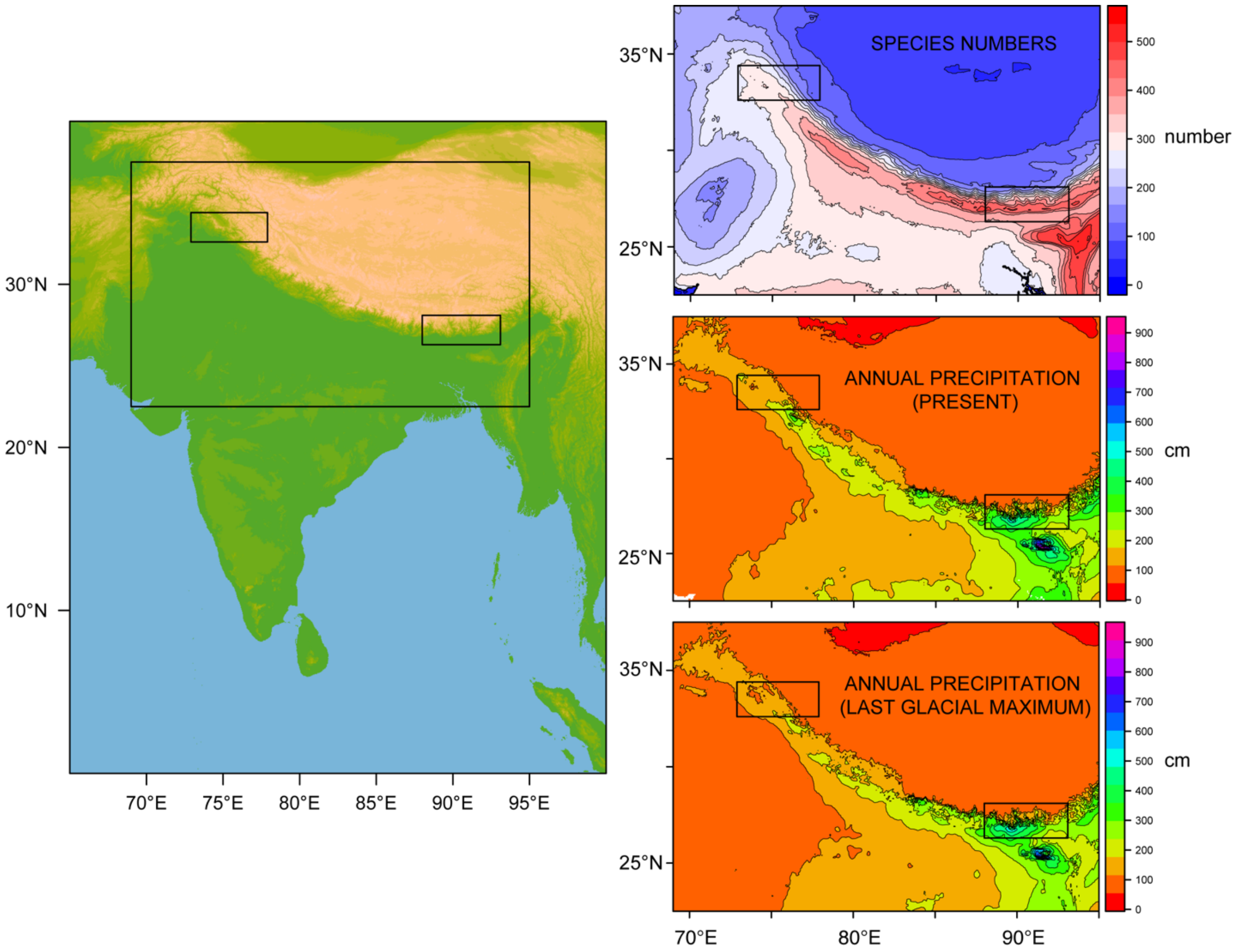


Figure 1 | Map of India showing the topography of the regional study area and two smaller areas designated as the southeast (longitudinal and latitudinal limits, 88°E, 93.1°E, 26.3°N, 28.1°N) and northwest (72.9°E, 77.9°E, 32.6°N, 34.4°N) sub-regions. Right top, breeding bird species richness. Right middle, present annual precipitation. Right bottom, annual precipitation reconstructed for the last glacial maximum. Precipitation data are from the WorldClim database (Hijmans et al. 2005; http://www.worldclim.org). Species numbers were generated using global range maps for all bird species provided by BirdLife International and NatureServe (2014).

addition to elevational differences, species in open country habitats more often have ranges extending from southeast to northwest than those in forest habitats (Price et al. 2011). This may reflect presence of suitable habitat across the range (i.e. a persistence factor) or instead alternatives linked to dispersal: species in more open habitats may be more dispersive or more open habitats may be more connected. Second, Srinivasan et al. (2014) argued those Himalayan babblers (Aves: Timaliidae) that undergo greater altitudinal movements are more dispersive. They showed that these species tend to have ranges extending further to the northwest but were unable to eliminate various confounds, including climate tracking and the possibility that altitudinal movements were a response to range expansions into the more seasonal northwest rather than a predisposing factor leading to the range expansion. Here, I expand on these studies to directly compare a measure of intrinsic dispersal ability (i.e. a species' wing shape) and extrinsic landscape barriers (topographic complexity within a species elevational distribution) for a large number of Himalayan birds. I compare climatic niche explanations with those of dispersal limitation.

Besides climate tracking, other factors may affect persistence in new locations, including competition (MacArthur 1972; Pigot and Tobias 2013; Price et al. 2014) and hybridization between close relatives (Barton and Hewitt 1985). Although competition may be diffuse (MacArthur 1972), competitive interactions and especially hybridization are expected between closely related species and they can be most easily examined in this context. I studied recently diverged allopatric replacements, in which the range limit of a species occupying the southeastern Himalayas is plausibly set by an interaction with a close relative occupying the northwest. In these replacements, dispersal is not likely to directly set the range limit. However, I show that dispersal is important in affecting the conditions under which allopatric speciation may occur (Claramunt et al. 2012), thereby indirectly affecting range limits of closely related allopatric/parapatric forms.

1.3 Methods

1.3.1 Species

I consider the geographic distributions, morphological traits, and phylogenetic relationships for 444 bird species that breed in a 10,000 km^2 region encompassing eastern Sikkim, northern West Bengal, western Arunachal Pradesh, and Bhutan (Fig. 1). I obtained measurements for 339 species in the order Passeriformes (93% of oscines and 4 of 5 suboscines in the region) and 105 nonpasserines from the Piciformes, Bucerotiformes, Upupiformes, Trogoniformes, Coraciiformes, Cuculiformes, Psittaciformes, Apodiformes, Strigiformes and Columbiformes (70% of the total species in these groups in the region). I did not include Galliformes, Anseriformes and Falconiformes. Geographic distributions (polygons) and elevational range limits for all species were taken from Price et al. (2011), and a detailed description of how they were generated is discussed therein. Some updates were made to the original data, and the entire dataset, with updates and novel data described below, is deposited in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.ks633 (White 2016). Following the historical directionality of habitat expansion after the last glacial maximum (Trivedi et al. 2013), I maintain a southeast to northwest focus to analyze factors contributing to range expansion, and ignore species present in northwest but not the southeast (N = 125, from Price et al. 2014). Because the inference is that forested habitat was differentially lost from the northwest during the Pleistocene (Ray and Adams 2001), I also consider forest (N = 340) and open (N = 94) species separately (from Price et al. 2011: species were defined as open if, when perched or foraging, they would most generally not have foliage above them; the remaining 10 species in the dataset are riverine), to test if signals of dispersal limitation are confounded with the contractions of certain niches in the northwest. Phylogenetic corrections throughout used the time-dated molecular phylogeny in Price et al. (2014), and tree manipulation was done using the ape package in R (Paradis et al. 2004). Simple regressions not taking phylogeny into account are included and gave similar results in most cases.

1.3.2 Morphological measurements of dispersal

Wing morphology is correlated with dispersal in birds (reviewed in Lockwood et al. 1998). Although many different measures of wing morphology have been used, recent studies seeking to quantify dispersal potential have used the hand-wing index (Kipp 1959; Burney and Brumfield 2009; Claramunt et al. 2012; Pigot and Tobias 2014). This index quantifies the aspect ratio of the wing on a size independent scale (0-1) from short, rounded wings to long, narrow wings and is measured as the difference in length between the longest primary feather and the first secondary feather divided by the longest primary, which is the wing length (see Fig. 2 in Claramunt et al. 2012). Wings with high aspect ratio confer increased flight performance and greater dispersal (Lockwood et al. 1998), and the hand-wing index is considered a comparable quantitative measure of dispersal ability between bird species (Burney and Brumfield 2009). The measure has been established as a correlate of natal dispersal distance (distance between point of hatching and point of breeding), migratory distance (Lockwood et al. 1998), and the ability to fly overwater in an experimental setting (Claramunt et al. 2012). I measured the hand-wing index on museum specimens at the Field Museum of Natural History in Chicago, Illinois. I measured 2 males of 421 species and the single specimen available for 23 species. In order to ask if hand-wing index is affected by ecological niche occupied, I assessed the effects of habitat association (open versus closed) and five classes of foraging stratum (ground, bush, mid-canopy, top canopy, aerial from Price et al. 2011, 2014) on hand-wing index using phylogenetic ANOVAs in the package phytools in R (Revell 2011, R Core Team 2015).

1.3.3 Topographic complexity

In order to assess the potential limits to range expansion imposed by topographic complexity, I compared two ~10,000 km^2 regions at the terminal margins of the Himalayan mountain range, one in the southeast near West Bengal, in which all 444 species are present, and one in the northwest near Kashmir (Fig. 1). Using geographic and elevational data from worldclim.org (Hijmans et al. 2005), I randomly paired geographic locations, one selected from each terminal region that shared the same elevation to the nearest 10m. This was done for the mid-point of the eastern elevational range for all 444 species in the dataset. Using the gdistance package (van Etten 2015) in R, I measured the least cost distance of travel between the two points in each pair across an altitudinal raster of the Himalayas (30 arc seconds resolution, from worldclim.org). The least cost distance is the smallest total elevational range that must be traversed between two points, based on raster cell midpoints. While species theoretically could follow an elevational contour (least cost distance = 0), adjacent points in a raster typically differ in their mean elevation and least cost distance is calculated from one raster cell to the next. To ensure that least cost distance was not strongly affected by the number of cells in the raster, I ran the analysis with a subset of the data (N = 200 species) on a lower resolution altitude raster (2.5 arc-minutes resolution, from worldclim.org) and got qualitatively similar results. Moreover, to ensure elevational midpoint adequately described landscape complexity encountered by each species, I also ran this analysis by calculating the mean of 100 least cost distances measured between regionally matched elevations randomly sampled within the limits of each species eastern elevational range. These values were highly correlated (*r* = .98) with the midpoint results, and the midpoint results are reported here.

1.3.4 Climatic niche similarity

A previous analysis compared the availability of climates encountered by each species in the southeast with the availability of that climatic regime in the northwest (Price et al. 2011). Here, I instead compare the climates at the same elevations in the east and west (results are similar to the previous analysis by Price et al. 2011, which was also recreated here for comparison, Table 3). I used the alternative approach as both a complement to the previous analyses and to allow direct comparison with the analysis of topographical complexity, which also used matching elevations in these regions. I used 4 bioclimatic variables from WorldClim for each of the 10,000 km^2 regions: maximum temperature, annual precipitation, temperature range, and precipitation seasonality; other climatic variables in WorldClim are not reliable for the northwest. For each species range between 200 – 5000 m, I drew 200 random co-ordinates in both the southeast and northwest regions and extracted the bioclimatic data at those points. To do this, for each species I drew 200 elevations from a normal distribution with mean equal to the species' mean elevation and standard deviation equal to one-sixth the species' elevational range, and then estimated climate for each of the points by choosing a cell on the climate raster closest to the elevation of that point. Following Broennimann et al. (2012), I calibrated a principal components analysis using the entire climate dataset (excluding the sampled points) in both 10,000 km^2 focal regions combined and calculated principal component scores for each of the 200 sampled points in each region. I then calculated the niche center for each set of 200 principal component scores using the technique described by Broennimann et al. (2012) and measured the Euclidean distance between the southeast and northwest niche centers for each species. These niche distances characterize the climatic differences between local conditions at similar elevations at the terminal ends of the mountain range and thus provide a lens with which to assess the potential for climatic niche tracking in the absence of other limits to range expansion.

1.3.5 Statistical analysis of range limits

I used the maximum northern latitude of each southeastern species' modern geographic range as a surrogate for range expansion, and I model maximum northern latitude as the dependent variable in all analyses. Because of the southeast to northwest orientation of the Himalayas, northern range limits are generally synonymous with western limits within the region of interest, but western range limits for lowland species traverse far beyond the Himalayas. Superspecies (Amadon 1966) containing young allopatric or parapatric replacements along the mountain chain (< 4 my diverged, mean = 2.54 my, N = 12) were treated as one species (i.e., hand-wing index was averaged for the two component species in the superspecies and the northern most limit was based on that of the western allospecies). Results were similar when I used the 12 eastern species instead.

Modeling maximum northern latitude, I first present results for three variables individually and then a full model where I account for correlations and potential interactions. I included 3 main effects and 1 interaction in the final multiple regression: hand-wing index, climatic shift between eastern and western squares, least cost distance of movement between the eastern and western squares (as calibrated for each species elevational mid-point in the east), and the interaction between least cost distance and hand-wing index. The interaction term was chosen to assess the hypothesis that the effect of intrinsic dispersal capability depends on the strength of extrinsic barriers. All variables were centered to mean zero prior to running analyses. The 2 other interaction terms (*hand-wing index: climate* and *climate: least cost distance*) do not appear to have readily interpretable biological meaning and, despite their inclusion improving the model fit (with a lower AICc), are not included in the final analysis (see below, Table 5). I performed the analyses first with all species included and then with open and forest species separately in order to account for potential open habitat refuges that may have remained farther to the northwest and correlations of habitat with intrinsic dispersal measures. Statistical analysis of geographic distributions was done using phylogenetic generalized least squares regression (PGLS) in the caper package in R (Orme et al. 2013). The relative importance of each main effect is reported by the change in the AICc score when each effect was left out of the full model (i.e., the model ranked third highest in Table 5), and the contribution of each effect to the total explained variance was calculated by averaging the sequential (type 1) sums of squares over all orderings of the regressors (Kruskal 1987).

1.3.6 Indirect effects of dispersal on sister species pairs

As noted above, 12 superspecies contain allopatric replacements that are <4 my diverged. I focus on these young species pairs with the logic that the northwestern range limit of the eastern species may well be set by hybridization or competition with the northwestern relative, rather than through the direct limitation of dispersal or climate. Allopatric speciation requires both range expansions and range fragmentation in sequence (Claramunt et al. 2012), processes that at the very least require successful movement across a landscape (i.e. expansion) and subsequent limitations of movement across the same region (i.e. fragmentation). In the Himalayas, the formation of these species complexes along the mountain chain should therefore be directly influenced by dispersal limitation. I ask if these 24 species are differentiated from other regional species by the hand-wing index using a phylogenetic logistic regression (Tung Ho and Ane 2014). Following Claramunt et al. (2012), I predict that intermediate dispersal ability favors the formation of the superspecies because high dispersal enables range expansion, but low dispersal enables range fragmentation. Intermediate dispersal combines these effects.

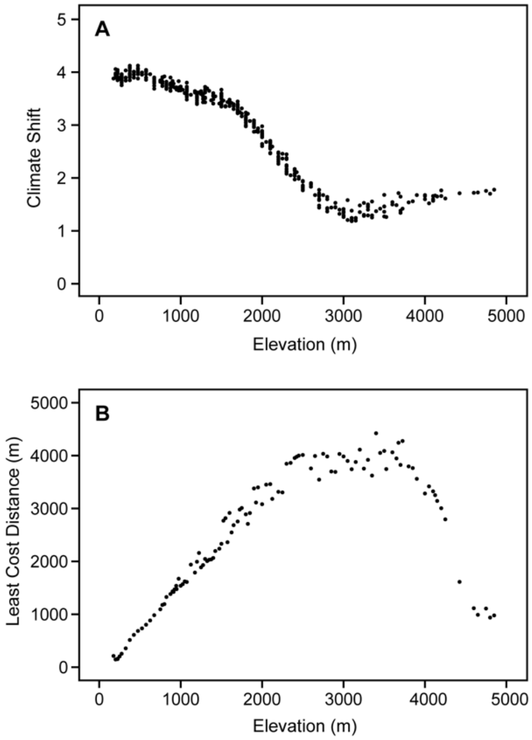


Figure 2 | Regional comparisons of climatic differences and landscape complexity between the southeast and northwest**.** Elevations are paired between the southeastern and northwestern terminal ends of the Himalayan range. A) Niche distance measured between southeastern and northwestern climatic niches based on species' elevational distributions in the southeast. B) Least cost distance measures the cost of movement (elevation traversed) between matched points at the same elevation in both regions, with the starting point located in the southeast. For each graph, the points represent the mean elevation each species occupies in the southeast.

1.4 Results

1.4.1 Limits to range expansion – climate tracking

Climatic niche differences are greatest at low elevations (Fig. 2A) where warm, wet, subtropical seasonality in the southeast gives way to cooler, drier, temperate climates in the northwest. Climatic niches become increasingly similar moving up in elevation and are most similar at 3000

– 3100 m. Between 3100 – 5000 m differences increase again, although relatively slightly compared to those differences measured at low elevations.

Climatic similarity between matching elevations in the southeast and northwest predicts northerly range extent (GLM: $P < 0.0001$, PGLS: $F_{1,442} = 50.35$; $P < 0.0001$). In other words, species with an eastern elevational distribution that requires a small climatic shift are more likely to range further to the northwest than those species with eastern elevational distributions that would require a large climatic shift. However, climate is correlated with elevation, and elevation is correlated with mountain barriers (Table 1), potentially confounding evidence for climate with dispersal limitation.

Table 1 | Correlations among main effects and maximum northern range limit

Effect	wing shape	least cost distance	climate shift	elevation	max. northern latitude
wing shape		0.11	-0.14	0.15	0.22*
least cost distance	-0.07		-0.81*	0.78*	0.22*
climate shift	0.05	-0.86*		-0.90*	-0.31*
elevation	-0.03	0.82*	-0.94*		0.33*
max. northern latitude	0.28*	0.25*	-0.36*	0.39*	

Note: Pearson correlations among the main effects and the dependent variable, maximum northern latitude (N = 444 species). Upper triangular elements are correlations that remain after accounting for phylogeny (correlations between independent contrasts). Lower triangle elements are raw correlations. Significant correlations are marked with *.

1.4.2 Intrinsic limits to range expansion – wing morphology

Birds in the Himalayas have highly variable wing-shapes, with a distribution roughly normal with outliers at the upper end (mean = 0.24, median = 0.21). The hand-wing index positively predicts northern range extent (regression of hand-wing index on northwestern range limit, GLM**:** $P < 0.0001$: PGLS: $F_{4,424} = 21.95$; $P < 0.0001$). Eastern species with long, narrow wings range further to the northwest than those with shorter, more rounded wings.

This result suggests that intrinsic dispersal ability affects range limits, but as in the case of climate tracking considered above there are potential confounds. Notably, dispersal ability may itself evolve in response to ecological factors, which could then be the primary determinant of range expansions. To assess ecological associations, I compared species morphology across primary habitat and foraging strata. Wing morphology differs both between open ($\bar{x} = 0.28$, SE = 0.017, N = 94) and forest ($\bar{x} = 0.22$, SE = 0.006, N = 340) habitats (phylogenetic t-test: $t_{432} = 3.98$, $P = 0.045$) and, separately, between five foraging strata (phylogenetic ANOVA: $F_{4,426} = 86.48$; $P = 0.001$, Fig. 3). Thus, wing morphology and habitat tracking may be confounded.

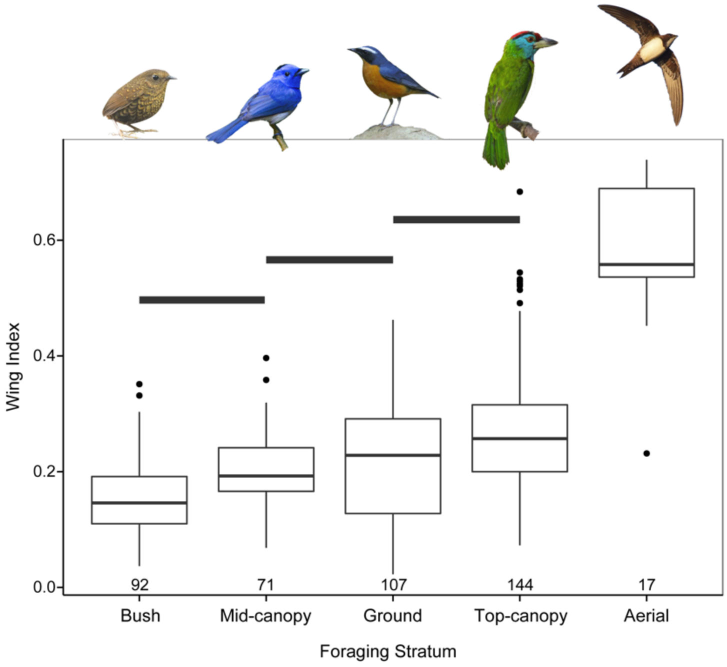


Figure 3 | Boxplot of hand-wing index within foraging strata**.** Lines represent medians, boxes span the inter-quartile range (IQR), and whiskers represent 1.5 × IQR, beyond which outliers are plotted. Phylogenetic ANOVA shows that wing shapes are differentiated among foraging strata ($F_{4,426} = 86.48$; $P = 0.001$), and solid bars indicate groups that are not differentiated by Tukey's *post hoc* test. Representative species from each stratum are (pictured from left to right) *Pnoepyga pusilla, Hypothymis azurea*, *Luscinia brunnea, Megalaima asiatica,* and *Tachymarptis melba*. Numbers indicate the sample size in each stratum.

1.4.3 Extrinsic limits to range expansion – topographic complexity

Connectivity between matching elevations at the terminal ends of the Himalayas is highly variable. Topographical complexity (i.e. least cost distance) is highest at about 3000 m (Fig. 2B). The low complexity at the highest and lowest elevations is associated with the relatively simple topography of the Tibetan plateau and the plains of India respectively. Despite this, least cost distance is actually positively correlated with northwestern range limits for southeast Himalayan birds (GLM: $P < 0.0001$: PGLS: $F_{1,442} = 23.31$; $P < 0.0001$). In other words, eastern species living at elevations where there are the greatest topographical barriers to dispersal are more likely to extend their ranges to the northwest. This may be explained because climatic similarity between the southeastern and northwestern regions is also greatest around 3000m, approximately the same elevation where the connectivity analyses suggest the physical landscape is most disjointed (Table 1). These contrasting effects of climate similarity and landscape connectivity imply that they should be separable in a generalized linear model, which is considered in the next section.

In summary, climate, wing shape, and topography are all associated with northern range limits. These variables are correlated (summarized in Table 1), so I now consider a full model including the possible interaction between hand-wing index and topography.

1.4.4 Interactions between intrinsic and extrinsic limits

The effects of hand-wing index and climate each remained important in the full model (Table 2). The main effect of least cost distance was no longer important, but the interaction between hand-wing index and least cost distance was a strong positive predictor of northwestern range limits (PGLS: $F_{4,439} = 23.17$; $P < 0.0001$, Table 2). The significant positive interaction shows that as

Table 2 | Phylogenetic generalized least-squares regression showing the effects of landscape and morphology on the northern range limit of southeastern species

	Estimate	Std. Error	P	Δ AICc	Prop. R^2
All species, N = 444					
wing shape	11.636	3.0706	0.0002 *	31.33	0.24
climate shift	-1.5840	0.3102	<0.0001 *	23.57	0.19
topography	-0.0003	0.0002	0.1236	17.78	0.35
wing shape : topography	0.0046	0.0010	<0.0001 *	18.17	0.22
Forest species, N = 340					
wing shape	11.632	3.2041	0.0003 *	24.40	0.25
climate shift	-1.3534	0.3383	<0.0001 *	13.81	0.15
topography	-0.0001	0.0002	0.5092	12.41	0.37
wing shape : topography	0.0048	0.0012	<0.0001 *	14.06	0.23
Open species, N = 94					
wing shape	9.0863	7.4378	0.2251	2.47	0.20
climate shift	-1.7080	0.7352	0.0224 *	3.30	0.15
topography	-0.0005	0.0005	0.3779	0.88	0.34
wing shape : topography	0.0047	0.0021	0.0324 *	2.63	0.21

Note: Prop. R^2 refers to the proportion of the multiple R^2 that each effect contributes, averaged across all orderings of the regressors in a PGLS. Δ AICc is calculated as the difference of the full model AICc and the AICc of a model calculated without the effect (the interaction term was also dropped when leaving out the *wing shape* and *topography* linear terms). For all species: Adjusted $R^2 = 0.1668$; $F_{4,439} = 23.17$; $P < 0.0001$; N = 444 species. For forest species: Adjusted $R^2 = 0.1619$; $F_{4,335} = 17.37$; $P < 0.0001$; N = 340 species. For open species: Adjusted $R^2 = 0.1480$; $F_{4,89} = 5.038$; $P = 0.001$; N = 94 species. Significant effects ($P < 0.05$) are marked with *.

least cost distance increases, hand-wing index becomes an increasingly important factor affecting range expansion in the Himalayas (Fig. 4). When least cost distance is low (which is at both high and low elevations, Fig. 2B), wing index has little influence on expansions, but when it is high, species that are particularly dispersive are more successful at expanding ranges. The full model explained 17% of the variance in northern range limits (as calculated by R^2), and 82% of that explained variance was accounted for by factors associated with dispersal (wing shape, topographic complexity, and their interaction). Climatic shifts between the southeast and northwest accounted for the other 19%. By order of relative importance (as calculated by Δ AICc), wing shape, climatic shift, and the *wing shape:topography* interaction were all important contributors to the model.

Price et al. (2011) compared the availability of climates encountered by each species in the southeast with the availability of that climatic regime in the northwest using a MaxEnt (Phillips et al. 2006) species distribution model (a detailed description of the method is available in Price et al. 2011). To ensure that the climate metric used here was not biasing the interpretation of the model, I re-ran the full model using the metric for climate availability in the

Table 3 | Phylogenetic generalized least-squares regression using niche availability instead of niche similarity

	Estimate	Std. Error	*P*	Δ AICc	Prop. R^2
All species, N = 444					
wing shape	13.0220	3.0976	0.0000 *	31.31	0.25
climate availability	0.3123	0.0838	0.0002 *	11.78	0.16
topography	0.0002	0.0002	0.3442	13.93	0.37
wing shape : topography	0.0043	0.0010	<0.0001 *	14.99	0.22
Forest species, N = 340					
wing shape	11.7590	3.2125	0.0003 *	21.37	0.26
climate availability	0.3335	0.0895	0.0002 *	11.75	0.17
topography	0.0002	0.0002	0.2656	9.42	0.38
wing shape : topography	0.0044	0.0012	0.0004 *	10.85	0.19
Open species, N = 94					
wing shape	11.9300	7.6406	0.1220	4.11	0.25
climate availability	0.1201	0.1997	0.5492	-1.85	0.09
topography	0.0003	0.0004	0.4707	2.03	0.31
wing shape : topography	0.0050	0.0022	0.0265 *	3.0	0.35

Note: Prop. R^2 refers to the proportion of the multiple R^2 that each effect contributes, averaged across all orderings of the regressors in a PGLS. Δ AICc is calculated as the difference of the full model AICc and the AICc of a model calculated without the effect. For all species: Adjusted R^2 = 0.1444; $F_{4,439}$ = 19.69 ; $P < 0.0001$; N = 444 species. For forest species: Adjusted R^2 = 0.1568; $F_{4,335}$ = 16.76; $P < 0.0001$; N = 340 species. For open species: Adjusted R^2 = 0.1000; $F_{4,89}$ = 3.583; $P = 0.009$; N = 94 species. Significant effects (P < 0.05) are marked with *.

northwest obtained using the same methods as Price et al. (2011). Results are qualitatively the same as those obtained using climatic shift (Table 3), except that climate availability is a poor predictor of range limits for open species. This may be because southeastern and northwestern climates are very different, and many species are not predicted to range into the west at all despite many actual ranges spanning the entire mountain range.

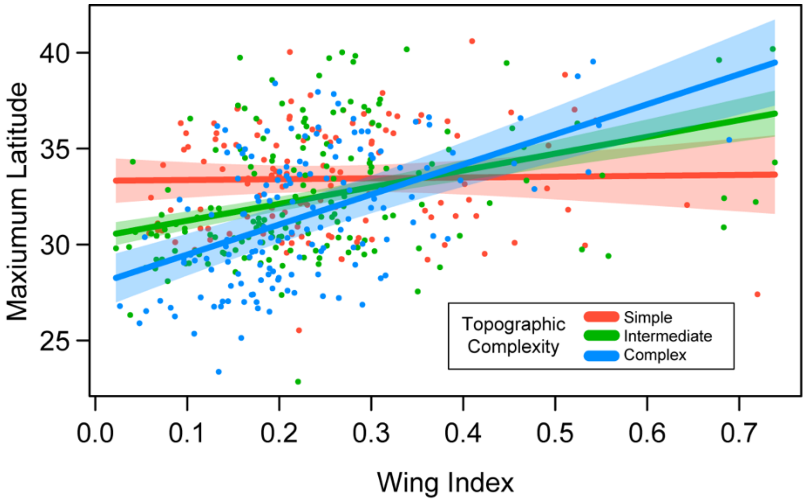


Figure 4 | Interaction between wing morphology and least cost distance affects northern range limits**.** Including all species, the effect of the interaction term *wing shape : topography* from Table 2 is detailed, showing the increasing limitation of wing index on the maximum latitude (degrees) of a species range as topographic complexity (least cost distance) increases. To aid interpretation, the range of least cost distance is cross-sectioned into three at the 30th and 70th quantiles. Each line represents regression model output within each section. Red: Species with elevational midpoints that correspond to least cost distance values between 225 – 1428 cost units. Green: Species with elevational midpoints that correspond to least cost distance values between 1429 – 3380 cost units. Blue: Species with elevational midpoints that correspond to least cost distance values between 3381 – 4487 cost units. Shaded regions correspond to 95% confidence intervals for each cross-section of the data. Lines are plotted using the package visreg (Breheny and Burchett 2016) in R.

1.4.5 Effects of dispersal within habitats

While these results suggest roles for both dispersal and climate, the hand-wing index itself correlates with habitat and hence part of the effect of hand-wing index on range limits may be through climate and habitat tracking, not dispersal. One way to assess the effects of habitat is to consider open and forest species separately. For forest species, the effects of wing shape, topography, climate, and the *wing shape:topography* interaction influenced northern range limits in much the same way as the model including all species together (Table 2). Open species, with greater mean hand-wing index overall and possible greater dispersal ability within the group, showed no direct limitation of wing shape, yet the *wing shape:topography* interaction, along with climate, continued to be an important predictor of range limits. In each case, the main effect of topography was not important, but the interaction between hand-wing index and least cost distance was a positive predictor of northwestern range limits.

Comparing full models for the foraging groups illustrated in Fig. 3 shows that the effect of the interaction between hand-wing index and least cost distance appears to be largely driven by top-canopy foragers. In this group as well, climate similarity was a strong positive predictor of northwestern range limits (Table 4), however the *wing shape:topography* interaction was a more important contributor to the model. Separate models for bush and ground foraging species instead show that hand-wing index alone, without the interaction, was a positive predictor of range limits, suggesting that even the weakest topographic barriers are strong enough to limit dispersal for species with extremely blunted wings (Fig. 3). Mid-canopy (N = 71) and aerial (N = 17) species show no effects of hand-wing index, topography, or climate on northern range limits (Table 4).

The variance explained by these models (R^2) and individual contributions of each effect to the explained variance are not consistent across foraging groups, as evidenced by the bush and top-canopy species. In these groups, the final model explained 25% of the variance in northern range limits for bush species and 22% of the variance for top-canopy species, but the

Table 4 | Phylogenetic generalized least-squares regression showing the effects of landscape and morphology on northern range limit of southeastern species among foraging strata

	Estimate	Std. Error	P	Δ AICc	Prop. R^2
Top-canopy species, N = 144					
wing shape	5.5963	4.9809	0.2631	16.86	0.22
climate shift	-1.7193	0.6311	0.0072 *	5.34	0.06
topography	-0.0003	0.0004	0.3824	16.66	0.25
wing shape : topography	0.0096	0.0020	<0.0001 *	18.57	0.47
Ground species, N = 107					
wing shape	20.5256	6.8403	0.0034 *	4.80	0.29
climate shift	-1.1645	0.5865	0.0497 *	1.86	0.18
topography	-0.0002	0.0004	0.5800	-3.94	0.53
wing shape : topography	-0.0005	0.0022	0.8000	-2.13	<0.00
Bush species, N = 92					
wing shape	28.6925	7.8115	0.0004 *	9.10	0.26
climate shift	-2.6060	0.6272	<0.0001 *	14.41	0.43
topography	-0.0015	0.0006	0.0217 *	5.81	0.31
wing shape : topography	-0.0001	0.0043	0.9800	-2.24	<0.00
Mid-canopy species, N = 71					
wing shape	0.5100	8.7853	0.9538	-3.79	0.23
climate shift	-1.3889	0.6272	0.0757	1.1	0.39
topography	-0.0003	0.0005	0.6470	-3.12	0.25
wing shape : topography	0.0039	0.0046	0.4071	-2.0	0.13
Aerial species, N = 17					
wing shape	-12.3492	26.0761	0.6442	-5.73	0.47
climate shift	-2.7198	2.8503	0.3587	-2.88	0.06
topography	-0.0033	0.0065	0.6211	-5.34	0.41
wing shape : topography	0.0108	0.0123	0.3975	-3.06	0.07

Note: Prop. R^2 refers to the proportion of the multiple R^2 that each effect contributes, averaged across all orderings of the regressors in a PGLS. Δ AICc is calculated as the difference of the full model AICc and the AICc of a model calculated without the effect (the interaction term was also dropped when leaving out the *wing shape* and *topography* linear terms). For top-canopy species: Adjusted $R^2 = 0.2259$; $F_{4,139} = 11.43$; $P < 0.0001$; N = 144 species. For ground species: Adjusted $R^2 = 0.1619$; $F_{4,102} = 5.57$; $P = 0.0004$; N = 107 species. For bush species: Adjusted $R^2 = 0.2520$; $F_{4,87} = 8.65$; $P < 0.0001$; N = 92 species. For mid-canopy species: Adjusted $R^2 = 0.0210$; $F_{4,66} = 1.38$; $P = 0.252$; N = 71 species. For aerial species: Adjusted $R^2 = 0.2890$; $F_{4,12} = 2.63$; $P = 0.0870$; N = 17 species. Significant effects (P < 0.05) are marked with *.

contributions of each effect to the explained variance within these groups were extremely varied. For top-canopy species, dispersal related effects (wing shape, topography and their interaction) accounted for 94% of the explained variance, showing little influence of climate, while for bush species, dispersal effects accounted for 57% of the explained variance, with climate contributing 43% to the explained variance.

I evaluated support for each potential set of model parameters in a phylogenetic least squares model using corrected Akaike information criterion (AICc, Table 5). The "full model" used in the analysis of northern range limits, with three main effects and the *wing: topography* interaction, has the third highest support. Two additional interaction terms (*wing shape:climate* and *climate:topography*) were also well supported by the model selection procedure, however they were not included in the general analysis of northern range limits due to the fact that it is not

Table 5 | Models to predict northern range limits of Himalayan birds

Parameters in model	df	AICc	Δ AICc	Model weight
wing shape + climate + topography + wing:topo + clim:topo	6	2383.9	0	0.614
wing shape + climate + topography + wing:clim + wing:topo + clim:topo	7	2385.9	2.03	0.222
wing shape + climate + topography + wing:topo	5	2388.1	4.23	0.074
wing shape + climate + topography + wing:clim + clim:topo	6	2389.2	5.29	0.044
wing shape + climate + topography + wing:clim + wing:topo	6	2390	6.13	0.029
wing shape + climate + wing:clim	4	2392.1	8.21	0.01
wing shape + climate + topography + wing:clim	5	2392.4	8.57	0.008
wing shape + climate + topography + clim:topo	5	2402.2	18.28	0
wing shape + climate	3	2405.9	22.01	0
wing shape + climate + topography	4	2406.3	22.41	0
wing shape + topography + wing:topo	4	2411.7	27.8	0
climate + topography + clim:topo	4	2413.6	29.74	0
climate	2	2419.1	35.26	0
climate + topography	3	2419.4	35.56	0
wing shape + topography	3	2428.1	44.23	0
topography	2	2444.2	60.34	0
wing shape	2	2445.5	61.64	0

clear what hypothesis these parameters test in the context of range limits. While these parameters indeed improve the predictive power of the model, their effects and the interpretation of those effects are perhaps biologically meaningless or at the very least their relevance is poorly understood.

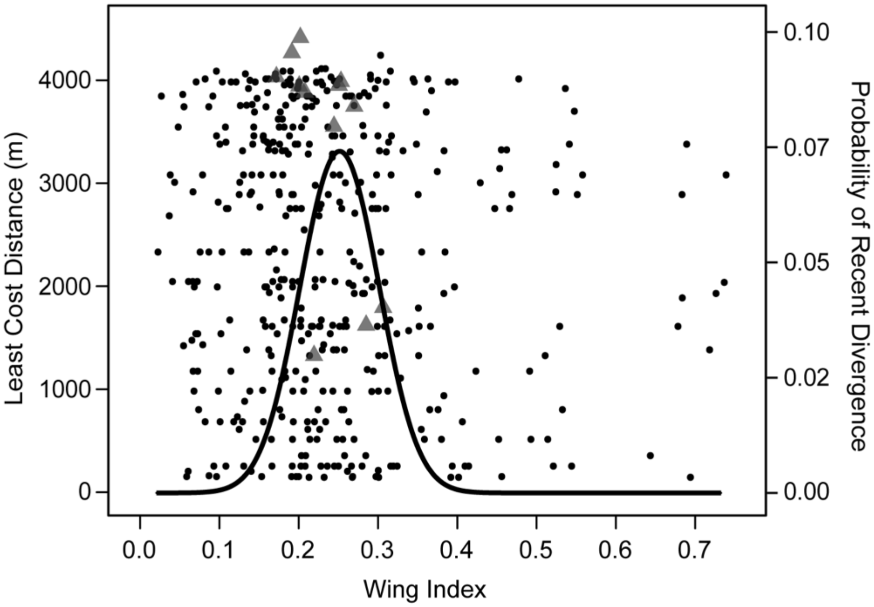


Figure 5 | Recent allopatric divergence is predicted by both intermediate dispersal ability and complex topography. The relationship between wing shape and least cost distance is detailed for all birds in the southeast (N = 444), showing the effects of both intrinsic and extrinsic factors on recent divergence. Grey triangles represent the 12 southeastern species that have recently diverged (<4 Ma) from a parapatric or allopatric relative in the northwest. Of these 12 species, 9 occupy elevational distributions characterized by the extreme of topographical complexity between the southeast and northwest Himalayas. Superspecies have intermediate dispersal ability when compared to the entire regional avifauna (phylogenetic logistic regression shows significant linear [hand-wing index: $\beta = 119.19 \pm 50.2$ S. E., $P = 0.018$] and quadratic [hand-wing index^2: $\beta = -228.81 \pm 100.01$, $P = 0.022$] influence of wing index). The curve shows the output from the phylogenic logistic regression. Black points represent all other species that breed in the southeast.

1.4.6 Superspecies are intermediate dispersers

Superspecies are defined here as a pair of closely related (< 4 my diverged) allospecies, with one each in the southeast and northwest regions. Eastern species that have intermediate levels of dispersal are more likely to belong to superspecies (Fig. 5, phylogenetic logistic regression, linear [hand-wing index, $P = 0.018$] and quadratic [hand-wing index^2, $P = 0.022$] terms are significant). The significance of the quadratic term implies intermediate dispersers are more likely to be in superspecies. These results are in accordance with the intermediate dispersal hypothesis, which states that high dispersers and low dispersers are less likely to form superspecies than intermediate dispersers, because of the opposing effects of gene flow and range expansions (Claramunt et al. 2012). For 9 of the 12 pairs, the southeastern species inhabits the slopes at elevations encompassing the peak of least cost distance (Fig. 5). This implies that while dispersal ability may be important, it must be interpreted in the context of landscape interactions.

1.5 Discussion

In a static framework assuming equilibrium, the limits of a species' range reflect those factors that influence the persistence of a species in a given environment (Sexton et al. 2009). These factors are likely to include climate (Parmesan and Yohe 2003) as well as biotic interactions including competition, predation, or facilitation (Wethey 2002; Briers 2003; Garrick et al. 2013). However, in a dynamic framework, factors associated with dispersal influence species ranges. This is particularly likely given the history of Pleistocene climatic perturbation across terrestrial environments. In the Himalayas, Price et al. (2011) framed their analyses in a static framework and found that climate tracking and occupancy of open habitat were important positive predictors of northern range limits. The analyses presented corroborate these findings (e.g., the influence of climate differences between southeast and northwest), but also show an important role for dispersal limitation assuming expansion out of an eastern refuge. Factors I have measured associated with dispersal explain four times the variance in northerly range limits than climate.

These results suggest that range size fluctuations associated with the climatic perturbations in the Pleistocene have been influenced by the ability of a species to disperse. As noted in the introduction, recent evidence from modern European plant distributions also suggests that range recolonization after the last Pleistocene glacial maximum was influenced by the ability of plant species to disperse north and the process may be ongoing (Svenning and Skov 2004; Normand et al. 2011). Normand et al. (2011) showed that while modern climatic conditions are a strong predictor of modern range limits for ~1000 plant species, 50% of those species are still affected by post-glacial recolonization lag. Further, for about one-third of these 50%, dispersal limitation, as measured by the spatial accessibility of new habitat, was a better predictor of range recolonization than modern climate. Southeastern Himalayan birds show a similar pattern. Dispersal has been shown to contribute to delayed range expansion for other animal groups as well, notably flightless beetles (Smith and Farrell 2005) and amphibians (Munguía et al. 2012), and there is evidence that bird distributions in Europe do not completely match their predicted climatic envelope, although it is not known how that mismatch is influenced by dispersal limitation (Devictor et al. 2008). In the Himalayas, both factors – dispersal and persistence – play a role, and dispersal increasingly plays a role when topographic complexity increases along the elevational gradient and when only forest species are examined.

The general link between geographic range size and intrinsic dispersal ability is evident when comparing species with the extreme measures of the hand-wing index: the blue-winged laughingthrush (*Trochalopteron squamatum,* hand-wing index = 0.02) and the alpine swift (*Tachymarptis melba*, hand-wing index = 0.74). *Trochalopteron squamatum* is endemic to the southeastern Himalayas and the nearby Hengduan mountains and only breeds in sub-tropical montane forests between 1000 – 2000 m. *Tachymarptis melba* on the other hand is broadly distributed between 150 – 2500 m across the Himalayas, as well as in Europe, the Middle East, and Africa. It is perhaps not surprising that *Tachymarptis melba* exhibits the greatest predicted dispersal ability within the entire Himalayan assemblage, as one individual in Europe was recently observed to have stayed in flight for over 200 days without landing (Liechti et al. 2013). On the other hand, experimental evidence from overwater flight tests found that species with similar hand-wing indices to *Trochalopteron squamatum* in Panama only flew on average about 200 m when released over water before crash-landing (Moore et al. 2008; Claramunt et al. 2012).

Seasonal migratory distance may contribute to differences in dispersal ability between species, and ultimately range expansions out of refuges. Migratory behavior may also contribute to population persistence by allowing species to avoid inhospitable conditions during the non-breeding season. Although not assessed in this study, migratory distance correlates with wing shape in other systems (Dawideit et al. 2009), and migrants tend to have longer natal dispersal distances than residents (Paradis et al. 1998). Srinivasan et al. (2014) suggested that the extent of altitudinal migration plays a role in setting the northern range limits of babblers.

It is possible that wing shape and dispersal themselves reflect adaptations to different ecological resources (Fig. 3) and the resources associated with low dispersal ability disappear to the northwest. For example, hand-wing index associates with foraging strata. More open foraging niches, where increased aerial agility is required to capture prey, correlate with longer, narrower wings, and closed foraging niches, where vegetative structure may obstruct movement and prevent efficient foraging, correlate with more blunted (i.e., less cumbersome) wings (Fig. 3). Swifts exploit aerial insects, present everywhere, while the thick undergrowth favored by babblers becomes scarcer to the northwest (Price et al. 2011). This may account for some of the correlation of wing shape with northerly extent, but several lines of evidence suggest declines in resources are not entirely responsible: (1) birds have extreme differences in dispersal ability as described above; (2) the effect of the interaction remains when looking within forest or open species only; and (3) the effects of wing shape become most important where topographic barriers are strong (elevations between 2000 – 4000 m). Finally, while it is true that foraging substrate appears to influence the evolution of wing shape (Fig. 3), niche-based predictions themselves do not account for different range expansions to the northwest. For example, frugivores in the top-canopy contain both green pigeons (Columbidae, *Treron* sp.) and barbets (Megalaimidae), yet only one group, the green pigeons, drops out to the northwest. Thus a species' eastern foraging niche, at least as far as it has been measured, is not entirely predictive of the species' northern range limit, although within one dispersive group it has been shown to be important (Ghosh-Harihar and Price 2014). In general, it should be noted that wing shape is by no means a perfect measure of dispersal ability. For any given species there are a number of factors that potentially drive individuals to move across the landscape, and there are likely other morphological contributions (i.e., body mass and wing loading) which may help resolve the true dispersal ability of a given species.

While dispersal and climate contribute to setting range limits in this system, a large fraction of the variance (>80%) in northern range limits remains to be explained. For example, green pigeons are much more dispersive, with longer, narrower wings than barbets, yet, as noted above, it is the green pigeons that drop out to the northwest. One possible effect that is not accounted for in the models is interactions between related species, notably competition or hybridization (Barton and Hewitt 1985; Price and Kirkpatrick 2009). While it is difficult to fully address this, an attempt was made to examine sister species, where the range of one species is likely to be set by the other, and here too there appears to be a role for dispersal. Claramunt et al. (2012) suggested that the production of allospecies is limited by the contrasting forces of range expansions (driven by high dispersal) and range fragmentation (driven by low dispersal), implying intermediate dispersal optimizes the production of allospecies. However, Claramunt et al. (2012) found little evidence for this, instead showing a role for low dispersal ability (as measured by the hand-wing index) in increasing the number of species among clades within Furnariidae (Aves). Part of the reason for their result may have been that the production of allospecies (favored by low dispersal) overwhelms any signal of range expansions (favored by high dispersal). Here, by considering superspecies and an explicit hypothesis of range expansion, results reveal a role for intermediate dispersal. Superspecies in the Himalayas were (1) characterized by an intermediate wing shape and (2) associated with a landscape of high topographical complexity. Especially for the 9 superspecies located in the highly complex topographical landscape at about 3,000m, dispersal limitation imposed by the landscape appears to be an important factor in initiating the formation of superspecies in the Himalayas.

Thus, intermediate dispersal may indirectly contribute to limiting ranges through the creation of interacting allospecies. The relatively small number (N = 12) of recently diverged allopatric species pairs conforms to the modern understanding that Pleistocene divergence between Himalayan lineages is uncommon (Päckert et al. 2012; Price et al. 2014). However, the data presented here do show that the high topographic complexity in mountains is an important driver of the production of allospecies (Davies et al. 2007; Fjeldså et al. 2012), and I argue this is primarily because of the limits imposed on dispersal. The roles of topography and dispersal in forming allospecies in the Himalayas appear to add an important nuance to the commonly discussed “sky island” model of diversification – while mountain peaks may be responsible for isolation, for species to fragment on sky islands, they need to arrive on those isolated habitats in the first place. Hence, species with strong dispersal limitation will not produce a large number of allospecies, or alternatively, as suggested by Claramunt et al. (2012), habitats that do not effectively isolate highly dispersive species will be similarly ineffective at generating allospecies. These results provide new insights about this general model of diversification but conclusions are limited with such small sample size. Genetic sampling has not been exhaustive or even necessarily sufficient to identify all cryptic species in the Himalayas, and biased genetic sampling could theoretically contribute to the result. However, sampling is likely to have been greater at low and high elevations, which tend to be visited more, where allopatric replacements appear less common.

Predictions about species range movements under future climate change require precise inputs as to the factors that control range expansions (Thomas et al. 2004; Loarie et al. 2008). In general, the importance of the climatic influence on range boundaries should not be taken as evidence that dispersal limitation plays no part, or that all species are at equilibrium with current climatic conditions. In their meta-analysis of species richness gradients, Field et al. (2009) identified 115 out of 297 studies that tested a hypothesis for a species richness gradient concerning dispersal or history. Of those, none appeared to have measured or compared intrinsic biological traits that govern locomotion, and instead these studies relied on tentative surrogates of dispersal limitation like geographic distance between source pools and colonized areas. Obtaining more accurate measures of dispersal limitation is critical in evaluating richness gradients more generally and specifically the hypothesis that geographic ranges are limited by dispersal.

2 Regional influences on community structure across the tropical-temperate divide[2]

2.1 Abstract

The varying composition of the fauna and flora at different locations across the globe delineates discrete biogeographic regions. Wallace identified six regions (Wallace 1876), such as the Nearctic and Neotropical, and based his scheme on the distinctive groups of species (i.e., biotas) found within each region. Recent updates have supported Wallace's classification while adding further partitions (Vilhena and Antonelli 2015), notably by incorporating phylogenetic information (Holt et al. 2013). Several partitions align with the transition between tropical to temperate climates, which is expected according to a number of historical models to explain differences in tropical and temperate species diversity (Hillebrand 2004; Wiens and Donoghue 2004; Mittelbach et al. 2007; Fine 2015). However, biogeographers have also long recognized that biotas overlap geographically (Pelseneer 1904) - the discrete assignation of geographic locations to single biotas has thus prevented a fundamental understanding of the mechanisms partitioning biotas across space, particularly across major biotic transitions such as the tropical-temperate divide. Here, we identify biotas and assign them to geographical regions using a method that allows for continuous turnover and, furthermore, quantifies affinities of species to biotas. We show that even across the steepest tropical-temperate gradient in the world, that of the Himalaya, the boundary separating temperate and tropical biotas is narrow, correlates with the freezing line, and is sharpened further when species abundances are taken into account. These patterns are retained when clades or traits rather than species are considered. The strong linkage

---

[2] A version of this chapter in review as: White, A.E., Dey, K.K., Mohan, D.M., Stephens, M., and T.D. Price. 2018. Regional influences on community structure across the tropical-temperate divide. In Review.

of local communities to regions defined by climate suggests important historical influences on community structure and function (Lessard et al. 2011; Cavender-Bares et al. 2016). The reality of the sharp tropical-temperate boundary adds credence to the argument that exceptional species richness in the tropics reflects species accumulation over time, with limited transgressions of species and clades into the temperate.

2.2 Introduction

The presence of more species in the tropics than the temperate regions has regularly been ascribed to the stability of tropical climates over millions of years (Wiens and Donoghue 2004; Mittelbach et al. 2007; Fine 2015). This idea implies that tropical and temperate biotas are discretely different, with limited overlap. However, evolutionary transitions in both plants (Zanne et al. 2014) and birds (Price et al. 2011) readily occur across the tropical-temperate boundary, and multiple correlated biotic and abiotic influences (Price et al. 2011), geographical barriers (Ficetola et al. 2017), and intrinsic dispersal ability (White 2016) have also been implicated in affecting individual species' ability to expand into the temperate from the tropics. Together, these findings suggest the tropical-temperate transition may be more gradual than implied from both previous biogeographical studies and historical explanations for the latitudinal gradient. Here, we evaluate biogeographic affinities for birds. We identify biotas and their spatial distributions to evaluate the reality of biogeographic regions and ask how contributions from different regions affect attributes of bird communities along the sharp tropical-temperate climatic gradient in the Himalaya.

Our methods are based on Grade of Membership models (Erosheva and Fienberg 2005), which extend earlier cluster-based models to allow that the units of analysis (e.g. species or geographic locations) have partial membership in multiple clusters. These probabilistic models are used extensively in population genetics (Pritchard et al. 2000) and Natural Language Processing (Blei et al. 2003), and have also been applied in ecology (Valle et al. 2014). Unlike previous approaches to partitioning global diversity (Vilhena and Antonelli 2015; Holt et al. 2013), in the Grade of Membership model the species present at each location can involve mixing of one or more biotas. The method simultaneously infers clusters of co-occurring species, which we term *species motifs,* and the proportional contribution of each motif to a location; this inference defines both biogeographical regions and their potential interdigitation. The contributions can be visualized on a map, identifying regions of admixture and patterns of turnover, thereby summarizing geographical structuring of global species. We use the general term *motif* because the same analytical framework can be applied to other features of biodiversity including species abundances in communities, analyses of clades, and community function, as we show below.

2.3 Results and Discussion

To infer species motifs, we start with an $N \times G$ data matrix of 0s and 1s denoting the absence and presence (BirdLife International and NatureServe 2014) of all terrestrial breeding bird species ($g = 1, 2, \ldots, G$, here $G = 9{,}518$) resolved to 1° × 1° across the globe ($n = 1, 2, \ldots, N$; here $N = 17{,}441$). We fix the number of motifs (K) *a priori* and estimate (1) the proportional contribution of the k^{th} motif to the n^{th} map cell (ω_{nk}) and (2) the probability of membership of the g^{th} species in the k^{th} motif (θ_{kg}, Fig. 9 illustrates the general method). The ω_{nk} values are plotted using a spatial STRUCTURE plot (Rosenberg et al. 2002) where each cell contains a pie chart with colours indicating the contributions from each species motif.

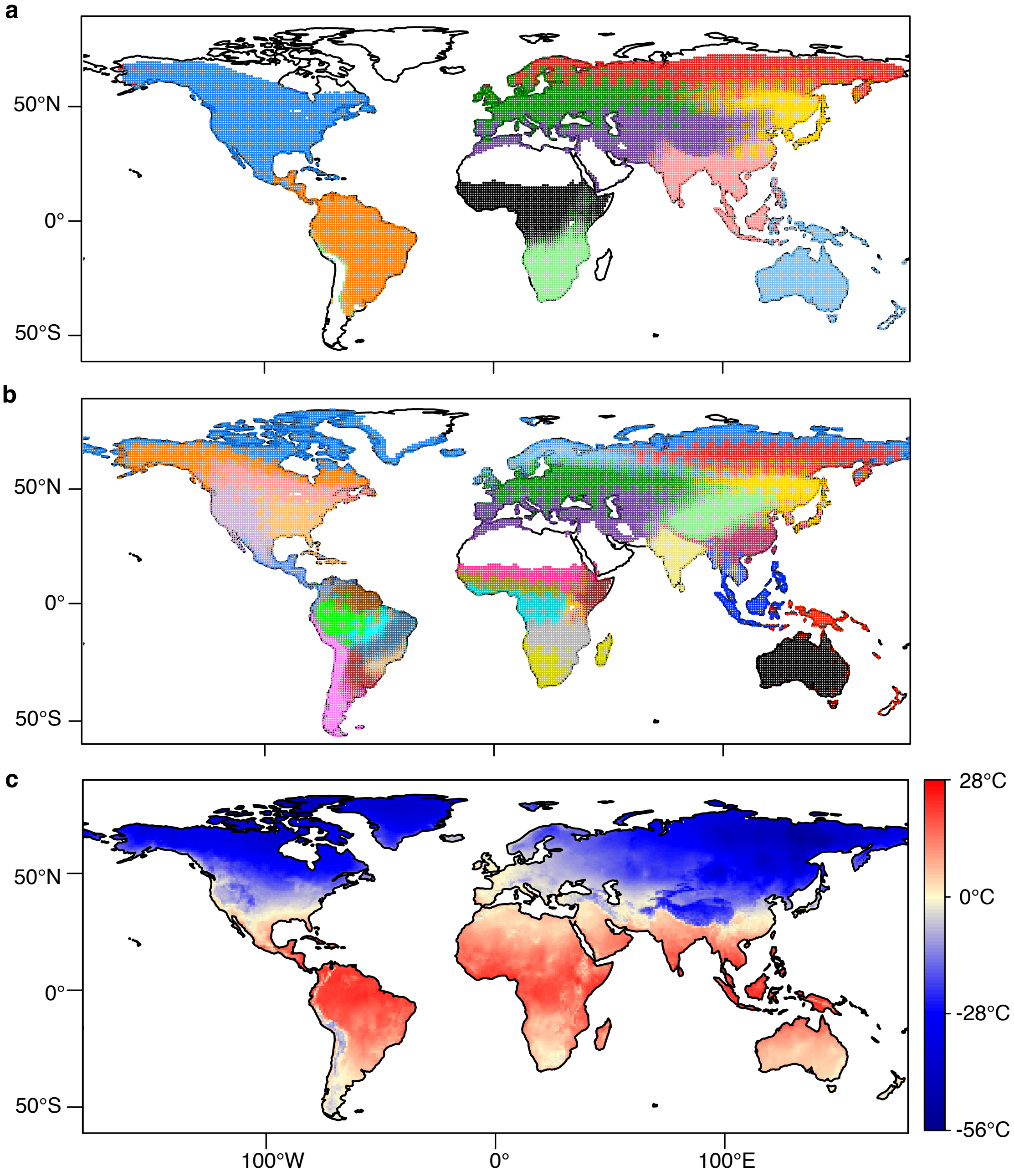


Figure 6 | Species motifs for breeding birds are organized by freezing temperatures. a, Pie charts represent the proportional contribution of species motifs ($K = 11$) to each 1° × 1° global map cell (i.e., ω_{nk}). Colours represent the different motifs. Blank regions (no pie chart) contain relatively few species and are dominated a motif that summarizes low richness. **b,** Same as panel **a** but with $K = 33$. **c**, Mean minimum temperature for the coldest month for each map cell (from worldclim.org (Hijmans et al. 2005). The freezing line is indicated by the pale-yellow coloured cells. Note the correspondence between sharp turnover of species motifs in **a** and **b** and the freezing line in panel **c**. A high resolution version of this figure, where the pie charts for each map cell can be clearly viewed, is available for download on figshare.com.

We fit the model for $K = 2 \ldots 35$ (Fig. 6). A feature of the model is that disparate locations with low species richness generally become united in one species motif. This unification arises because these locations have an especially low probability of contributing a species to global bird diversity. Other species motifs largely contribute to geographically cohesive locations (Fig. 6). Recognizing the long history of examining species presence-absence matrices *by species* rather than *by sites* (Simberloff and Connor 1979; Arita et al. 2008), we also computed the motifs on a matrix where the rows define species and the columns define sites. Fitting the model to this matrix clusters locations according their contributions to species ranges. Results are broadly similar, but, as theoretically predicted(Arita et al. 2008; Arita 2017), one motif now identifies locations based on high species endemism rather than richness (Fig. 10). A useful feature of this second approach is that we can map the contributions of each location to a given motif. We can then identify features of those map cells (e.g., climate) that are associated with, and are potentially responsible for, generating the primary spatial patterns of global bird diversity (Fig. 5).

The *by species* approach shows that, as expected, different continents are associated with different species motifs (Fig. 6). Other motifs are demarcated within continents and show varying degrees of interdigitation. Turnover between motifs is associated with climatic gradients (e.g. in precipitation across the southern edge of the Sahara, see black motif in Fig. 6a). As can be seen in Mexico (orange to blue in Fig. 6a), South America (southern border of orange), and south Asia (pink to purple), particularly sharp turnover corresponds approximately to the line of regular freezing (compare Figs. 6a and 6b with Fig. 6c). This freezing line runs right along the Himalaya at ~1,500 m. and putatively brings a temperate and a tropical motif into close contact. In the Himalaya, locations associated with one climatic regime appear to be separated from locations associated with the other regime by only a few km. and no obvious barriers. Ficetola et al.(Ficetola et al. 2017) inferred that the turnover of species across the Himalaya was due in part to barriers imposed by the mountains themselves (i.e. orography), but the coincidence of this boundary with the freezing line instead implies a role for temperature.

To examine the fine structure of this boundary, we conducted fields studies at all elevations across the entire Himalaya, the region which brings tropical and temperate climates in closer geographical proximity than anywhere else in the world. We censused bird communities on 38 five ha. sites (Fig. 7a, Table S1). Unlike the global analysis, we are now restricted to these 38 sites, and yet the questions we address require that we input a broader geographic context. Therefore, to incorporate geography, we define each community based on its dispersion field, which is the number of species from that community found in each map cell (Graves and Rahbek 2005; Arita et al. 2008; Carstensen et al. 2013, see Fig. 11 for a visualization). Input for this analysis is again an $N \times G$ matrix, but here each row represents the dispersion field ($N = 38$ rows), and the cells in the matrix are populated by species counts, rather than *0*s and *1*s. In this case, clustering depends only on the extent to which communities have similar dispersion fields and is agnostic with respect to species identity. For example, high elevation dispersion fields across the entire Himalaya look very similar (Fig. 11) even though they may share only 50% of their species.

We refer to the motifs generated by this analysis as *geographical motifs* (Fig. 9 illustrates the method explicitly for this case). Geographical motifs are defined as a weighted distribution of map cells (θ_{kg}). The contribution of a geographic region to a community is defined by the proportional contribution of the k^{th} geographical motif to the n^{th} community's dispersion field (ω_{nk}). Note that this analysis is robust to the precision of the range maps (see Geographical motifs in methods). We use Bayes factors to evaluate the strength of evidence for each value of K when compared to a model with $K = 1$ (i.e., no clustering, Fig. 12). Partitioning with $K = 2$ regional motifs has the highest support. Proportional contributions of each geographical motif to each community are shown in Fig. 7a, where communities are arranged by elevation along two gradients, one in the east and one in the west. Consistent with the global analysis, turnover is by elevation, not by geography, and occurs between 1200m – 1800m, i.e., straddling the freezing line (green-grey in Fig. 7a; Fig. 13e).

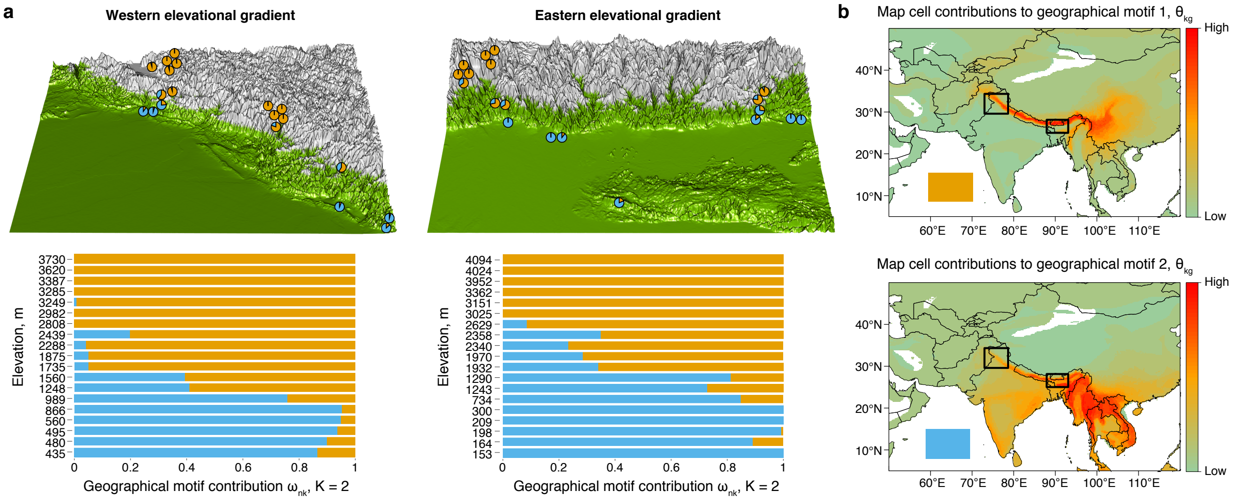


Figure 7 | Assignment of regional features to 38 local communities of Himalayan birds (K = 2 motifs). a, Communities arranged by elevation within geographical regions. Colours of each bar represent the proportional contribution of the two regional motifs to the community's dispersion field. These same proportions are plotted as pie charts at the locality of each community on 3D maps of the west and east region; the silver shading represents the "freezing line" where mean minimum January temperatures averages < 0 ˚C (Fig. 13). The spatial extent of dispersion fields was confined to the area mapped in panel **b** (G = 201,600). **b,** Heat map of proportional contributions of a cell in the raster to each motif (θ_{kg}). Rectangles indicate boundaries of 3D maps in panel **a**.

Maps in Fig. 7b show the proportional contribution of each cell to the two geographical motifs. A temperate motif extends out of the Himalaya to the mountains of western China and a tropical motif extends into peninsular India and southeast Asia. High elevation communities from the west have similar dispersion fields to those from the east (Fig. 11). Evidently, a temperate fauna associated with western China tracks climate across the Himalaya (Price et al. 2011; White 2016). At lower elevations, tropically adapted species have colonized from the plains. The plains represent climatic regimes that differ somewhat in precipitation (India: west, drier; South-east Asia: east, wetter, Fig. 13d) and the influence of these climatic differences becomes apparent when we set K = 4 (Fig. 8c). In this case the high elevation geographical motif running the length of the Himalaya is retained. The other motifs are low west, low east, and mid-elevation east. Regions associated with each of these four motifs are shown in Figure 8d, and climatic affinities of the low elevation motifs, which differ in precipitation, are shown in Figure 13b-d.

Attributes of communities depend not only on what species are present, which we have shown are strongly determined by geographic affinity, but also by their abundances, largely set by local conditions. Therefore, we constructed *species abundance motifs*, which differ from species motifs in that the entries are the number of individuals of each species ($G = 304$ species) in the community censuses ($N = 38$), rather than *0*s and *1*s. Here, θ_{kg} is the proportional contribution of the g^{th} species to the k^{th} motif, and ω_{nk} represents the proportional contribution of the k^{th} motif to the n^{th} community's species abundance distribution. Fig. 8a visualizes results for $K = 4$ (Fig. 12 for $K = 2$ to $K = 5$). An analysis of species motifs (i.e., presence/absence rather than abundance) results in a more gradual transition (see Fig. 15), suggesting abundance sharpens transitions between communities. The sharp turnover of species abundance motifs, when compared with both geographical motifs and those based on presence/absence (Fig. 15), implies species sourced from one region generally have low abundance when occurring in a community sourced primarily from another region. When $K = 4$, the four species abundance motifs (Fig. 8a) are largely congruent with the geographical motifs (Fig. 8c), a congruence that can be visualized in an interactive network of species co-occurrences (Fig. 14). This shows how the top contributing species to each species abundance motif lie along orthogonal axes defined by elevational and east-west partitioning (species that strongly contribute to each motif are listed in Table S2).

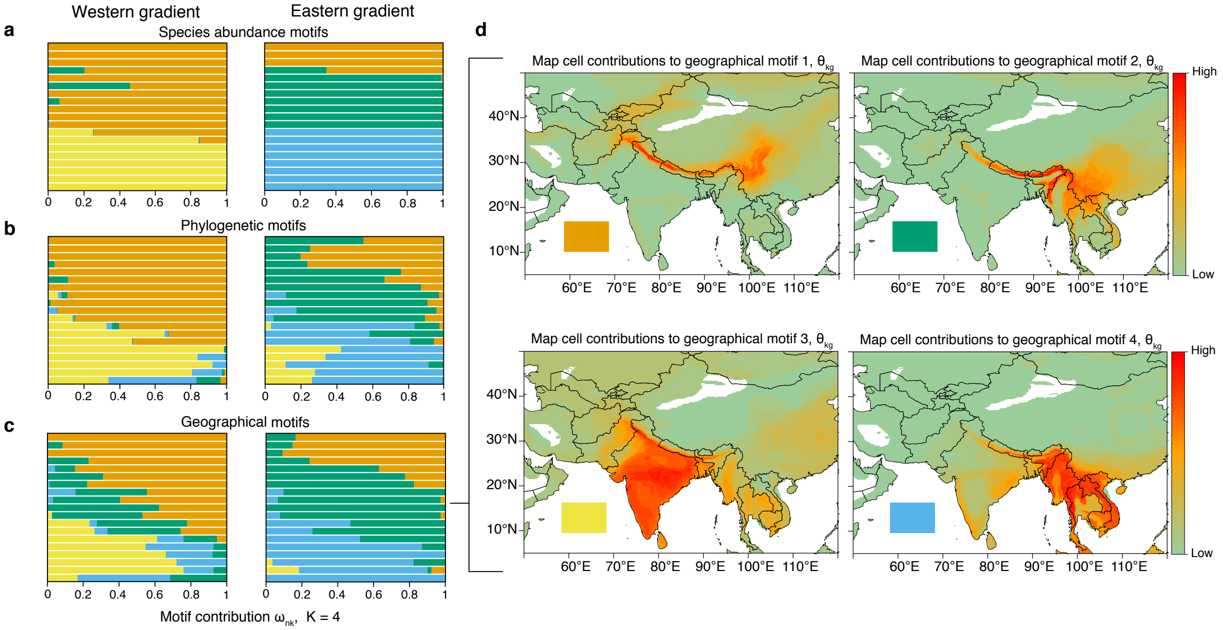


Figure 8 | Motif contributions to local communities when $K = 4$. The arrangement is as in Fig. 6. **a,** Taxonomic motif contributions to local communities. **b,** Phylogenetic motifs, for clades subtended by the 20 million-year timeline (n = 67 clades, Fig. 16). **c,** Regional motifs. Maps show raster cell contributions to the regional motifs (θ_{kg}).

Several theories to explain the difference between tropical and temperate regions assume many clades are restricted to one or another region, implying a deep historical imprint on communities (Wiens and Donoghue 2004; Fine 2015). We assessed the evolutionary relationships of species within communities by generating *phylogenetic motifs*, continuing to weight by abundance. For each community, we summed the abundances of all species belonging to a given clade and proceeded to fit the Grade of Membership model, as in the species abundance motifs. Clades were defined with respect to a given time slice in the phylogeny (Fig. 16). Even at the 20 Ma timeline, where the collapsed phylogeny has just $G = 67$ tips, the ensuing phylogenetic motifs show strong structure, which matches that of the geographical and species abundance motifs (Fig. 8; Fig. 12). Given the persistence of the overall pattern of structure between the phylogenetic motifs and the geographical and species abundance motifs, we conclude the effects of climate in shaping modern communities have a deep history.

The geographical distribution of phylogenetic motifs reflects restriction of whole clades to tropical or temperate associated portions of the elevational gradient (e.g. *Cyornis* flycatchers at low elevation are replaced by ecologically similar, higher elevation *Ficedula* flycatchers; barbets, *Megalaima,* are only found up to the freezing line). However, in other clades, closely related species do contribute to different motifs (because one occurs above and one below the freezing line (Price et al. 2011; White 2016; Fig. 17). Hence, depending on the taxa compared, climatic niche is conserved or labile (Price et al. 2011), as in plants (Zanne et al. 2014). We suggest the differences in evolutionary lability are driven by when and where ecological opportunity arose within the Himalaya, and the ease of exploitation of this opportunity by ancestral forms. Notably if unexploited resources are similar in one climate zone to those in another, climatic niches may readily evolve within certain clades (e.g., *Aethopyga* sunbirds, *Pericrocotus* minivets, Fig. 17). Nevertheless, the freezing line appears to exhibit a strong control on individual species distributions even within those clades.

Ultimately, understanding community structure requires an evaluation of traits. Thus, we developed methods to define *trait motifs.* Here we study a composite of 5 morphological traits (beak, tarsus and wing measurements). From these measurements, we constructed a distance matrix, converted it into a dendrogram, and proceeded as in the phylogenetic case. Subtending tips in the dendrogram requires arbitrary cut-offs; here we infer trait motifs by collapsing nodes to leave ~50% of the original taxonomic diversity (i.e., $G = 152$ tips out of the original 304 species). The distribution of trait motifs across the Himalaya is similar to the pattern of the species abundance motifs and geographical motifs (Fig. 12). The structural commonalities of the Himalayan gradient persist not only across different values of K, which show a nested pattern (Fig. 12), but also among different kinds of motifs, suggesting the primacy of the tropical-temperate partition across multiple axes of biodiversity.

Across many environmental gradients, including within the Himalaya, community taxonomic composition appears to vary more or less continuously, both across elevation and geography (e.g., Fig. 17). Based on observations such as these, Ricklefs (2008) argued that continuous turnover implies communities are artificial constructs, whose structure is heavily impacted by dispersal from elsewhere. Results presented here imply a more integrated local community than may be inferred from species lists. First, common species are shared over parts of the gradient, overlapping with common species that are present in other communities, thereby generating a network that can be partitioned into well-supported motifs (Fig. 14). Second, communities may be similar in traits, even as species make-up differs. For example, along the eastern elevational gradient a species of sunbird, and a species of minivet (Fig. 17) are found in every forest community we surveyed. The elevational distributions of species in these distinctive groups generally differ, generating many different combinations of species in different communities, but the communities themselves are similar in the functions represented by these genera. Finally, if dispersal strongly affects community structure, then communities associated with more than one geographical motif should show elevated species, phylogenetic, and trait variance with respect to communities associated with just one geographical motif. We examined this with respect to phylogenetic diversity and found no support (Fig. 18). Instead, phylogenetic diversity monotonically declines with elevation, reflecting the increasing dominance of a single order, the Passeriformes, which are thought to have become established ~35 Myr ago in association with a rapidly cooling climate (Price et al. 2014). Communities with tropical affinities are sourced by older lineages, and this overwhelms any potential effect of mixing on phylogenetic diversity.

In conclusion, by developing methods that allow for continuous variation, we confirm two major insights of Wallace (Wallace 1876) – that the world is organized into discrete biomes and that tropical biotas have a long history distinct from that of the temperate. Biological diversity is partitioned hierarchically across the globe, with shallower climatic transitions producing nested and more continuous turnover within these major biomes (Figs. 6a, 6b, 12, 13). The deep history and strong geographical restriction of biotas is consistent with explanations for tropical-temperate differences rooted in hypotheses of age, area, and climatic stability (Fine 2015). The discrete contribution of biomes and their boundaries persist even when considering collections of locally interacting species, and associated traits, across very steep climatic gradients, confirming the importance of historical influences on community structure (Cavender-Bares et al. 2016). These influences may work through conservatism in the direct effects of temperature on bird species ranges and abundances (Elsen et al. 2017), or through similar conservatism in the distributions of producers, such as trees (Fine 2015), which in turn drive consumer coexistence (Price et al. 2014).

2.4 Methods

2.4.1 The Grade of Membership model

The Grade of Membership model (Erosheva and Fienberg 2005) is a clustering model that allows each sample to have proportional memberships in more than one cluster, where a cluster is defined by the underlying features of interest. Such models are widely recognized in population genetics as ADMIXTURE models (Pritchard et al. 2000), where they are used to determine an individual's ancestry from multiple populations, with each population having a distinct SNP-level allele frequency distribution. In Natural Language Processing, these models are popularly known as *topic models* or *Latent Dirichlet Allocation models* and are used to decompose the semantic structure of text documents into one or more *topics*, where each topic is defined by a distinct frequency distribution of words in a vocabulary.

Here, we illustrate how the Grade of Membership model can be extended to decompose the structure of surveyed ecological communities into component building blocks of community structure. Valle et al. (Valle et al. 2014) previously applied such a model to species abundance data, where they referred to the clusters as "component communities". In this paper, we develop a flexible framework that can be used to cluster both taxonomic presence-absence data as well as community abundance data, thereby integrating analyses at the global (i.e., biogeographic) scale with assessments at the community level. At the global level, we assess the structure of species presences to identify patterns of bioregionalization across space. At the local level, we assess structure for various axes of biodiversity – namely species abundances, lineages (phylogenetic), maps (geographic) and traits – to examine how biogeographic patterns are reflected within ecological communities. We generically refer to the clusters generated by these analyses as *motifs*. A *motif* can be viewed as an analog of ancestral populations in ADMIXTURE models or a topic in topic models.

2.4.2 Species motifs

In the global context, we constructed a binary presence-absence matrix by overlapping the geographic distributions of breeding birds for each 1° × 1° global map cell meeting two conditions: 1) the cell contained at least 10% land cover and 2) the cell was overlapped by at least 3 species breeding distributions. The resulting matrix comprised 9,518 bird species in 17,441 map cells that passed the processing steps. Polygons of species breeding distributions were obtained from BirdLife International (BirdLife International and NatureServe 2014).

We applied a Bernoulli version of the Grade of Membership model on the presence-absence data matrix $M_{N \times G} = ((m_{ng}))$ where m_{ng} is 0/1 based on if the species g is absent/present in map cell n.

$$m_{ng} \sim Ber\,(p_{ng}) \tag{1}$$

where p_{ng} is the probability that bird species g is present in the map cell n. We assume a lower dimensional representation for p_{ng}.

$$p_{ng} = \sum_{k=1}^{K} \omega_{nk}\, \theta_{kg}$$

where

$$0 \leq \omega_{nk} \leq 1 \qquad \sum_{k=1}^{K} \omega_{nk} = 1 \qquad \forall n$$

$$0 \leq \theta_{kg} \leq 1 \qquad \forall k \ \forall g$$

Here, K represents the number of underlying motifs fitted in the model, ω_{nk} represents the proportional contribution of the k^{th} motif to map cell n and θ_{kg} is the probability that the g^{th} species is a member of the k^{th} motif. We assume non-informative Dirichlet priors on the proportions vector $\omega_{n.}$ and non-informative beta priors for each θ_{kg}.

We fit this model for different values of K ranging from 2 to 35, and models for different K were compared using Bayes factors. The membership proportions vector $\omega_{n.}$ for each map cell n was displayed using a pie chart, placed at the exact latitude and longitude of map cell n. At a 1° × 1° resolution, this visualization shows both the spatial distribution of the motifs and, crucially, the transition between motifs across space (see Fig. 6). The $\theta_{k.}$ vector of probabilities represents the probability that a bird species is a member of species motif k and can be used to extract the bird species that differentially contribute to the different species motifs.

2.4.3 Species abundance motifs

Here, the input data comprises community censuses of species abundances in 38 5 ha. forest sites across the Himalaya (field methods and data availability described below). We observed 304 species in total and recorded their abundances in a table of counts $C_{N\times G} = ((c_{ng}))$ where c_{ng} is the number of individuals of species g observed in site n. Here, we applied a multinomial version of the Grade of Membership model that allows each surveyed site to have contributions from multiple motifs, where each motif is defined by the proportional contribution from each species. Each row of the data matrix is modeled as follows.

$$(c_{n1}, c_{n2}, \cdots, c_{nG}) \sim Mult\,(c_{n+}, p_{n1}, p_{n2}, \cdots, p_{nG})$$

where $c_{n+} = \sum_{g=1}^{G} c_{ng}$ and $\sum_{g=1}^{G} p_{ng} = 1$ for all n. As in case of species motif model, we assume that p_{ng} has a low rank structure and can be decomposed as

$$p_{ng} = \sum_{k=1}^{K} \omega_{nk}\, \theta_{kg}$$

where

$$0 \leq \omega_{nk} \leq 1 \qquad \sum_{k=1}^{K} \omega_{nk} = 1 \qquad \forall n$$

$$0 \leq \theta_{kg} \leq 1 \qquad \sum_{g=1}^{G} \theta_{kg} = 1 \qquad \forall k$$

Here, K represents the number of motifs, ω_{nk} represents the membership proportion of the k^{th} motif in site n and θ_{kg} is the relative frequency of the g^{th} species in the k^{th} community. We assume non-informative Dirichlet priors on the membership proportions vector $\omega_{n.}$ for sample n and the proportional motif contribution vector $\theta_{k.}$ for the cluster k.

We fit the Grade of Membership model for different values of K from 2 to 10. For each value of k, the model was run 10,000 times with different starting points from which the fit with the highest Bayes factor was selected as the best model for that value of K. We used the *maptpx* package (Taddy 2012) in R to fit the model for each K.

We displayed the membership proportions vector $\omega_{n.}$ for each community n using a *Block Structure* plot, a slightly modified version of the “Structure plot” representation used in population genetics for visualizing admixed individuals (Rosenberg et al. 2002). In the *Block Structure* plot, the vectors of membership proportions are represented as horizontal stacked bar charts with samples from the eastern and western Himalaya displayed in parallel blocks and samples for each block ordered by site elevation (elevations are given in Fig. 7). This aids in identifying the motif transition points with respect to elevation and comparing the motif structure between regions. The $\theta_{k.}$ vector provides the normalized abundance profiles of the different species in motif *k* and can be used to determine bird species that contribute most heavily to the local species abundance motifs.

In this case, species are considered as independent entities, but in reality, they may be related phylogenetically, or in terms of morphological characteristics or geographic range distribution. To capture these characteristics, we consider data with the same surveyed sites along the rows but with the features along the columns placed in appropriate units for different axes of diversity.

2.4.4 Phylogenetic motifs

We obtained a phylogenetic tree of species in our surveyed sites (Price et al. 2014), and cut the tree at a certain time slice, T, at which point some lineages in the tree would subtend one or more related species (Fig. 16). We obtained a count per site for each tip present in the phylogeny at time slice T by summing over the counts of the species that tip subtends. We then fit the multinomial Grade of Membership model (see Species abundance motifs above) to the counts data matrix with 38 sites along the rows and the common ancestors to species at time T along the columns. We considered values of T varying from the present day (the tips of the phylogenetic tree) to 20 million years before present. Results presented (Fig. 8b) correspond to an analysis with T = 20 million years before present.

2.4.5 Geographical motifs

For the geographical case, we used rasters, one for each surveyed 5 ha site, to represent the geographic distributions of the species present in each community. We counted the number of species from the community that are present in each raster cell in the region represented in Fig.

11 (longitudinal and latitudinal limits, 50°E, 120°E, 5°N, 50°N). This raster is commonly referred to as an "assemblage dispersion field" (Graves and Rahbek 2005), but here we focus on the dispersion field for breeding birds recorded in a small 5 ha. area, rather than all birds recorded in the entire raster cell as used in other applications. Polygons of species breeding distributions were provided by BirdLife International (BirdLife International and NatureServe 2014). After overlaying the species geographic distributions on a 360 by 560 raster ($1/8^{th}$ square degree cells), we obtained a count per raster cell, corresponding to number of species breeding polygons overlapping that cell. We subsequently vectorized the raster and stacked these vectors together to form a 38 by 201,600 matrix of counts, to which we then applied the multinomial Grade of Membership model. Note that these distribution maps are coarse grained and it has been suggested that they cannot be used at resolutions less than 110 km^2 ($1° \times 1°$, Hurlbert and Jetz 2007; Pollock et al. 2017). However, this is not an issue in our case, where we only need to robustly capture the general features of the dispersion fields in a limited number of motifs and each species list is known from a site-specific survey.

2.4.6 Trait motifs

Five morphological measurements were obtained from museum specimens for each species: bill length, bill width, bill depth, tarsus length and wing length (Price et al. 2014). As a pre-processing step, we constructed a distance matrix using the Z-scores of the $\log_{10}$ of each trait and clustered the 304 species into 152 trait clusters (i.e., 50% of the original species richness) using complete-linkage hierarchical clustering with Euclidean distance. We then obtained the abundance per site for each of the trait clusters by summing over the counts of the species within each cluster, to which we again fit the multinomial Grade of Membership model as above.

2.4.7 Modeling considerations

For the Himalayan data, we compared Grade of Membership models for different choices of *K* ranging from 2 to 10. As noted above, to account for the random initialization biases, for each *K*, we ran 10,000 iterations for each of the trait, phylogenetic and taxonomic counts data, and 100 for the regional level counts data, and we chose the model fit with the highest Bayes factor for that value of *K*. Because of computational constraints, we ran one iteration for each value of *K* for the global analysis. The results are coherent across different values of *K* (e.g., patterns are often nested for higher values of *K*), implying the revealed structure is robust.

The Dirichlet prior assumption in the Grade of Membership model assumes that the features along the columns of the data matrix are only weakly dependent. This assumption is not true for the species or species abundance motifs studies, as it ignores the phylogenetic relatedness and morphological similarity of species. The phylogenetic and trait-based analyses capture this relatedness to some extent, but a more comprehensive model that accounts for all inherent relatedness would be more optimal. Also, in the geographic context, the model ignores the spatial proximity of the raster cells to one another. The models we propose thus have room for improvement, but the simplicity of the models make them computationally fast, and as is evident from Figs. 1-3, the approach is successful in decomposing biodiversity of the global map cells or the local Himalayan sites into biologically interpretable motifs. Because of the simplicity of the model, Bayes factors used to compare model fits should be interpreted with caution. General patterns that are consistent across different values of *K* are likely more reliable to interpret than patterns associated with one “best fit” model.

2.4.8 Model comparisons

A longstanding concern in community ecology is whether the data conform to an appropriate null expectation. To address this issue, we randomly permuted the Himalayan census data according to four different null model approaches – *frequency*, *richness*, *independent swap*, and *trial swap* (Gotelli 2000; Miklós and Podani 2004). These permutations were generated using the R package *picante* (Kembel et al. 2010). We then fitted the multinomial Grade of Membership model on the permuted dataset and compared, for the same values of K, the observed and the null model generated distributions using Bayes factors. For each approach, the null model Bayes factor distribution came from fitting Grade of Membership model for a fixed number of clusters K on 1000 null model matrices. We compared the Bayes factor from fitting Grade of Membership models for same number of clusters K on the observed census data with respect to the null model Bayes factor distribution. The p-values were significant with respect to 5% level of significance in all comparisons ($P < 0.01$).

2.4.9 Comparison to other clustering methods

The Grade of Membership model uses a low dimensional factorization to cluster the samples, which calls for a comparative analysis of our approach with other clustering algorithms and dimension reduction methods. In the context of RNA-sequencing data, Dey et al. (2017) showed that the multinomial Grade of Membership model is more efficient in distinguishing samples from known groupings better than hierarchical clustering. A possible explanation for this is that the multinomial model is able to capture sparse count data better than a non-model based hierarchical clustering approach, as exemplified by RNA-seq. Sparse count-based data are typical characteristics of species abundance data too, with our data having close to 90% sparsity;

hence the multinomial Grade of Membership model would seem like a natural choice over hierarchical clustering in this case.

For further evaluation, we compared the taxonomic Grade of Membership results to patterns obtained using two traditional ordination techniques - Principal Components Analysis (PCA) and non-metric Multi-dimensional Scaling (NMDS). We also compared these results against patterns obtained using t-distributed Stochastic Neighbor Embedding (t-SNE), a more recently developed machine learning method for low-dimensional data projection (van der Maaten and Hinton 2008, see Fig. S2). Since t-SNE is sensitive to the choice of the perplexity parameter (Wattenberg et al. 2016), we performed additional visualizations of the taxonomic data using t-SNE at various levels of perplexity. Results were consistently less informative. Unlike PCA, NMDS, or t-SNE, the Block Structure plot representation of the Grade of Membership model displays site membership patterns for multiple motifs in one plot and also highlights change in these motif patterns across elevation and geography much more effectively than other approaches.

2.4.10 Field methods

Details on the bird surveys are in Price et al. (Price et al. 2014), who present data and describe the methods for 18 east Himalayan 5 ha sites. For this study, in order to account for geographical variation between the east and west Himalaya, TDP and/or DM surveyed one additional site in the northeast Indian hills of Meghalaya, and 19 sites in the west using similar methods (Fig. 7a). Some sites were surveyed in multiple years. For each site, we use results from a single observer from a single year, but results are not affected if we use other years or switch observers for sites which both TDP and DM surveyed (which were the majority). We measured abundance of species belonging to 11 orders (excluding game birds and birds of prey, Price et al. 2014). In total we encountered 304 species out of 621 species in these orders that breed in the Himalaya (Price et al. 2014).

2.4.11 Code availability

Code for fitting Grade of Membership models, analyzing motifs, and generating visualizations is publicly available in the new R package *ecostructure*. The package contains: 1) functions for processing local community data using a phylogeny or a morphological dataset to fit phylogenetic and trait motifs, 2) functions for generating presence-absence matrices and dispersion fields from GIS data sources to fit species and geographic motifs, and 3) functions for visualizing the output of those model fits. Tools for null model comparisons and comparisons of geographic motifs with climatic data are also provided. The package, and a detailed vignette of its functionality, is available for download at https://kkdey.github.io/ecostructure.

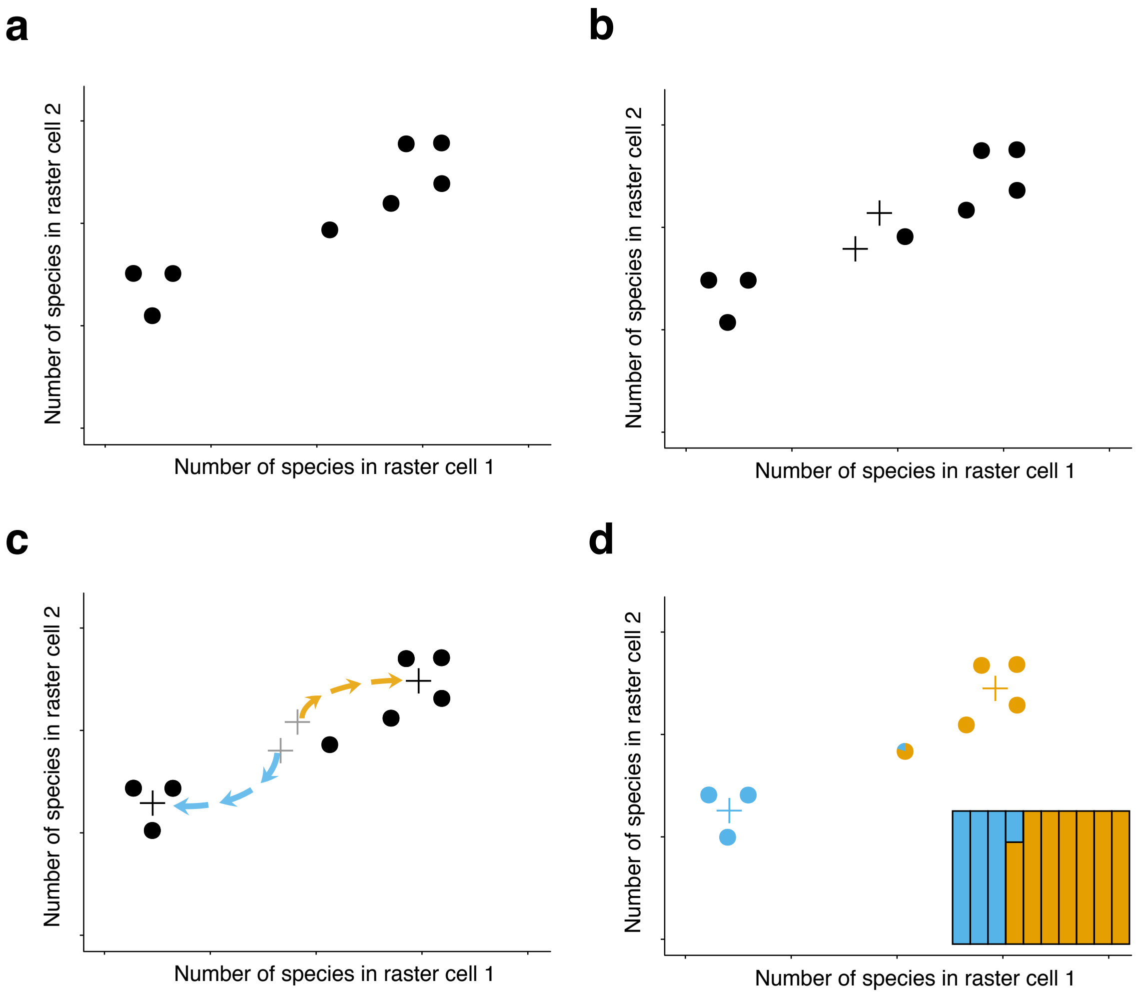


Figure 9 | Graphical illustration of the method used to assign species in a localized area or community to geographical motifs. a, The points represent the number of species from each of 8 hypothetical dispersion fields (associated with 8 localities) that are drawn from two cells in a hypothetical raster (note, in our analysis there are 201,600 cells). **b,** The procedure starts with a specific number of motifs, $K = 2$ in the illustrated case. The points are randomly assigned to a motif with equal probability, and their values averaged across the localities assigned to that motif. By chance, the motif means, illustrated by crosses, tend to each lie closer to one of the natural clusters in the data than the other. **c,** In an iterative procedure, localities are again assigned to motifs, but with a probability that is positively correlated with proximity to each. The new average values move towards natural clusters. **d,** The procedure reaches an equilibrium, at which point relative probabilities of a community being assigned to a given motif are recorded (vertical bars, ω_{nk}). Contributions of map cells to motifs (θ_{kg}) can be similarly assigned, e.g. map cell 2 more strongly contributes to the blue cluster than map cell 1 (e.g. see Fig. 7b, Fig. 8). This basic process can be generalized to evaluate other types motifs explained in this paper.

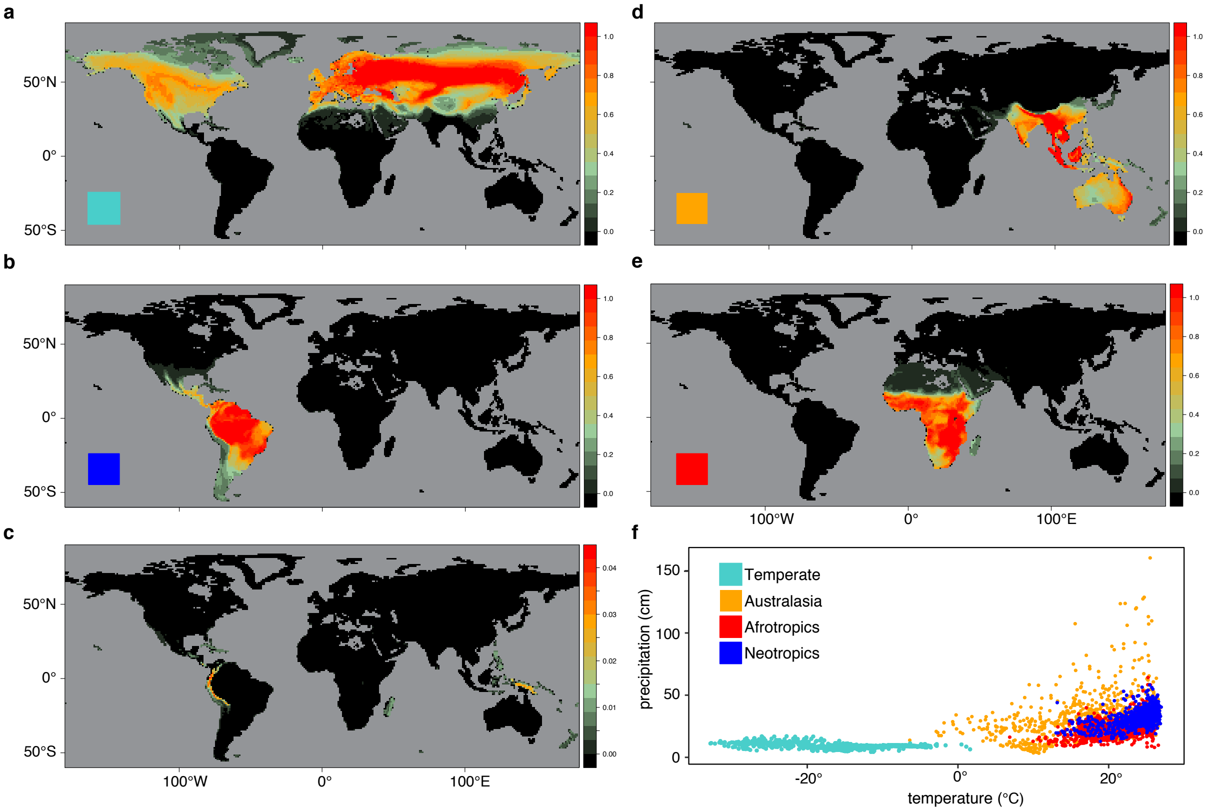


Figure 10 | Map cell contributions to motifs defined by species ranges. Grade of Membership model fit (K=5) on the same data matrix as Fig. 6, but with rows and columns transposed. This model identifies motifs that are formed according to shared species ranges. **a-e,** Maps show the probability of membership for each map cell to the given motif (the θ_{kg} from these model fits). Map in the lower left (**c**) is for the motif that captures regions of endemism. **f,** Mean temperature of the coldest 3 months of the year (°C) against precipitation of the wettest month (cm, from worldclim.org, Hijmans et al. 2005) for the top 5% of locations contributing to the 4 major motifs.

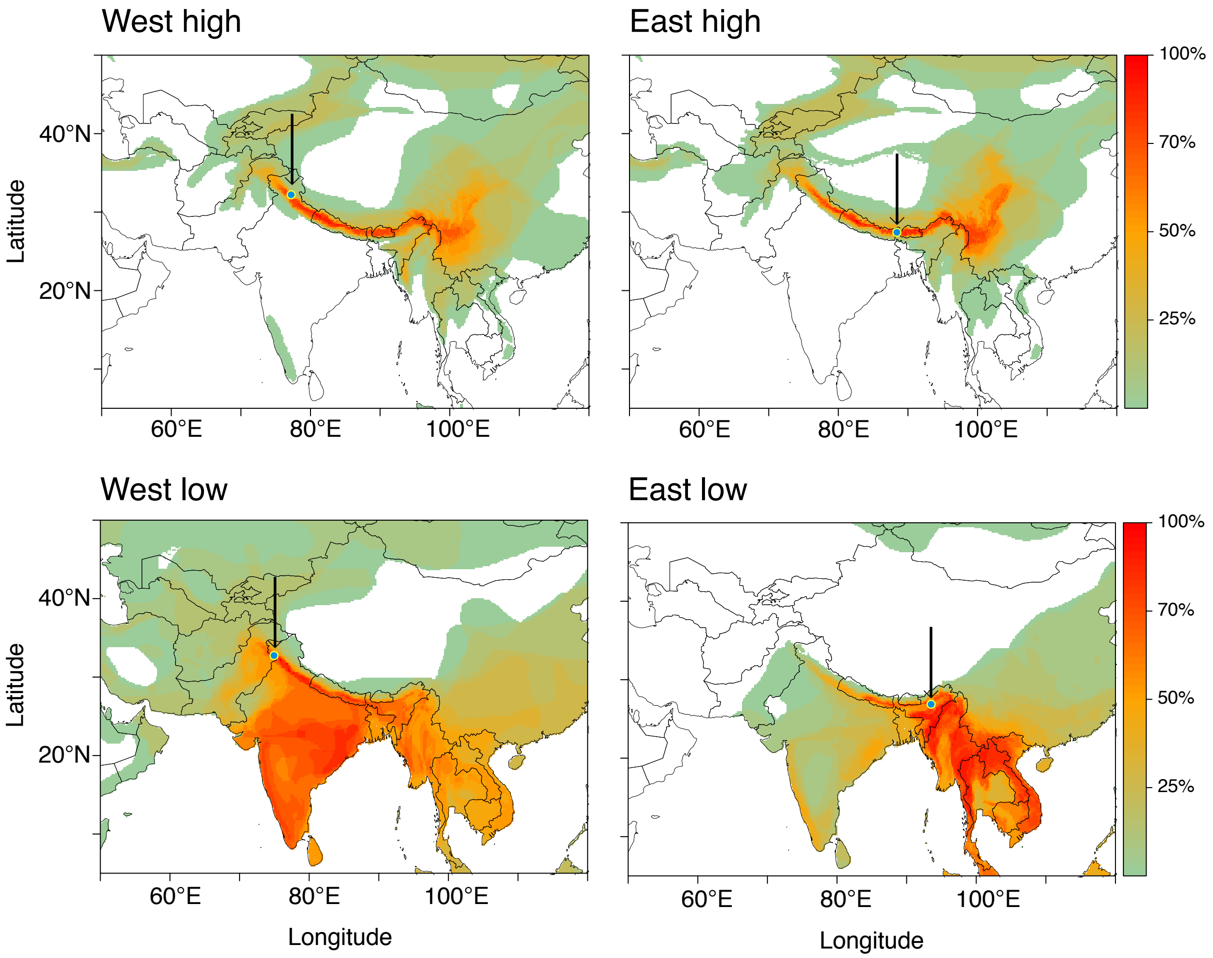


Figure 11 | Dispersion fields for four communities. West high (Site: M2, N = 26 species, elevation = 3285 m; see Table S1), East high (N2, 20 species, 3952 m), West low (J1, 27 species, 480 m), East low (A8, 28 species, 300 m). Colours indicate the proportion of species from the local site that are present in each raster cell. Locations of each site are indicated by an arrow.

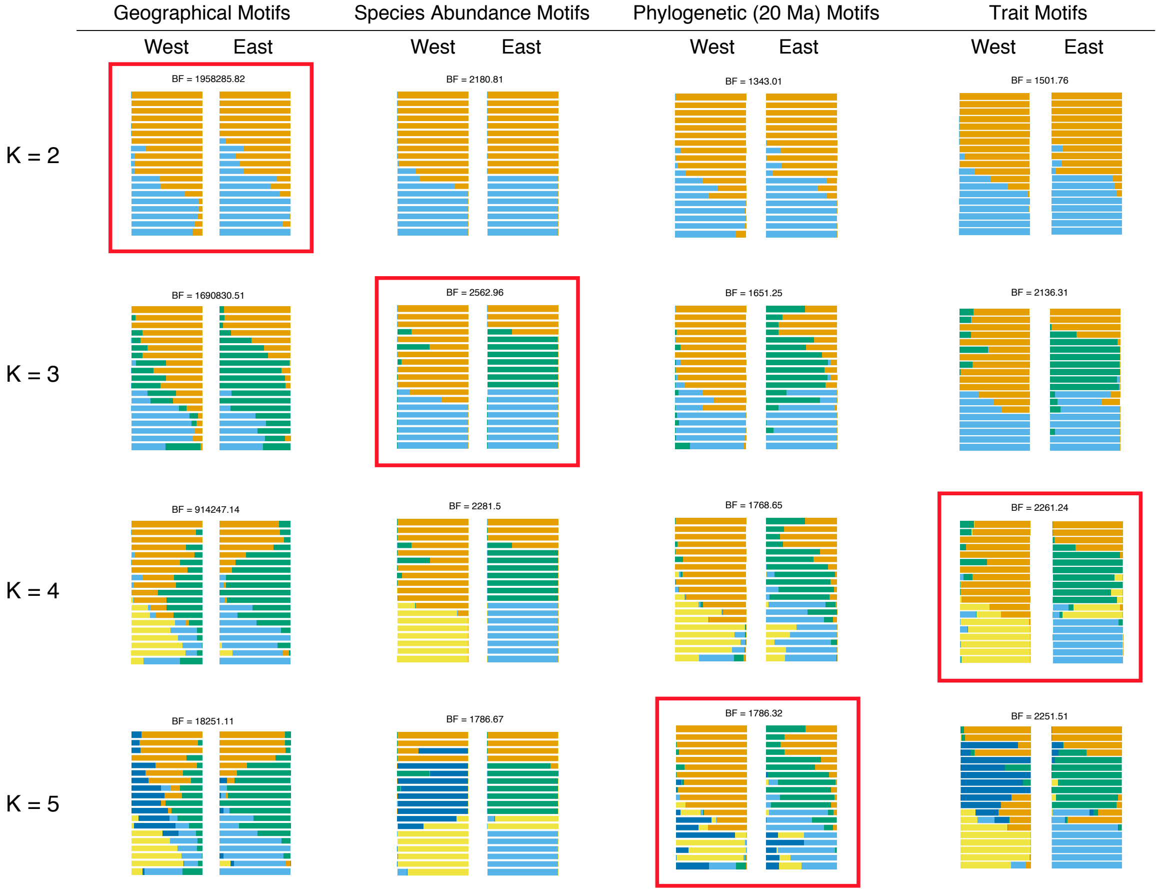


Figure 12 | Motifs for *K* = 2 - 5. Motif contributions to Himalayan bird communities for geographical, species abundance, phylogenetic (clades subtended at 20 million years) and trait (50% of species diversity) motifs, with associated Bayes factors above each graph. For each column, the *K* value with the highest Bayes factor is highlighted with a red box. Communities are separated into west and east and ordered by elevation within a region.

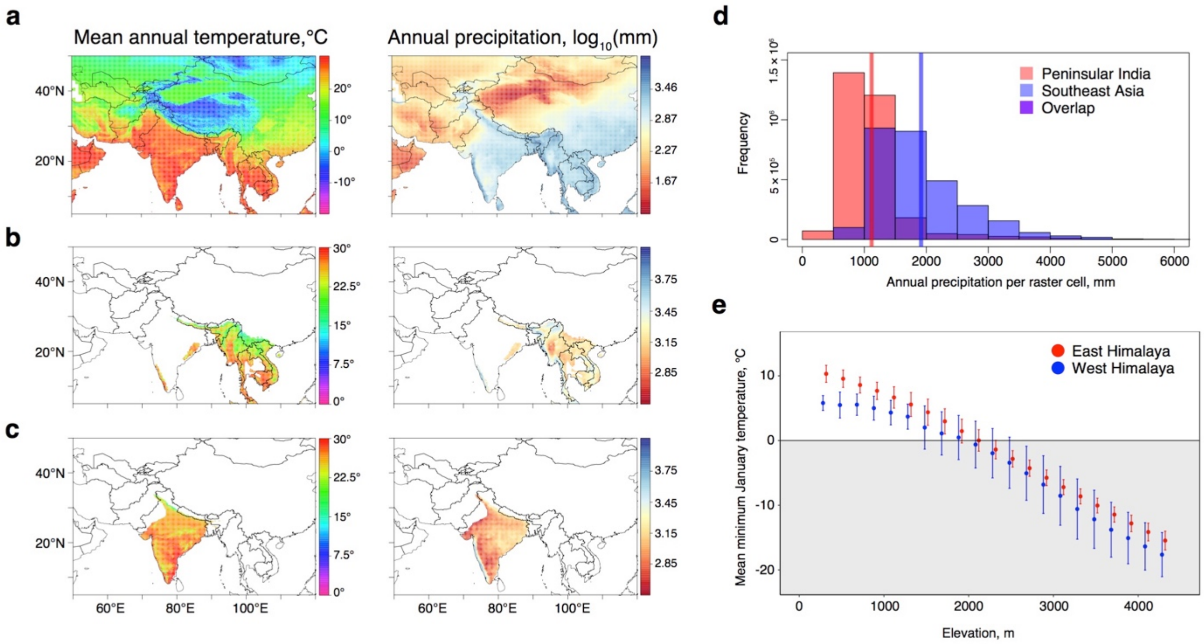
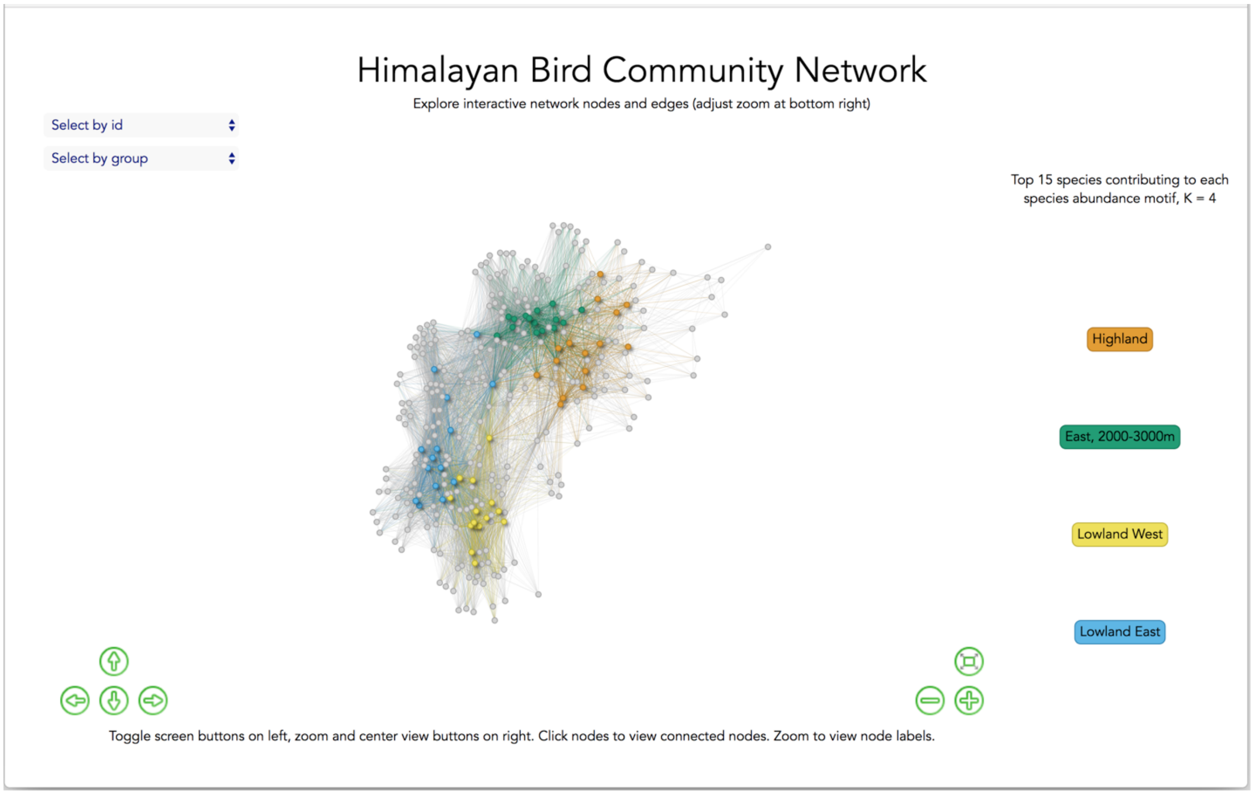


Figure 13 | Climate across the Himalaya. a, Mean annual temperature and total precipitation (from worldclim.org, Hijmans et al. 2005). **b, c,** The same climate variables for the top 10% of raster cells contributing to the low-west and low-east regional motifs respectively, when $K = 4$ (see Fig. 8). **d,** Histogram of the (**b**) and (**c**) cell entries for precipitation showing the east regional motif receives more rainfall. Lines represent mean precipitation in each region. **e,** Mean minimum January temperature against elevation for the west and east Himalayan regions (boundaries for climate were defined by the rectangles in Fig. 6b). Grey shading represents regular freezing.



Himalayan Bird Community Network

Explore interactive network nodes and edges (adjust zoom at bottom right)

Top 15 species contributing to each species abundance motif, K = 4

Toggle screen buttons on left, zoom and center view buttons on right. Click nodes to view connected nodes. Zoom to view node labels.

Figure 14 | Screen-shot of interactive network of species co-occurrences. A fully interactive version to be opened in an internet browser highlights the species names and their associations (Fig. S1). Coloured points represent the top 15 species contributing to each of the species abundance motifs (i.e. largest θ_{kg}), when $K = 4$ (Fig. 8a). In the few cases when a species was shared between motifs in the top 15, the species is illustrated with the motif for which it contributes with a higher θ_{kg} value. The next species down the list was then included for the other motif. Edges represent a co-occurrence between species in a community. If two species tend to co-occur or co-occur with similar species they are drawn closer together in the network. Note that species abundance motifs (i.e., colours in the network) are generated using abundances, yet the network structure represents species presences.

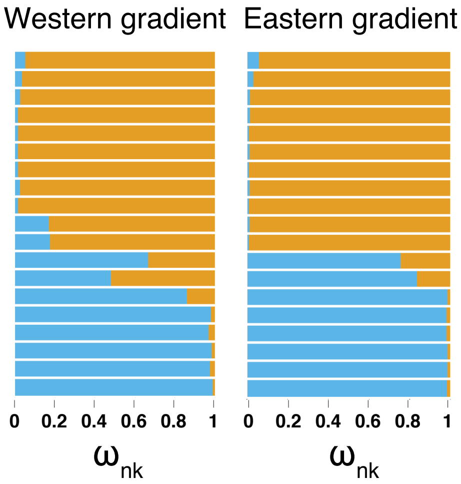


Figure 15 | Species motifs generated using a presence-absence model for *K* = 2. These motifs are constructed in a similar way to the species abundance motifs (Fig. 12), but species abundances are set to 0 or 1 (i.e., a binomial distribution underlies the Grade of Membership model). Clustering by elevation is retained for *K* = 2, but the sharp transition between tropical and temperate taxonomic motifs found using the multinomial model is more gradual, indicating that abundance distributions sharpen the partitioning associated with species presences in a community.

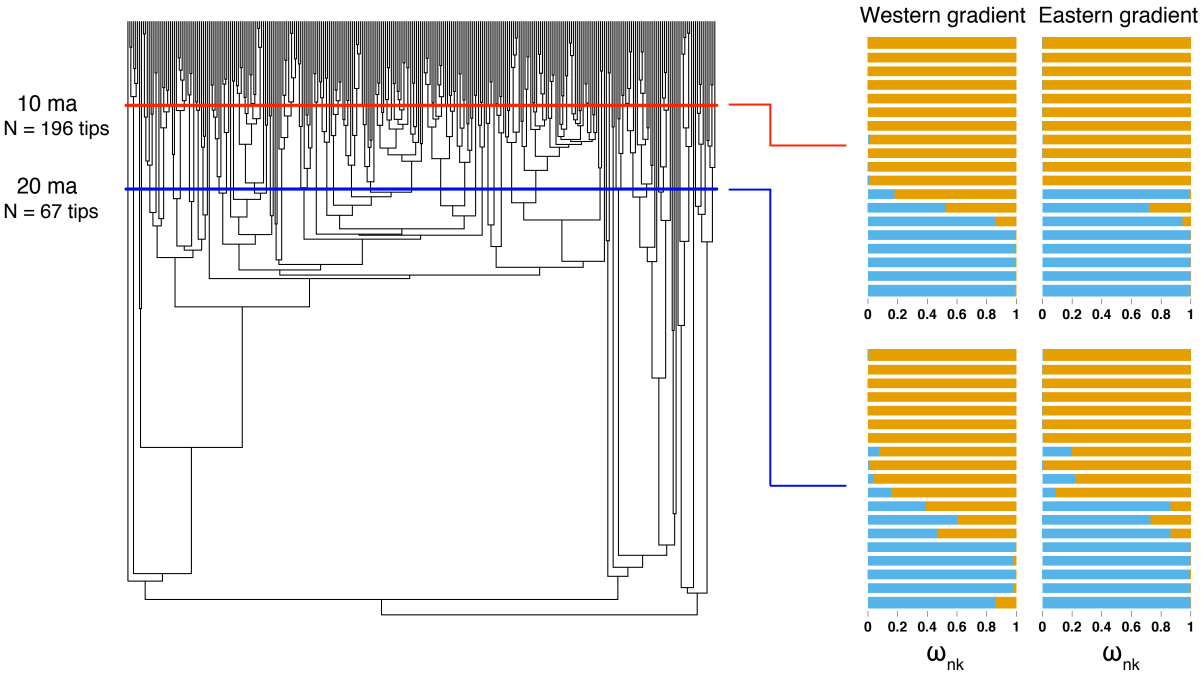


Figure 16 | Phylogenetic motifs at the 10 Ma and 20 Ma timelines. These motifs are constructed in a similar way to the species abundance motifs, but with all species in a clade subtended at a timeline collapsed into one lineage, whose abundance in the community is given by the summed abundances of all species in the clade. Strong clustering by elevation is retained for $K = 2$, implying that at least in some clades, the majority of species contribute mostly to one motif.

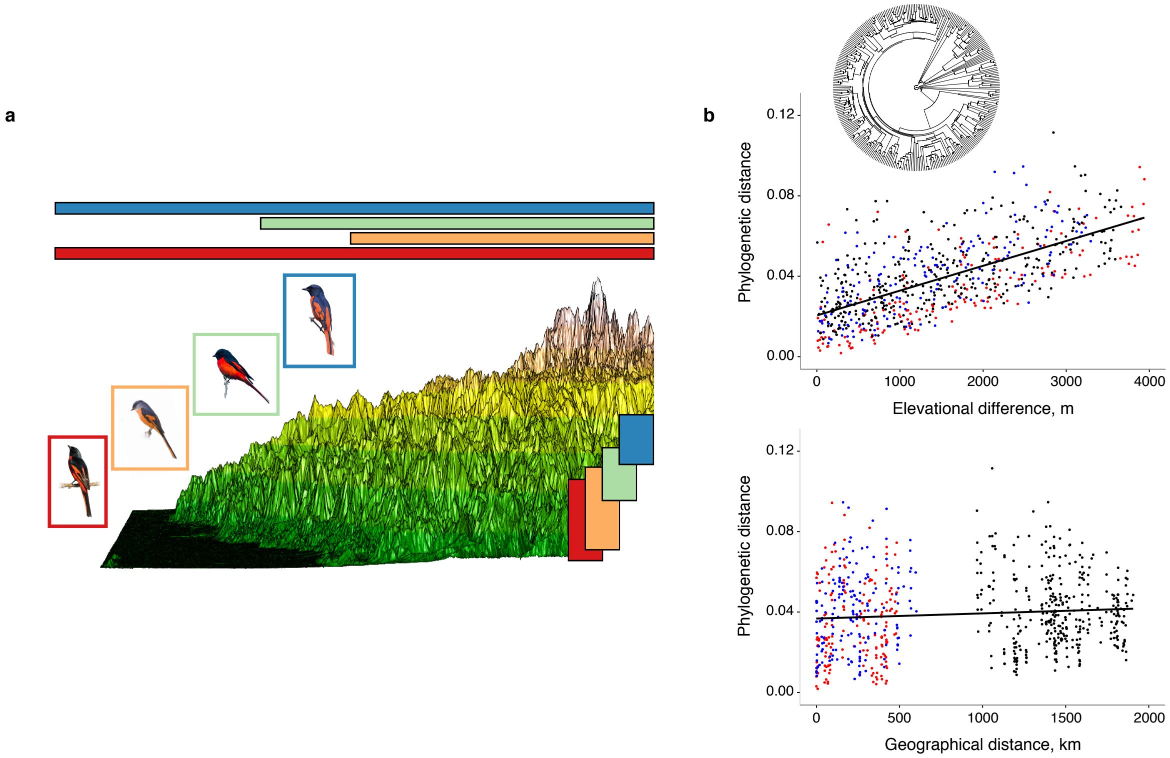


Figure 17 | Elevational and geographic similarities of communities. **a,** Horizontal bars at top indicate geographical extent of four species of minivets, *Pericrocotus,* a genus containing species across the temperate-tropical divide. Illustrated from left: *P. speciosus* (red, 150 – 2000 m in the east Himalaya), *P. solaris* (orange, 600 – 1900 m), *P. brevirostris* (light green, 1700 – 2800 m), and *P. ethologus* (blue, 2400 – 3600 m). Vertical bars at bottom left indicate elevational distributions of the four species in the east Himalaya **b,** Phylogenetic beta diversity, a metric based on phylogenetic distance between species in two communities, weighted by species abundances (phylogeny from Price et al (2014) is illustrated). Methods for assessing pairwise similarities or distances between communities have proliferated (Tucker et al. 2017). We use a similarity metric which is inversely proportional to the average within site species phylogenetic distances relative to the average between site species phylogenetic distances, weighted by species abundance (S_Q calculated using $S_{\text{Sørensen}}$, Pavoine and Ricotta 2014). Red points compare sites in the east Himalaya, blue in the west Himalaya and black between the east and west Himalaya. Lines are least squares regression lines fit to all the points (elevation: $\beta = 1.24 \times 10^{-2}$ units/km geography: $\beta = 2.56 \times 10^{-6}$ units/km). Communities separated by 500 m. of elevation differ as much as they do from one end of the Himalaya to the other (~2,000 km), partly because many species have long, elevationally-restricted ranges, as illustrated by the minivets in panel **a**.

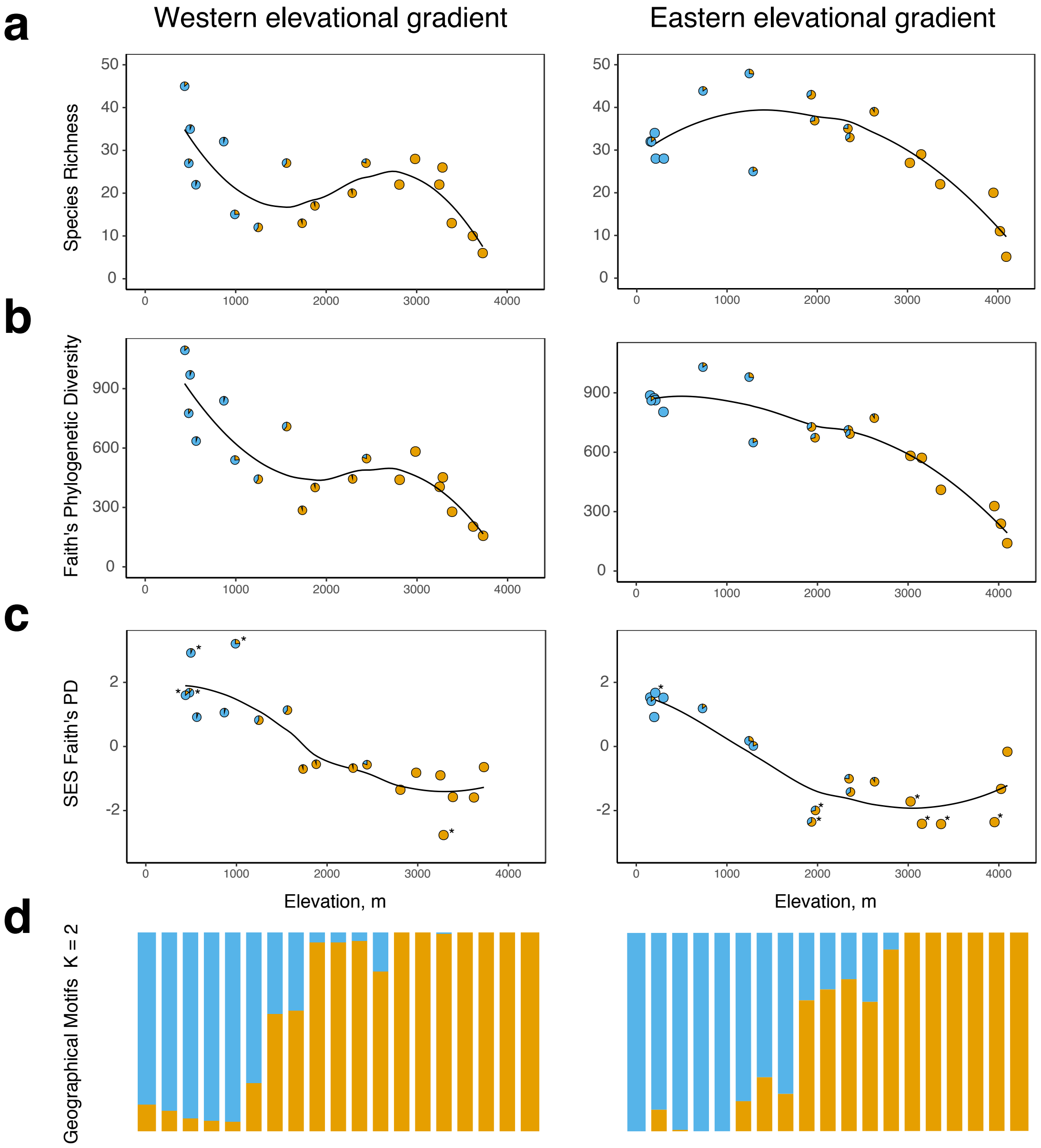


Figure 18 | Species richness and phylogenetic alpha diversity in communities. a, Species richness plotted against the elevation of each community, arranged by region. **b,** Faith's phylogenetic diversity (PD, abundance weighted)(Faith 1992) plotted against elevation. **c,** Standardized effect size of Faith's PD vs. elevation. PD measures the total branch length contained in a phylogeny that connects all species in the sample. Asterisks (*) designate significantly dispersed or clustered communities compared to 1000 simulated communities of the same richness. **d,** Regional motifs with sites ordered along the elevation axis in each region (K = 2). All points are colour coded according to the proportional contribution of regional motifs when K = 2. Note that regional mixing does not elevate richness or phylogenetic diversity.

3 Quantifying geographic and phylogenetic structure in birds and mammals across Wallace's line

3.1 Abstract

Visualizing patterns of biodiversity traditionally involves organizing locales into discrete categories according to the species shared between them (i.e., biotas). It is not required, however, that species should be organized into geographically coherent biotas, as there may be local ecological conditions to which individual species are well adapted and a sporadic distribution of those habitats may result in a geographically disjunct biota. Traditional clustering methods for identifying biotas are poorly suited to identify these patterns, as they may only appear as a component of an individual location, especially at the resolutions at which macroecological studies are commonly carried out. Here, we leverage a continuous clustering method to identify biotas of birds and mammals in the Indo-Pacific, where we specifically address the turnover of biotas from Australia to Asia across Wallace's line. We incorporate species level phylogenies for each group to evaluate the phylogenetic diversity of biotas, as well as to generate a clustering that incorporates phylogenetic binning. We compare results to the historical debates around Wallace's original partitioning of the region and the subsequent proposed partitions of Huxley, Weber, and Lydekker. We find geographically incoherent biotas in birds generally related to coastlines and wetlands but find strong geographical partitioning in mammals, suggesting dispersal plays a key role in the potential for biotas to form in the face of unique local conditions.

3.2 Introduction

Terrestrial biodiversity is structured according to unique geographical associations that are shared among disparate taxa; historically these associations have been identified as *realms*

(Wallace 1876; Holt et al. 2013). These realms, however, are not geographical *per se* but rather the spatial representation of distinctive co-occurrence patterns among species (i.e., biotas). Spatial turnover between biotas is most often associated with climatic transitions or large impassable gaps (Antonelli 2017; Ficetola et al. 2017), as exemplified by the tropical-temperate transition or large phylogenetic differences between tropical species on different continents. These patterns suggest that biotas form as a consequence of geographical proximity; despite this common interpretation, however, biotas are by no means required to be geographically cohesive. For example, unique ecological conditions in a sporadically dominant habitat (e.g., a wetland or coastal dune) may drive the formation of a unique assemblage, particularly if species are able to access these locations without dispersal limitation. Across a large geographic expanse (e.g., a continent), such an assemblage may appear in a patchy distribution (thus geographically disjunct) and yet still remain an important component of the overall structure and partitioning of the region into biologically relevant groups. The connectivity of these patches is the integral consideration, and the extent to which assemblages form as a consequence of unique local conditions may therefore be related to dispersal. For terrestrial consumers, volant taxa like birds may be more likely to form geographically incoherent biotas. Alternatively, dispersal may be sufficiently permitted such that local conditions pervasively account for the species present in a location.

A fundamental challenge in identifying biotas formed as a consequence of local conditions is that they may be hidden within low-resolution datasets that describe the presences of species in a wide geographic area (e.g., map cell resolution > 1° × 1°). Traditional methods for identifying biotas using presence-absence data (Kreft and Jetz 2010) may overlook these species associations, particularly if the species represent only a portion of the overall diversity in the general location. A critical limitation of such methods is the requirement that locations be categorized as containing a single biota, even if they would be better explained by the mixing of two or more. Here, we leverage a technique for clustering species and locations using continuous clustering to overcome this challenge and identify biotas for birds and mammals across zone of integral importance to the study of biotic boundaries: Wallace's line.

Observing a sharp biotic transition among both birds and mammals across Indonesia, Wallace designated a line approximating a point of contact between faunal assemblages of Asian and Australian origin (Wallace 1863). This partition is of central conceptual importance to biogeography (Simpson 1977) and is still studied extensively today (Wainwright et al. 2018). Wallace's line has been attributed to historical geographic isolation owing to the geological history of the region, as the physical environments on either side of Wallace's line are roughly the same. Nevertheless, among volant species there have been multiple transgressions across the line, and biologists have long recognized that the hard boundary implied by Wallace's line is not absolute (Pelseneer 1904; Mayr 1944; Esselstyn et al. 2010), i.e., the "line is not without width" as Pelseneer (1904) observed. Indeed, the faunas of Asia and Australia, historically more isolated by higher sea level during the Pleistocene, interdigitate along the Wallace's line; there is no fine boundary across which the Australian fauna turns over to Asian. Noted intermixing across the putative line thus led to an active debate among early biogeographers about the validity of Wallace's line: seven different early biogeographers proposed some modification of Wallace's original boundary (see Simpson 1977 for a detailed review of each line in turn).

That all of these modifications hinged on different interpretations of the precise boundary of faunal turnover is unsurprising; the biological reality is that species presences and absences and their abundances do not adhere to strict geographical partitions. Simpson (1977) laid bare the unspoken rules for estimating these faunal breaks with instructional detail; "*1) If reference is to terrestrial faunas…every land area must be placed definitely in one region or another. 2) The boundaries of such regions much be definite single lines on a map where, as Mayr put it, we must 'replace one color by another.'*" Largely, these rules of discrete categorization were motivated by necessity; if one regards the exercise of designating two biotas as subjective then the quantification of their intermixing seems even more dubious. Mayr (1944) suggested that Weber's line (Pelseneer 1904; Mayr 1944) was one of "faunal balance", revealing that the lines themselves are heuristics of faunal turnover; there remains a need to interpret these boundaries as such.

Our approach is to generate both the biotas and quantify their intermixing using a probabilistic model that maximizes the likelihood of both components, finding the most likely lower dimensional factorization of the dataset into groupings, which we term motifs, and the likely proportions of mixing of those motifs in each location (Erosheva and Fienberg 2005). Leveraging this method, we jettison the rules that Simpson describes; quite literately, the visualization of the model fit is achieved using pie charts that show the intergradation of biotic contributions to each location as a mixture of colors. The model is similar to that applied in population genetics to evaluate the mixing of ancestry within individuals (ADMIXTURE, Pritchard et al. 2000) and models used in Natural Language Processing (Blei et al. 2003; Blei 2012); generally these types of models are referred to as Grade of Membership models (Erosheva and Fienberg 2005). Other modern techniques for estimating biotas include network-based approaches that, while a significant advance, still rely on clustering biotas into discrete categories (Vilhena and Antonelli 2015); these methods may identify zones of transition yet unsatisfactorily assign the species in those locations into clusters all their own.

In the analysis below, we combine two publicly available datasets describing the geographic distributions of all known birds (BirdLife International NatureServe 2014) and mammals (IUCN 2017) together with molecular phylogenies of both groups (Fritz et al. 2009; Jetz et al. 2012) to assess the biotic mixing of these groups species across the region encompassing Wallace's line (from northern Australia to the southern tip of mainland Asia). We fit our model to generate biotas, which we term species motifs, for different chosen values of partitioning (K) in both birds and mammals separately. We generate phylogenetic motifs for the same values of K to understand how modern partitions are either upheld or obliterated by the underlying phylogenetic history of birds and mammals in the region. We evaluate the extent to which biotas are geographically contiguous finding that some are not, and evaluate the transitions between biotas to comment on the various historical lines of partition (Huxley 1868; Lydekker 1896; Pelseneer 1904) that have been proposed to describe the region.

3.3 Methods

3.3.1 Species presences

We overlaid global breeding range distributions for all birds and mammals on a 1° × 1° raster grid encompassing southern Thailand, Malaysia, Borneo, Indonesia, Philippines, Papua New Guinea and Northern Australia (longitudinal and latitudinal limits, 90°E, 160°E, 18°S, 20 °N). Birds ranges were obtained from BirdLife International and NatureServe (*Bird species distribution maps of the world.* Version 7, http://datazone.birdlife.org), and downloaded on April 17, 2015. Mammal ranges were obtained from the IUCN (*IUCN Red List of Threatened Species.* Version 3, May 2017, http://www.iucnredlist.org), downloaded on March 10th, 2018. We counted each species as present in a raster cell if the species' breeding range polygon overlapped the cell. We only counted cells that were composed of more than approximately 1/16 land (estimated by setting *precision* = 0.25 in the *dsp_create_from_gdb()* function in the R package *ecostructure*). This removed a number of islands of small area and low species diversity from our analysis – indeed any assessment of co-occurrence patterns (and the underlying co-diversity, i.e., compositional, patterns of sites) is contingent on the number of cells in the analysis (Arita 2017). We reduced our analysis to larger land masses so as to not bias our clustering towards species associated with island distributions. The resulting matrices comprised 2301 bird species in 664 map cells, and 1160 mammal species in those same cells. Manipulation of species distributions was done using the R package *ecostructure* (https://kkdey.github.io/ecostructure), which provides functionality for generating presence-absence matrices from GIS data using the *fasterize*, *raster*, and *sf* packages.

3.3.2 Estimating biotas in the Indo-Pacific

We estimate biotas using presence-absence matrices of birds and mammals separately. We applied a Bernoulli version of the Grade of Membership model on each presence-absence data matrix $M_{N\times G} = ((m_{ng}))$ where m_{ng} is 0/1 based on if the species g is absent/present in map cell n.

$$m_{ng} \sim Ber\ (p_{ng})$$

where p_{ng} is the probability that species g is present in the map cell n. We assume a lower dimensional representation for p_{ng}.

$$p_{ng} = \sum_{k=1}^{K} \omega_{nk}\, \theta_{kg}$$

where

$$0 \leq \omega_{nk} \leq 1 \qquad \sum_{k=1}^{K} \omega_{nk} = 1 \qquad \forall n$$

$$0 \leq \theta_{kg} \leq 1 \qquad \forall k \ \forall g$$

Here, K represents the number of underlying motifs fitted in the model, ω_{nk} represents the proportional contribution of the k^{th} motif to map cell n and θ_{kg} is the probability that the g^{th} species is a member of the k^{th} motif. We assume non-informative Dirichlet priors on the proportions vector $\omega_{n.}$ and non-informative beta priors for each θ_{kg} (see Chapter 2 of this thesis).

We fit this model for different values of K ranging from 2 to 10. The membership proportions vector $\omega_{n.}$ for each map cell n is displayed using a pie chart, placed at the exact latitude and longitude of map cell n. At a 1° × 1° resolution, this visualization shows both the spatial distribution of the motifs and, crucially, the transition between motifs across space. The $\theta_{k.}$ vector of probabilities represents the probability that a given species is a member of species motif k and can be used to extract the species that differentially contribute to the different species motifs.

3.3.3 Phylogenetic motifs

We obtained a phylogenetic tree for all birds (Jetz et al. 2012) and a phylogenetic tree for all mammals (Fritz et al. 2009), trimming each to include only the species present in the region of interest. Due to taxonomic discrepancies between the spatial and phylogenetic data (e.g., there are 5020 species in the phylogeny, but 5406 mammals in the spatial dataset), we were left with 2167 bird species and 1022 mammal species for this portion of the analysis. Leveraging each phylogeny separately, we cut the tree at a certain time slice, T, at which point some lineages in the tree would subtend one or more related species. We obtained a binary presence/absence at each location for each tip present in the phylogeny at time slice T by determining if any of the subtended tips were > 0 in the original presence absence matrix. We then fit the Grade of Membership model to the resulting matrix data matrix with 664 map cells along the rows, and the common ancestors to species at time T along the columns. We considered values of T varying from the present day (the tips of the phylogenetic tree) to 50 Myr ago at intervals of 5 ma. At 50 million years, 174 mammal clades are retained (out of 1160 species in the present phylogeny), as are 59 bird clades (out of 2167 species).

We also assessed the phylogenetic differences (Graham and Fine 2008) between the Asian and Australian biotas using the output from the original Grade of Membership model fits for $K = 2$ (i.e., those not incorporating phylogenies). Using the $\theta_{k.}$ vector (the probabilities that each species is a member of species motif k) we sampled 250 species for each of the 2 motifs and assembled a phylogenetic tree for each. We then calculated the total branch length (Faith's PD, (Faith 1992) for each tree and the common branch length between them. We repeated this process 1000 times to generate a mean PD measure for each of the two motifs, and a mean common branch length to assess the phylogenetic overlap between them. Metrics for the phylogenies were calculated using the package *PhyloMeasures* in R (Tsirogiannis and Sandel 2015).

3.4 Results

3.4.1 Species motifs

The Grade of Membership model captures the strong biotic turnover across Wallace's line for both birds and mammals, revealing the characteristic sharp transition between the islands

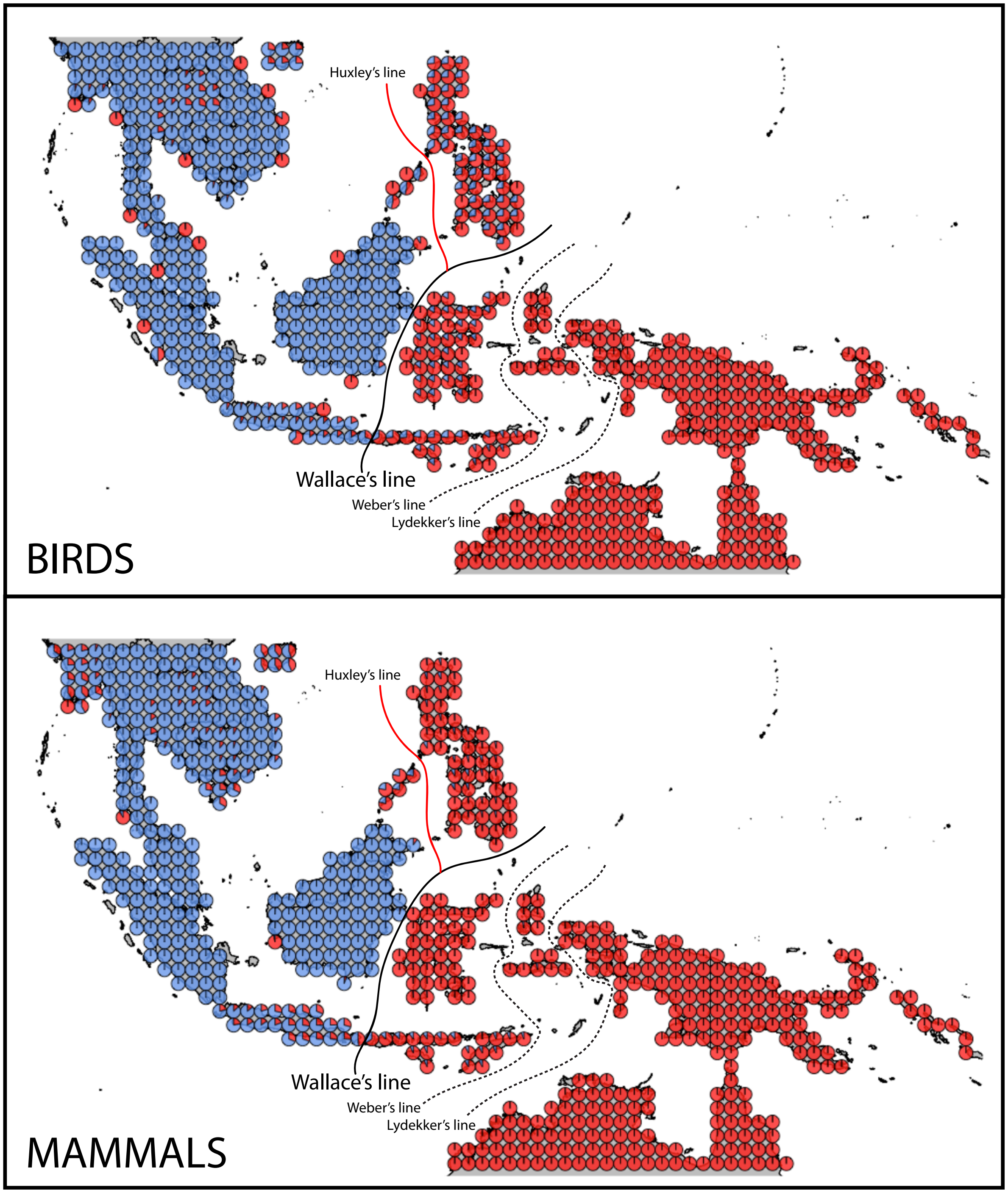


Figure 19 | Species motifs for birds and mammals in the Indo-Pacific, K=2. Each color represents membership of the species composition of the location in the 2 regional biotas

of Bali and Lombok for *K* = 2 (Fig. 19). Notably, nearly all of the species in each taxonomic subset were assigned either a 0 or 1 probability (θ_{kg}) of belonging to either the Asian or

Australian species motif. Only 43 mammal species and 97 bird species had a probability of membership between 0.01 and 0.9 (i.e., membership was mixed) for at least one of the motifs, ~4% of the total diversity of each group. For mammals, nearly all of these species are bats (*Chiroptera*), with the addition of 2 civets (*Carnivora*), 1 primate (*Macaca fascicularis*), and 4 human commensals (3 rodents, *Rodentia*, and 1 shrew, *Eulipotyphla*). The 97 bird species found to have a mixture of memberships between the two motifs represent 18 different orders of birds: *Galliformes* (4 species), *Anseriformes* (3), *Podicipediformes* (1), *Columbiformes* (5), *Caprimulgiformes* (4), *Cuculiformes* (3), *Gruiformes* (6), *Procellariformes* (4), *Ciconiiformes* (2), *Pelicaniformes* (7), *Suliformes* (4), *Charadriiformes* (10), *Strigiformes* (2), *Accipitriformes* (6), *Coraciiformes* (4), *Falconiformes* (3), *Psittaciformes* (1), *Passeriformes* (28).

Increasing values of *K* continue to partition the region into unique geographic groupings. A partition with *K*=10 shows that the motifs are largely restricted to large islands or continents for both birds and mammals (Fig. 20). However, small islands to the east of Wallace's line (e.g., Timor-Leste) show considerable mixture of the various species motifs. These islands were the focus of early biogeographers debating Wallace's original partition – results show that they are in fact an admixture of species motifs from the region. The strong partition between Borneo and Sulawesi remains for both groups at *K* = 10.

For *K* = 10, a geographically incoherent motif for birds (Fig. 20, blue motif) appears to contribute to locations from eastern Australia to mainland Asia – putatively this motif brings species together based on shared habitat (coasts or outlying islands) rather than geographic isolation. Indeed, the top 10 species contributing to this motif belong to highly dispersive orders, many of which are strongly linked to coastal habitats: *Columbiformes Prodilarriiformes, Pelicaniformes, Suliformes*, *Charadriiformes*, *and Accipitriformes*. These species include

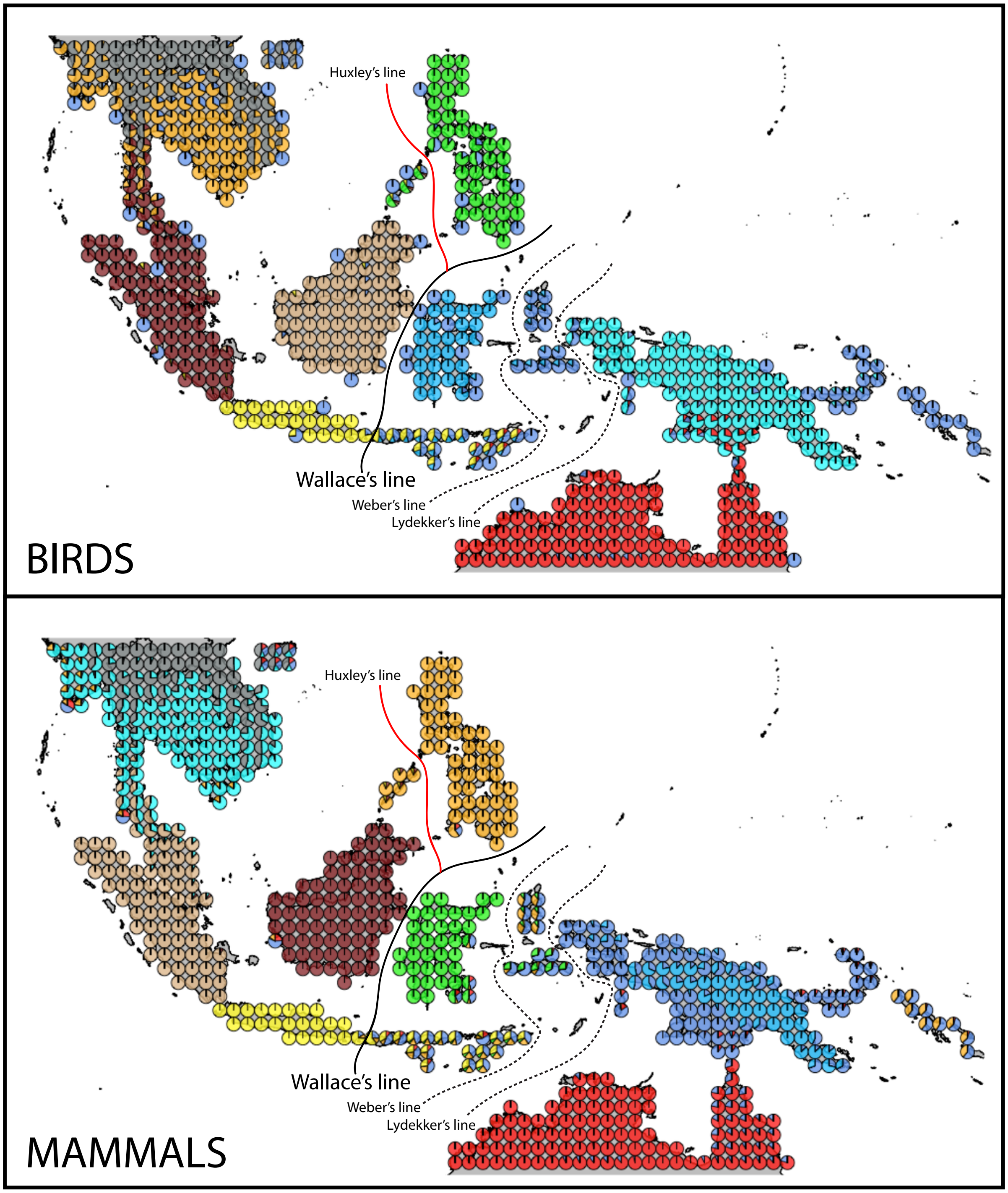


Figure 20 | Species motifs for birds and mammals in the Indo-Pacific, *K*=10. Each color represents membership of the species composition of the location in the 10 species motifs.

shearwaters, herons, boobies, a frigatebird, a tern, a pigeon, and a hawk – all either water birds, seabirds, or species with high dispersal ability associated with foraging across large areas (e.g., the pigeon and hawk), and all of which breed in large colonies excepting the hawk. Mammals show no such pattern, and nearly all of the mammal assemblages reflect the contribution of geographically contiguous motifs.

3.4.2 Phylogenetic comparison for *K* = 2

We examined the phylogenetic relationship between the Australian and Asian biotas for both mammals and birds by first taking the *K*=2 model fit from the standard analysis. Both birds and mammals show significant differences in phylogenetic diversity between the two regions, as calculated using 250 species subsets from each of the motifs and measuring PD within phylogenetic trees built from those 250 species (Fig. 21). For birds, the Asian species motif is more closely related with lower PD, owing to the more recent radiation and high diversity of oscine passerines in Asia (Price et al. 2014). Birds within the Australian motif are generally dominate the region. Mammals show an opposite pattern, where placental mammals radiated out

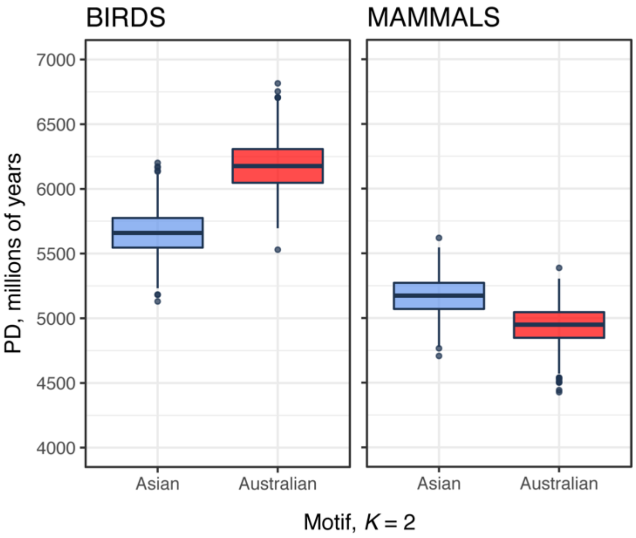


Figure 21 | Phylogenetic diversity of assemblages sampled from the membership proportions vector $\theta_{k.}$ for each of the two motifs. For each clade, 250 species were sampled and a phylogeny was generated for that subset. PD was calculated as the total branch length of the phylogeny.

older, as expected from the high number of non-passerines, basal oscines, and suboscines that of Australia and are present in large numbers in Asia. Australia, however, is dominated by a more closely related group, the marsupials, and their contribution the Australian species motif is reflected in Figure 21. A comparison of the common branches shared between the Australian and Asian motifs shows phylogenetic overlap (common branch length in millions of years) is generally low for both birds ($\bar{x} = 1663.2$, SE = 3.6, N = 1000) for mammals ($\bar{x} = 958.5$, SE = 2.2, N = 1000).

3.4.3 Phylogenetic motifs

We also generated phylogenetic motifs to observe how the spatial patterns of partitioning are upheld when species are clumped into clades of a given age. We report *K*=2 and *K*=10 for

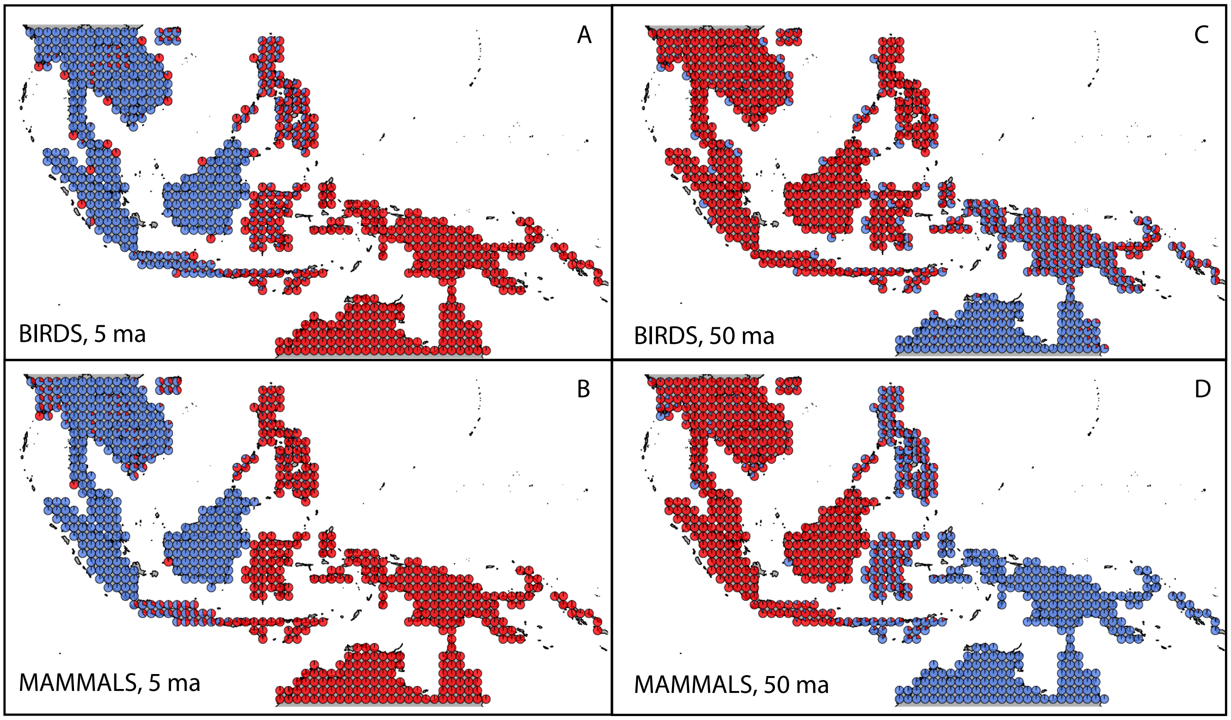


Figure 22 | Phylogenetic motifs for birds and mammals, *K* = 2. **A,B.** Motifs generated by subtended the tips of the phylogeny at 5 myr ago. **C,D.** Motifs generated by subtending the tips of the phylogeny at 50 myr ago.

both birds and mammals, with clade cutoffs at 5 myr ago and 50 myr ago. The primary partitions across the Indo-pacific have a deep history for both groups (Fig. 22), although mammals appear to have a deeper and more structured partition at the 50-ma timeline. Locations in New Guinea show association with both bird phylogenetic motifs at the 50-ma cutoff, yet mixture between the motifs only stretches to Sulawesi and Philippines for mammals.

A higher-level partition of the region into phylogenetic motifs ($K = 10$) shows much more mixing, in particular deep into the phylogeny at 50 myr (Fig. 23). Closely related birds on the Asian mainland collapse into one phylogenetic motif at 5 myr ago (compare Fig. 22 with Fig. 23) as do birds on Sumatra and Borneo, creating new partitions for the montane region of New Guinea (light blue, Figure 23A) and drier habitat birds on the interior of Australia (dark brown,

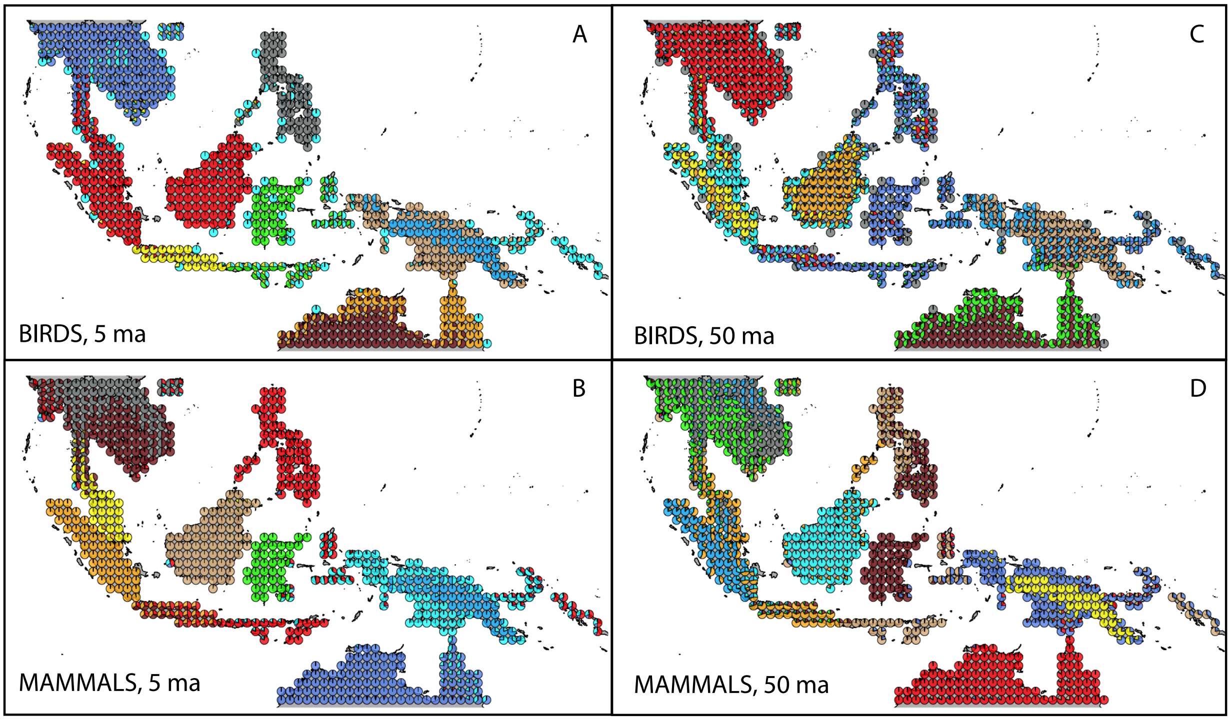


Figure 23 | Phylogenetic motifs for birds and mammals, $K = 10$. **A,B.** Motifs generated by subtended the tips of the phylogeny at 5 myr ago. **C,D.** Motifs generated by subtending the tips of the phylogeny at 50 myr ago.

Fig. 23A). Where Java was partitioned into a single species motif in present day mammals (yellow motif, Figure 20), a new split appears between mammals on Malaysia and Sumatra at 5 ma back, likely due to allospecies clumping within those regions, causing Java to be represented by multiple phylogenetic motifs (Fig. 23B). The geographically incoherent motif revealed in the present-day for birds (blue motif, Fig. 20) remains an important partition even back to 50 ma.

3.5 Discussion

Geographical structure in the biotic composition of a region is strongly expected in the face of large differences in climate or extreme barriers to dispersal; absent these factors, sharp biotic turnover begs further investigation. It is for this reason that Wallace and others were so struck by the sharp transition in biotic assemblages across Indonesia, Borneo, and the Philippines. In this study, we examined the extent to which bird and mammal assemblages turn over across the Indo-Pacific, finding that while these two groups shared a number of primary patterns in their geographic structure, differences appear to be related to gross differences in dispersal ability. Notably, our analysis reveals that intermixing between Asian and Australian biotas across Wallace's line is primary driven by bats in mammals, while a number of distantly related orders of birds contribute to interdigitation. That volant organisms overcome this biotic partition is perhaps unsurprising, however it is not known to what extent dispersal *per se* or associated ecological characteristics (e.g., foraging behaviors) are responsible for movement across the line. The difficulty in disentangling the ability to disperse from other explanations for the presence of a species in a given region owing to ecology remains a persistent challenge (White 2016; Kennedy et al. 2017).

One partition stands out in our analysis – the geographically incoherent motif of birds that appears to persist among different values of *K* not only in the present but also deep into the phylogeny. We suggest this motif is primarily associated with the local ecological conditions of coastal habitats and wetlands – this is most evident in evaluating the contribution of the cyan phylogenetic motif in Figure 23A, where one can observe a small contribution of this motif to locations covered in wetlands in central Thailand. We find this same motif contributing to outlying islands east of New Guinea and scattered coastlines throughout the region. The species forming this motif also appear to be unique phylogenetically, as the motif persists to the 50-ma cutoff (Fig. 23C).

While our clustering indeed supports the partition proposed by Wallace, the transition between Asian and Australian biotas shows various levels of intergradation across the line depending on which locations are compared (Fig. 19). For example, Borneo and Sulawesi are very well partitioned for mammals despite their proximity, while the Asian biota appears to make a larger contribution to Sulawesi in birds (Fig. 19). Java is not strictly Asian, as the various historical lines claim, showing a mixture of both Asian and Australian elements in both birds and mammals. As a general comparison, Huxley's modification of Wallace's line, as described by Mayr (1944), appears to more accurately predict the partitioning of mammals (with the exception of Palawan) than for birds, with Philippine birds having ~25% membership in the Asian biota. Palawan shows levels of intergradation for both taxonomic groups, and the analysis of phylogenetic motifs supports the conclusions of Esselstyn et al. (2010) that Palawan and Philippines are indeed closely associated in their biogeographic history. Huxley's line is thus less informative in light of molecular studies, but clearly was an obvious improvement over Wallace's initial suggestion that Philippines showed a stronger affinity to Asia.

Weber's and Lydekker's lines reflect the uncertainly around binning the outlying islands of Timor-Leste, Babar, and others west of New Guinea and south of Sulawesi. Our model fit shows these islands are in a zone of transition at higher levels of partitioning (*K*=10), with components from a number of regional biotas (Figure 20); however, for a primary partition of the region into 2 biotas, Wallace's original line (with Huxley's modification) conforms more strongly with our analysis (Figure 19) compared to those lines that were proposed later.

Data Accessibility

Geographical barriers and dispersal propensity interact to limit range expansions of Himalayan birds

- *Publication*: (White 2016)

Regional influences on community structure across the tropical-temperate divide

- *Publication*: (White, Dey et al. *In Review*)

Quantifying geographic and phylogenetic structure in birds and mammals across Wallace's line

- Data are publicly available from BirdLife International, IUCN, Fritz et al. (2009), and Jetz et al. (2012).

References

Amadon, D. The superspecies concept. 1966. Systematic Zoology 15:245–249.

Antonelli, A. 2017. Drivers of bioregionalization. Nature Ecology & Evolution 1:1–2.

Arita, H. T. 2017. Multisite and multispecies measures of overlap, co-occurrence, and co-diversity. Ecography 40:709–718.

Arita, H. T., J. A. Christen, P. Rodríguez, and J. Soberón. 2008. Species diversity and distribution in presence-absence matrices: mathematical relationships and biological implications. The American Naturalist 172:519–532.

Assis, J., E. A. Serrão, B. Claro, C. Perrin, and G. A. Pearson. 2014. Climate-driven range shifts explain the distribution of extant gene pools and predict future loss of unique lineages in a marine brown alga. Molecular Ecology 23:2797–2810.

Barton, N. H., and G. M. Hewitt. 1985. Analysis of Hybrid Zones. Annual Review of Ecology and Systematics 16:113–148.

Benn, D. I., and L. A. Owen. 1998. The role of the Indian summer monsoon and the mid-latitude westerlies in Himalayan glaciation: review and speculative discussion. Journal of the Geological Society 155:353–363.

Bhattacharyya, A., N. Mehrotra, S. K. Shah, N. Basavaiah, V. Chaudhary, and I. B. Singh. 2014. Analysis of vegetation and climate change during Late Pleistocene from Ziro Valley, Arunachal Pradesh, Eastern Himalaya region. Quaternary Science Reviews 101:111–123.

BirdLife International, NatureServe. 2014. Bird species distribution maps of the world. Cambridge, UK and Arlington, USA.

Blei, D. M., A. Y. Ng, and M. I. Jordan. 2003. Latent Dirichlet Allocation. Journal of Machine Learning Research 3:233–269.

Blei, D. M. 2012. Probabilistic topic models. Communications of the ACM 55:77–8.

Breheny, P., and W. Burchett. 2016. visreg: Visualization of Regression Models. R package version 2.2-2. https://CRAN.R-project.org/package=visreg

Briers, R. A. 2003. Range limits and parasite prevalence in a freshwater snail. Proceedings of the Royal Society B: Biological Sciences 270:S178–S180.

Broennimann, O., M. C. Fitzpatrick, P. B. Pearman, B. Petitpierre, L. Pellissier, N. G. Yoccoz, W. Thuiller, M. Fortin, C. Randin, N. E. Zimmermann, C. H. Graham, and A. Guisan. 2012. Measuring ecological niche overlap from occurrence and spatial environmental data. Global Ecology and Biogeography 21:481–497.

Burney, C. W., and R. T. Brumfield. 2009. Ecology Predicts Levels of Genetic Differentiation in Neotropical Birds. The American Naturalist 174:358–368.

Carstensen, D. W., J.-P. Lessard, B. G. Holt, M. Krabbe Borregaard, and C. Rahbek. 2013. Introducing the biogeographic species pool. Ecography 36:1310–1318.

Cavender-Bares, J., D. D. Ackerly, S. E. Hobbie, and P. A. Townsend. 2016. Evolutionary legacy effects on ecosystems: biogeographic origins, plant traits, and implications for management in the era of global change. Annual Review of Ecology, Evolution, and Systematics 47:433–462.

Claramunt, S., E. P. Derryberry, J. V. Remsen, and R. T. Brumfield. 2012. High dispersal ability inhibits speciation in a continental radiation of passerine birds. Proceedings of the Royal Society B: Biological Sciences 279:1567–1574.

Davies, R. G., C. D. L. Orme, D. Storch, V. A. Olson, G. H. Thomas, S. G. Ross, T. S. Ding, P. C. Rasmussen, P. M. Bennett, I. P. F. Owens, T. M. Blackburn, and K. J. Gaston. 2007. Topography, energy and the global distribution of bird species richness. Proceedings of the Royal Society B: Biological Sciences 274:1189–1197.

Davis, M. B. 1986. Climatic Instability, Time Lags, and Community Disequilibrium. Pages 269 - 284 in J. Diamond & T. J. Case, eds. Community Ecology. Harper and Rowe, New York.

Davis, M. B., and R. G. Shaw. 2001. Range Shifts and Adaptive Responses to Quaternary Climate Change. Science 292:673–679.

Dawideit, B. A., A. B. Phillimore, I. Laube, B. Leisler, and K. Böhning-Gaese. 2009. Ecomorphological predictors of natal dispersal distances in birds. Journal of Animal Ecology 78:388–395.

Demske, D., P. E. Tarasov, B. Wünnemann, and F. Riedel. 2009. Late glacial and Holocene vegetation, Indian monsoon and westerly circulation in the Trans-Himalaya recorded in the lacustrine pollen sequence from Tso Kar, Ladakh, NW India. Palaeogeography, Palaeoclimatology, Palaeoecology 279:172–185.

Devictor, V., R. Julliard, D. Couvet, and F. Jiguet. 2008. Birds are tracking climate warming, but not fast enough. Proceedings of the Royal Society B: Biological Sciences 275:2743–2748.

Dey, K. K., C. J. Hsiao, and M. Stephens. 2017. Visualizing the structure of RNA-seq expression data using grade of membership models. PLOS Genetics 13:e1006599–23.

Diamond, J. M. 1981. Flightlessness and fear of flying in island species. Nature 293:507–508.

Elsen, P. R., M. W. Tingley, R. Kalyanaraman, K. Ramesh, and D. S. Wilcove. 2017. The role of competition, ecotones, and temperature in the elevational distribution of Himalayan birds. Ecology 98:337–348.

Enzel, Y., L. L. Ely, S. Mishra, R. Ramesh, R. Amit, B. Lazar, S. N. Rajaguru, V. R. Baker, and A. Sandler. 1999. High-Resolution Holocene Environmental Changes in the Thar Desert, Northwestern India. Science 284:125–128.

Erosheva, E. A., and S. E. Fienberg. 2005. Bayesian mixed membership models for soft clustering and classification. Pages 11–26 in Classification - the Ubiquitous Challenge. Springer, Berlin/Heidelberg.

Esselstyn, J. A., C. H. Oliveros, R. G. Moyle, A. T. Peterson, J. A. McGuire, and R. M. Brown. 2010. Integrating phylogenetic and taxonomic evidence illuminates complex biogeographic patterns along Huxley's modification of Wallace's Line. Journal of Biogeography 37:2054–2066.

Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. Biological Conservation 61:1–10.

Ficetola, G. F., F. Mazel, and W. Thuiller. 2017. Global determinants of zoogeographical boundaries 1–7.

Field, R., B. A. Hawkins, H. V. Cornell, D. J. Currie, J. A. F. Diniz-Filho, J.-F. Guégan, D. M. Kaufman, J. T. Kerr, G. G. Mittelbach, T. Oberdorff, E. M. O'Brien, and J. R. G. Turner. 2009. Spatial species-richness gradients across scales: a meta-analysis. Journal of Biogeography 36:132–147.

Fine, P. V. A. 2015. Ecological and evolutionary drivers of geographic variation in species diversity. Annual Review of Ecology, Evolution, and Systematics 46:369–392.

Finkel, R. C., L. A. Owen, P. L. Barnard, and M. W. Caffee. 2003. Beryllium-10 dating of Mount Everest moraines indicates a strong monsoon influence and glacial synchroneity throughout the Himalaya. Geology 31:561–564.

Fjeldså, J., R. C. K. Bowie, and C. Rahbek. 2012. The Role of Mountain Ranges in the Diversification of Birds. Annual Review of Ecology, Evolution, and Systematics 43:249–265.

Fritz, S. A., O. R. P. Bininda-Emonds, and A. Purvis. 2009. Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. Ecology Letters 12:538–549.

Garrick, R. C., J. D. Nason, J. F. Fernández-Manjarrés, and R. J. Dyer. 2013. Ecological coassociations influence species' responses to past climatic change: an example from a Sonoran Desert bark beetle. Molecular Ecology 22:3345–3361.

Ghosh-Harihar, M., and T. D. Price. 2014. A test for community saturation along the Himalayan bird diversity gradient, based on within-species geographical variation. Journal of Animal Ecology 83:628–638.

Gotelli, N. J. 2000. Null model analysis of species co-occurrence patterns. Ecology 81:2606–2621.

Graham, C. H., and P. V. A. Fine. 2008. Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. Ecology Letters 11:1265–1277.

Grant, P. R., and B. R. Grant. 2008. How and why species multiply : the radiation of Darwin's finches. Princeton University Press, Princeton.

Graves, G. R., and C. Rahbek. 2005. Source pool geometry and the assembly of continental avifaunas. Proceedings of the National Academy of Sciences 102:7871–7876.

Harrison, S., and H. Cornell. 2008. Toward a better understanding of the regional causes of local community richness. Ecology Letters 11:969–979.

Harrison, S., and H. V. Cornell. 2007. Introduction: Merging Evolutionary and Ecological Approaches to Understanding Geographic Gradients in Species Richness. The American Naturalist 170:S1–S4.

Herzschuh, U. 2006. Palaeo-moisture evolution in monsoonal Central Asia during the last 50,000 years. Quaternary Science Reviews 25:163–178.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25:1965–1978.

Hillebrand, H. 2004. On the generality of the latitudinal diversity gradient. The American Naturalist 163:192–211.

Holt, B. G., J. P. Lessard, M. K. Borregaard, S. A. Fritz, M. B. Araujo, D. Dimitrov, P. H. Fabre, et al. 2013. An update of Wallace's zoogeographic regions of the world. Science 339:74–78.

Humboldt, A. V., and A. Bonpland. 1807. Essay on the Geography of Plants. (S. Romanowski, trans.). Paris.

Hurlbert, A. H., and W. Jetz. 2007. Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. Proceedings of the National Academy of Sciences 104:13384–13389.

Huxley, T. H. 1868. On the classification and distribution of the Alectoromorphae and Heteromorpha. Proceedings of the Zoological Society of London 36:622–637.

IUCN. The IUCN Red List of Threatened Species. Version 2017. (IUCN, ed.).

Jain, M., and S. K. Tandon. 2003. Fluvial response to Late Quaternary climate changes, western India. Quaternary Science Reviews 22:2223–2235.

James, H. F., and D. A. Burney. 1997. The diet and ecology of Hawaii's extinct flightless waterfowl: evidence from coprolites. Biological Journal of the Linnean Society 62:279–297.

Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann, and A. O. Mooers. 2012. The global diversity of birds in space and time. Nature 491:444–448.

Kadmon, R. 1995. Nested species subsets and geographic isolation: a case study. Ecology 76:458–465.

Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, et al. 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26:1463–1464.

Kennedy, J. D., M. K. Borregaard, K. A. Jønsson, B. Holt, J. Fjeldså, and C. Rahbek. 2016. Does the colonization of new biogeographic regions influence the diversification and accumulation of clade richness among the Corvides (Aves: Passeriformes)? Evolution 71:38–50.

Kipp, F. A. 1959. Der Handflügel-Index als flugbiologisches Maß. Vogelwarte 21:28–36.

Kreft, H., and W. Jetz. 2010. A framework for delineating biogeographical regions based on species distributions. Journal of Biogeography 37:2029–2053.

Kruskal, W. 1987. Relative importance by averaging over orderings. The American Statistician 41:6–10.

Lessard, J. P., M. K. Borregaard, J. A. Fordyce, C. Rahbek, M. D. Weiser, R. R. Dunn, and N. J. Sanders. 2011. Strong influence of regional species pools on continent-wide structuring of local communities. Proceedings of the Royal Society of London B: Biological Sciences 279:266–274.

Liechti, F., W. Witvliet, R. Weber, and E. Bächler. 2013. First evidence of a 200-day non-stop flight in a bird. Nature Communications 4:1–7.

Loarie, S. R., B. E. Carter, K. Hayhoe, S. McMahon, R. Moe, C. A. Knight, and D. D. Ackerly. 2008. Climate Change and the Future of California's Endemic Flora. PLoS ONE 3:e2502.

Lockwood, R., J. P. Swaddle, and J. M. V. Rayner. 1998. Avian wingtip shape reconsidered: wingtip shape indices and mophological adaptations to migration. Journal of Avian Biology 29:273–292.

Lydekker, R. 1896. A Geographical History of Mammals. University Press, Cambridge.

MacArthur, R. H. 1972. Geographical Ecology: Patterns in the Distribution of Species. Harper and Rowe, New York.

Mayr, E. 1944. Wallace's Line in the Light of Recent Zoogeographic Studies. The Quarterly review of biology 19:1–14.

Menéndez, R., A. G. Megias, J. K. Hill, B. Braschler, S. G. Willis, Y. Collingham, R. Fox, D. B. Roy, and C. D. Thomas. 2006. Species richness changes lag behind climate change. Proceedings of the Royal Society B: Biological Sciences 273:1465–1470.

Miklós, I., and J. Podani. 2004. Randomization of presence-absence matrices: comments and new algorithms. Ecology 85:86–92.

Mittelbach, G. G., D. W. Schemske, H. V. Cornell, A. P. Allen, J. M. Brown, M. B. Bush, S. P. Harrison, et al. 2007. Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. Ecology Letters 10:315–331.

Moen, D., and H. Morlon. 2014. Why does diversification slow down? Trends in Ecology & Evolution 29:190–197.

Moore, R. P., W. D. Robinson, I. J. Lovette, and T. R. Robinson. 2008. Experimental evidence for extreme dispersal limitation in tropical forest birds. Ecology Letters 11:960–968.

Munguía, M., C. Rahbek, T. F. Rangel, J. A. F. Diniz-Filho, and M. B. Araújo. 2012. Equilibrium of Global Amphibian Species Distributions with Climate. PLoS ONE 7:e34420.

Normand, S., R. E. Ricklefs, F. Skov, J. Bladt, O. Tackenberg, and J. C. Svenning. 2011. Postglacial migration supplements climate in determining plant species ranges in Europe. Proceedings of the Royal Society B: Biological Sciences 278:3644–3653.

Orme, D., R. Freckleton, G. Thomas, T. Petzoldt, S. Fritz, N. Isaac, and W. Pearse. 2013. caper: Comparative Analyses of Phylogenetics and Evolution in R. R package version 0.5.2. https://CRAN.R-project.org/package=caper

Overpeck, J., D. Anderson, S. Trumbore, and W. Prell. 1996. The southwest Indian Monsoon over the last 18,000 years. Climate Dynamics 12:213–225.

Owen, L. A., U. Kamp, J. Q. Spencer, and K. Haserodt. 2002. Timing and style of Late Quaternary glaciation in the eastern Hindu Kush, Chitral, northern Pakistan: a review and revision of the glacial chronology based on new optically stimulated luminescence dating. Quaternary International 97-98:41–55.

Paradis, E., S. R. Baillie, W. J. Sutherland and R. D. Gregory. 1998. Patterns of natal and breeding dispersal in birds. Journal of Animal Ecology 67:518–536.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20:289–290.

Parmesan, C., and G. Yohe. 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature 421:37–42.

Pavoine, S., and C. Ricotta. 2014. Functional and phylogenetic similarity among communities. (D. Orme, ed.)Methods in Ecology and Evolution 5:666–675.

Päckert, M., J. Martens, Y.-H. Sun, L. L. Severinghaus, A. A. Nazarenko, J. Ting, T. Töpfer, and D. T. Tietze. 2012. Horizontal and elevational phylogeographic patterns of Himalayan and Southeast Asian forest passerines (Aves: Passeriformes). Journal of Biogeography 39:556–573.

Pearson, R. G., and T. P. Dawson. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography 12:361–371.

Pelseneer, P. 1904. La ligne de Weber, limite zoologique de l'Asie et de l'Australie. Bulletin de la Classe des Sciences Académie Royale de Belgique 1904:1001–1022.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190:231–259.

Pigot, A. L., and J. A. Tobias. 2013. Species interactions constrain geographic range expansion over evolutionary time. Ecology Letters 16:330–338.

Pigot, A. L., and J. A. Tobias. 2014. Dispersal and the transition to sympatry in vertebrates. Proceedings of the Royal Society B: Biological Sciences 282:20141929.

Pigot, A. L., I. P. F. Owens, and C. D. L. Orme. 2010. The environmental limits to geographic range expansion in birds. Ecology Letters 13:705–715.

Pollock, L. J., W. Thuiller, and W. Jetz. 2017. Large conservation gains possible for global biodiversity facets. Nature 546:141–144.

Pounds, J. A., M. Fogden, and J. H. Campbell. 1999. Biological response to climate change on a tropical mountain. Nature 398:611–615.

Price, T. D., and M. Kirkpatrick. 2009. Evolutionarily stable range limits set by interspecific competition. Proceedings of the Royal Society B: Biological Sciences 276:1429–1434.

Price, T. D., D. M. Hooper, C. D. Buchanan, U. S. Johansson, D. T. Tietze, P. Alström, U. Olsson, M. Ghosh-Harihar, F. Ishtiaq, S. K. Gupta, J. Martens, B. Harr, P. Singh, and D. Mohan. 2014. Niche filling slows the diversification of Himalayan songbirds. Nature 509:222–225.

Price, T. D., D. Mohan, D. T. Tietze, D. M. Hooper, C. D. L. Orme, and P. C. Rasmussen. 2011. Determinants of Northerly Range Limits along the Himalayan Bird Diversity Gradient. The American Naturalist 178:S97–S108.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Quintero, I., S. González-Caro, P.-C. Zalamea, and C. D. Cadena. 2014. Asynchrony of Seasons: Genetic Differentiation Associated with Geographic Variation in Climatic Seasonality and Reproductive Phenology. The American Naturalist 184:352–363.

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Ray, N., and J. Adams. 2001. A GIS-based vegetation map of the world at the Last Glacial Maximum (25,000-15,000 BP). Internet Archaeology 11. http://dx.doi.org/10.11141/ia.11.2

Revell, L. J. 2011. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3:217–223.

Ricklefs, R. E. 2004. A comprehensive framework for global patterns in biodiversity. Ecology Letters 7:1–15.

Ricklefs, R. E. 2008. Disintegration of the ecological community. The American Naturalist 172:741–750.

Ronce, O. 2007. How Does It Feel to Be like a Rolling Stone? Ten Questions about Dispersal Evolution. Annual Review of Ecology, Evolution, and Systematics 38:231–253.

Rosenberg, N. A., J. K. Pritchard, J. L. Weber, H. M. Cann, K. K. Kidd, L. A. Zhivotovsky, and M. W. Feldman. 2002. Genetic structure of human populations. Science 298:2381–2385.

Sandel, B., L. Arge, B. Dalsgaard, R. G. Davies, K. J. Gaston, W. J. Sutherland, and J. C. Svenning. 2011. The Influence of Late Quaternary Climate-Change Velocity on Species Endemism. Science 334:660–664.

Sexton, J. P., P. J. McIntyre, A. L. Angert, and K. J. Rice. 2009. Evolution and Ecology of Species Range Limits. Annual Review of Ecology, Evolution, and Systematics 40:415–436.

Simberloff, D., and E. F. Connor. 1979. Q-mode and R-mode analyses of biogeogrpahic distributions: Null hypotheses based on random colonization. Pages 123–138 in G. P. Patil and M. L. Rosenzweig, eds. Contemporary quantitative ecology and related ecometrics. Burtonsville, MD.

Simpson, G. G. 1977. Too many lines; the limits of the Oriental and Australian zoogeographic regions. Proceedings of the American Philosophical Society 121:107–120.

Singh, G., R. J. Wasson, and D. P. Agrawal. 1990. Vegetational and seasonal climatic changes since the last full glacial in the Thar Desert, northwestern India. Review of Palaeobotany and Palynology 64:351–358.

Smith, B. T., J. E. McCormack, A. M. Cuervo, M. J. Hickerson, A. Aleixo, C. D. Cadena, J. Pérez-Emán, C. W. Burney, X. Xie, M. G. Harvey, B. C. Faircloth, T. C. Glenn, E. P. Derryberry, J. Prejean, S. Fields, and R. T. Brumfield. 2014. The drivers of tropical speciation. Nature 515:406–409.

Smith, C. I., and B. D. Farrell. 2005. Range expansions in the flightless longhorn cactus beetles, Moneilema gigas and Moneilema armatum, in response to Pleistocene climate changes. Molecular Ecology 14:1025–1044.

Srinivasan, U., K. Tamma, and U. Ramakrishnan. 2014. Past climate and species ecology drive nested species richness patterns along an east-west axis in the Himalaya. Global Ecology and Biogeography 23:52–60.

Svenning, J.-C., and F. Skov. 2004. Limited filling of the potential range in European tree species. Ecology Letters 7:565–573.

Taddy, M. A. 2012. On estimation and selection for topic models. Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics 22:1184–1193.

Thomas, C. D., A. Cameron, R. E. Green, M. Bakkenes, L. J. Beaumont, Y. C. Collingham, B. F. N. Erasmus, M. F. de Siqueira, A. Grainger, L. Hannah, l. Hughes, B. Huntley, A. S. van Jaarsveld, G. F. Midgley, L. Miles, M. A. Ortega-Huerta, A. T. Peterson, O. L. Phillips, and S. E. Williams. 2004. Extinction risk from climate change. Nature 427:145–148.

Trivedi, A., M. S. Chauhan, A. Sharma, C. M. Nautiyal, and D. P. Tiwari. 2013. Record of vegetation and climate during Late Pleistocene–Holocene in Central Ganga Plain, based on multiproxy data from Jalesar Lake, Uttar Pradesh, India. Quaternary International 306:97–106.

Tsirogiannis, C., and B. Sandel. 2015. PhyloMeasures: a package for computing phylogenetic biodiversity measures and their statistical moments. Ecography 39:709–714.

Tucker, C. M., M. W. Cadotte, S. B. Carvalho, T. J. Davies, S. Ferrier, S. A. Fritz, R. Grenyer, et al. 2017. A guide to phylogenetic metrics for conservation, community ecology and macroecology. Biological Reviews 92:692–715.

Tung Ho, L. S., and C. Ane. 2014. A Linear-Time Algorithm for Gaussian and Non-Gaussian Trait Evolution Models. Systematic Biology 63:397–408.

Valle, D., B. Baiser, C. W. Woodall, and R. Chazdon. 2014. Decomposing biodiversity data using the Latent Dirichlet Allocation model, a probabilistic multivariate statistical method. Ecology Letters 17:1591–1601.

van der Maaten, L., and G. Hinton. 2008. Visualizing data using t-SNE. Journal of Machine Learning Research 9:2579–2605.

van Etten, J. 2015. gdistance: Distances and Routes on Geographical grids. R package version 1.1-9. https://CRAN.R-project.org/package=gdistance

Vilhena, D. A., and A. Antonelli. 2015. A network approach for identifying and delimiting biogeographical regions. Nature Communications 6:1–9.

Wainwright, B. J., G. L. Zahn, I. S. Arlyza, and A. S. Amend. 2018. Seagrass-associated fungal communities follow Wallace's line, but host genotype does not structure fungal community. Journal of Biogeography 45:762–770.

Wallace, A. R. 1863. On the physical geography of the Malay Archipelago. Proceedings of the Royal Geographical Society 7:206–213.

Wallace, A. R. 1876. The Geographical Distribution of Animals. Cambridge University Press, Cambridge.

Wang, Y., X. Liu, and U. Herzschuh. 2010. Asynchronous evolution of the Indian and East Asian Summer Monsoon indicated by Holocene moisture patterns in monsoonal central Asia. Earth-Science Reviews 103:135–153.

Wattenberg, M., F. Viégas, and I. Johnson. 2016. How to use t-SNE effectively. Distill 1.

Wethey, D. S. 2002. Biogeography, Competition, and Microclimate: The Barnacle Chthamalus fragilis in New England. Integrative and Comparative Biology 42:872–880.

White, A. E. 2016. Geographical barriers and dispersal propensity interact to limit range expansions of Himalayan birds. The American Naturalist 188:99–112.

Wiens, J. J., and M. J. Donoghue. 2004. Historical biogeography, ecology and species richness. Trends in Ecology & Evolution 19:639–644.

Wood, J. R., N. J. Rawlence, G. M. Rogers, J. J. Austin, T. H. Worthy, and A. Cooper. 2008. Coprolite deposits reveal the diet and ecology of the extinct New Zealand megaherbivore moa (Aves, Dinornithiformes). Quaternary Science Reviews 27:2593–2602.

Wünnemann, B., D. Demske, P. Tarasov, B. S. Kotlia, C. Reinhardt, J. Bloemendal, B. Diekmann, K. Hartmann, J. Krois, F. Riedel, and N. Arya. 2010. Hydrological evolution during the last 15kyr in the Tso Kar lake basin (Ladakh, India), derived from geomorphological, sedimentological and palynological records. Quaternary Science Reviews 29:1138–1155.

Zanne, A. E., D. C. Tank, W. K. Cornwell, J. M. Eastman, S. A. Smith, R. G. FitzJohn, D. J. McGlinn, et al. 2014. Three keys to the radiation of angiosperms into freezing environments. Nature 506:89–92.

Appendix

Figure S1 | Interactive network of species co-occurrences. Open the .html file in an internet browser. See Figure 14 and supplemental files.

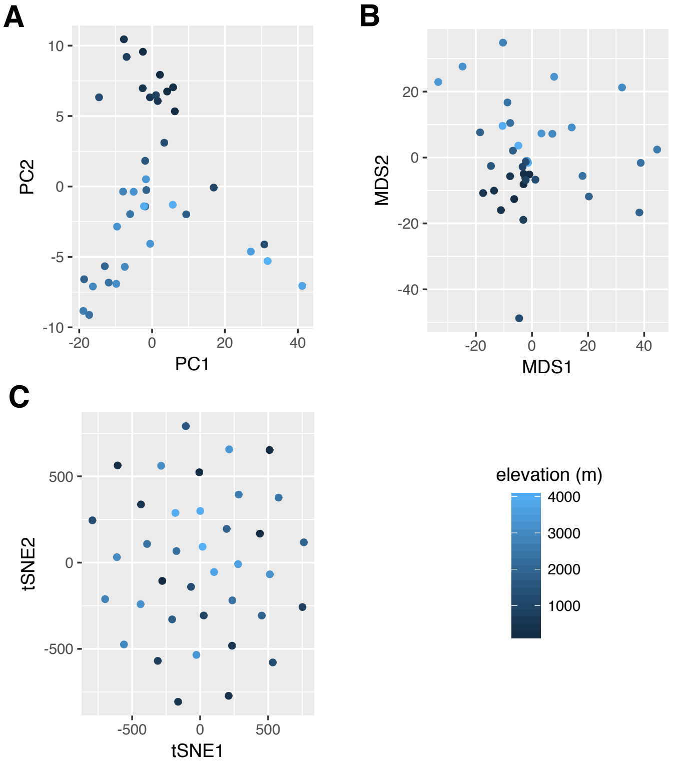


Figure S2 | Comparisons with other methods to project similarities of local Himalayan communities. Each panel represents different projections of the local Himalayan site by species census data – **A,** Principal components analysis, **B,** Non-metric multidimensional scaling, and **C,** t-SNE. Each community is coloured according to its elevation. Note that the clustering of sites and patterns of turnover that are revealed when fitting the Grade of Membership model are basically uninterpretable when using these approaches.

Table S1 – Site metadata for locations of Himalayan bird community surveys
Table S2 – Species contributions to Himalayan species abundance motifs
Table S3 – Species abundances in local Himalayan communities

Tables S1, S2, S3, and Fig. S1 are all available to download as supplemental files.