

## Research



**Cite this article:** van Holstein L, Foley RA.

2020 Terrestrial habitats decouple the relationship between species and subspecies diversification in mammals. *Proc. R. Soc. B* **287**: 20192702.

<http://dx.doi.org/10.1098/rspb.2019.2702>

Received: 19 November 2019

Accepted: 27 February 2020

**Subject Category:**

Evolution

**Subject Areas:**

taxonomy and systematics, evolution, ecology

**Keywords:**

speciation, subspecies, mammals, diversification, evolution

**Author for correspondence:**

Laura van Holstein

e-mail: [lav22@cam.ac.uk](mailto:lav22@cam.ac.uk)

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.4881018>.

# Terrestrial habitats decouple the relationship between species and subspecies diversification in mammals

Laura van Holstein and Robert A. Foley

Leverhulme Centre for Human Evolutionary Studies, Cambridge, UK

LvH, 0000-0003-1222-9593

Darwin proposed that lineages with higher diversification rates should evidence this capacity at both the species and subspecies level. This should be the case if subspecific boundaries are evolutionary faultlines along which speciation is generally more likely to occur. This pattern has been described for birds, but remains poorly understood in mammals. To investigate the relationship between species richness (SR) and subspecies richness (SSR), we calculated the strength of the correlation between the two across all mammals. Mammalian taxonomic richness correlates positively, but only very weakly, between the species and subspecies level, deviating from the pattern found in birds. However, when mammals are separated by environmental substrate, the relationship between generic SR and average SSR in non-terrestrial taxa is stronger than that reported for birds (Kendall's  $\tau = 0.31$ ,  $p < 0.001$ ). By contrast, the correlation in terrestrial taxa alone weakens compared to that for all mammals (Kendall's  $\tau = 0.11$ ,  $p < 0.001$ ). A significant interaction between environmental substrate and SR in phylogenetic regressions confirms a role for terrestrial habitats in disrupting otherwise linked dynamics of diversification across the taxonomic hierarchy. Further, models including species range size as a predictor show that range size affects SSR more in terrestrial taxa. Taken together, these results suggest that the dynamics of diversification of terrestrial mammals are more affected by physical barriers or ecological heterogeneity within ranges than those of non-terrestrial mammals, at two evolutionary levels. We discuss the implication of these results for the equivalence of avian and mammalian subspecies, their potential role in speciation and the broader question of the relationship between microevolution and macroevolution.

## 1. Introduction

Speciation across all forms of life is usually the process by which intraspecific population divergence culminates in sufficient discontinuity between populations to establish them on independent evolutionary trajectories. This process may be initiated by phenotypic differentiation, geographical isolation or both [1]. Given the link between intraspecific population divergence and speciation, it follows that the rates of population divergence and patterns of species richness (SR) should be linked and consequently, that taxonomic richness should be correlated between the species and subspecies levels. Darwin predicted such an association, hypothesizing that lineages comprising more species would 'oftener present varieties' [2, p. 55], or in other words, comprise more distinct populations, than less species-rich lineages. An association between the dynamics of divergence at and below the species level, however, is not a given: intraspecific populations may be too ephemeral for their origination to be linked to speciation [1], and other intraspecific processes such as population persistence or degree and nature of reproductive isolation between populations may affect the patterns of SR more [3]. Nonetheless, in a sample of 173 bird species, the rates of population divergence were shown to correlate positively with speciation rates, and there also seems to be a latitudinal effect

on the strength of this association with at- and below-species processes being more tightly linked in tropical lineages [1].

Estimating the rates of population divergence requires high-coverage genetic data and multiple samples per species, which are currently unavailable for most animal taxa. Insofar as subspecies represent spatially and phenotypically distinct populations within species [4], they can be used as a proxy for the product of intraspecific population divergence. Whether subspecies tend to represent ‘incipient’ species, and that average subspecies richness (hereafter ‘SSR’) and SR should thus be correlated, remains unresolved. Mayr [5], for example, conceived of the formation of subspecies as the second of five stages of speciation in birds, while Zink [6] made the contrasting case that avian subspecies nomenclature does not capture ‘real’ (that is, monophyletic) evolutionary units within species. In birds, however, SR and mean SSR do correlate positively with a Kendall’s *tau* of 0.23 [7].

While the relationship between SSR and higher taxonomic levels has received some attention among birds [1,6,7], much less is known about this in mammals. Here, we test Darwin’s prediction of a correlation between SR and SSR in mammals and, using avian studies as context, consider the effect of environmental substrate by comparing terrestrial and non-terrestrial mammals. We use multi-predictor phylogenetic regressions to test whether the relationship between SSR and SR differs significantly depending on environmental substrate, latitude or both. Finally, as it might be expected that range size is a key predictor of SSR, we examine the relationship between SSR and species range size.

## 2. Material and methods

### (a) Data collection

The number of subspecies per species for all mammals was extracted from Wilson and Reeder’s *Mammal species of the world* [8]. We separated mammals into two groups based on their environmental substrate; non-terrestrial mammals were the Orders: Chiroptera, Cetacea and Sirenia, and Families: Otariidae, Odobenidae and Phocidae. Species per genus and mean number of subspecies per species were calculated in R [9] using the *dplyr* package [10]. All further analyses were conducted in R. Species range data (total extent of species range (km<sup>2</sup>)) and median latitude was obtained from the PanTHERIA database [11]. Range data were not available for aquatic mammals. Tropical genera were defined as those with median latitude between 23.5° N and 23.5° S; temperate genera as those above 23.5° N and below 23.5° S.

### (b) Phylogenetic signal

We calculated Blomberg’s *K* and Pagel’s  $\lambda$  for generic average SSR using the ‘*phylosig*’ function in *phytools* [12] on 50 genus-level trees randomly sampled from Upham *et al.*’s [13] posterior sample. We generated these trees using the *VERTLIFE* subsetting tool [14] to produce trees with one species per genus, which avoids topological issues arising from any generic paraphyly. Supplemental analyses on the behaviour of Blomberg’s *K* and Pagel’s  $\lambda$ , given the right-skewed and long-tailed distribution of trait values, were carried out by simulating different data distributions and calculating Blomberg’s *K* and Pagel’s  $\lambda$  (see the electronic supplementary material, Section B).

### (c) Species and genus ‘ages’

As significantly different average species or genus ages between the two groups might affect the patterns of SSR, we compared the distributions of terrestrial versus non-terrestrial species and

genus branch lengths (as a proxy for taxon ‘age’) with Wilcoxon rank sum tests. We also calculated the Pearson correlation between species branch length and the number of subspecies for all mammals, and terrestrial and non-terrestrial groups separately, to see whether or not there was an effect of time on the accumulation of subspecies, and whether this differed in the two groups.

To obtain species ‘ages’, we extracted branch lengths from the maximum clade credibility (MCC) DNA-only node-dated tree from [13] using the ‘*edge.length*’ command in *phytools* [12]. We created an MCC genus-level tree from the species-level MCC tree using the ‘*drop.tip*’ command in *phytools* [12], leaving only one tip per genus, and extracted branch lengths in the same way as for the species-level tree.

### (d) Correlations between species richness and subspecies richness

We calculated Kendall’s *tau* between SR and mean SSR. Next, we ran phylogenetic regressions incorporating SSR, ecological substrate, latitude and species range on the 50 trees using the phylogenetic generalized least squares (PGLS) function in *caper* [15]. SR, SSR and species range (km<sup>2</sup>) were log transformed in all models. We optimized kappa and lambda branch length transformations (i.e. kappa = ‘ML’, lambda = ‘ML’) because Akaike information criterion scores for models with these transformations suggested these performed best out of all combinations of transformations (see the electronic supplementary material, table S1). Categorical predictors were coded as binary factors: for environmental substrate, these were ‘non-terrestrial’ and ‘terrestrial’; for latitude, these were ‘tropical’ and ‘temperate’. To test whether these categorical predictors interacted meaningfully with SSR to predict SR, we compared three linear models:

$$\text{simple: } y = \beta_0 + \beta_1 x + \varepsilon, \quad (2.1)$$

$$\text{variable intercept: } y = \beta_0 + \beta_1 x + \beta_2 x + \varepsilon \quad (2.2)$$

$$\text{and interaction: } y = \beta_0 + \beta_1 x + \beta_2 x + \beta_1 \beta_2 x + \varepsilon, \quad (2.3)$$

where *y* is SR, *x* is SSR and  $\beta_2$  is the categorical predictor of interest. The significance of the addition of the variable intercept (model 2), and the interaction term (model 3) was assessed by an ANOVA. Finally, we ran a model in which latitude and environmental substrate were allowed to interact, to ask whether the relationship between SR and SSR is mediated by latitude and substrate.

The same format was followed to explore the relationship between species’ range size and mean SSR, and whether this relationship is affected by the ecological substrate. The categories of environmental substrate were ‘terrestrial’ and ‘powered flight’ in this analysis, as no aquatic range data were available.

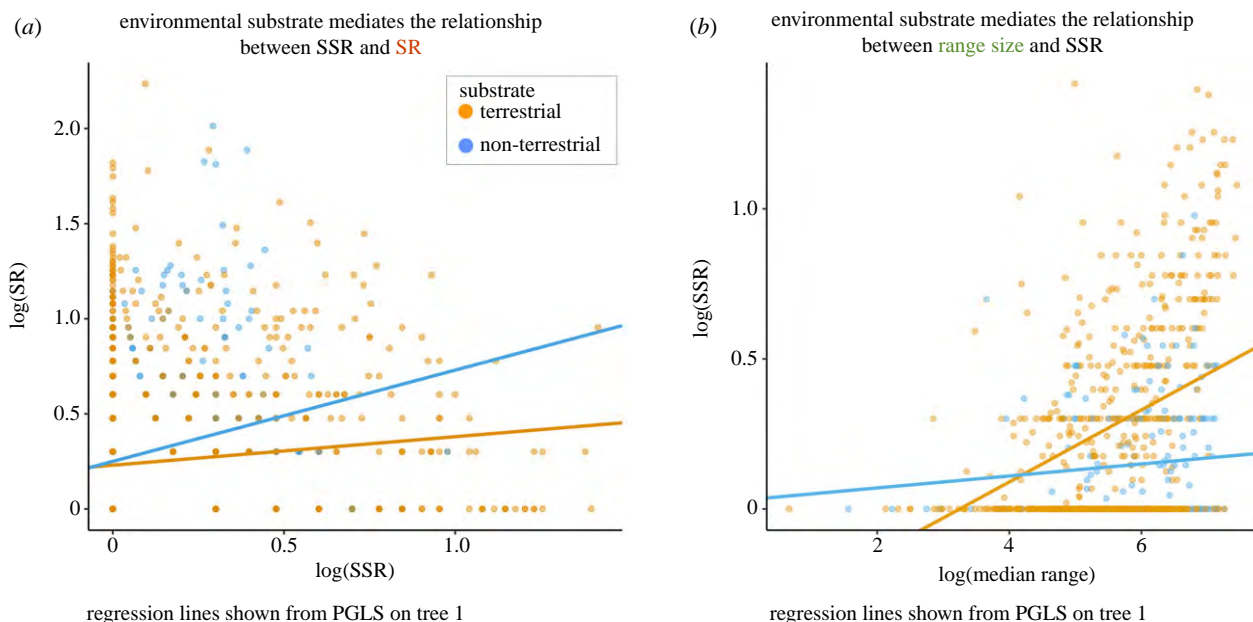
### (e) Testing for statistical artefacts

To confirm whether changes in statistical significance of correlations between SR and SSR are not simply a statistical artefact of subdividing a larger dataset, we took 10 000 random samples of 270 genera (because our non-terrestrial sample comprises 266 genera) from our data and calculated (i) Kendall’s *tau* between SR and average SSR and (ii) *p*-values for these correlation coefficients.

## 3. Results

### (a) Species richness and mean subspecies richness

Mammalian genera (*n* = 1251) comprise, on average, 4.4 species. Terrestrial mammalian genera (*n* = 985) tend to contain fewer species than non-terrestrial ones, but contain more subspecies per species (table 1).



**Figure 1.** Environmental substrate influences dynamics of diversification at two taxonomic levels. (a) shows the results from a phylogenetic regression on the first of our 50 trees between species richness (SR) and subspecies richness (SSR), where slope and intercept were allowed to vary depending on substrate. The interaction between SSR and substrate was found to be statistically significant, implying substrate mediates the relationship between average SSR and SR in mammals. (b) shows the results from a phylogenetic regression on the same tree between median species range and SSR, where slope and intercept were allowed to vary depending on substrate. Again, the interaction term was found to be significant, implying substrate mediates the relationship between range size and SSR within mammalian species. (Online version in colour.)

**Table 1.** Summary of results. (\*\*\*) $p < 0.001$ .

	mean species richness	mean subspecies richness	Kendall's $\tau$ correlation
all mammals	4.4	1.9	0.15***
terrestrial	4.3	1.9	0.11***
non-terrestrial	4.7	1.8	0.31***

### (b) Phylogenetic signal

Median Blomberg's  $K$  for the 50 trees was 0.09 (with  $p$ -values above and below 0.05), while that of Pagel's  $\lambda$  was 0.64 (all  $p < 0.05$ ) (see the electronic supplementary material, figure S1a). Given that Pagel's  $\lambda$  was significant in all cases we took into account phylogenetic structure in subsequent analyses by running phylogenetic regressions.

### (c) Clade 'ages'

Wilcoxon rank sum test with continuity correction indicated no difference between the distributions of genus branch length between terrestrial and non-terrestrial groups ( $p = 0.33$ ), and the same result was obtained for species ( $p = 0.38$ ). Further analyses were therefore performed without correcting for clade 'age'. Species branch lengths did not correlate with SSR in the whole mammal sample ( $R = -0.02$ ,  $p = 0.15$ ), nor when the sample was separated by substrate ( $R = -0.01$ ,  $p = 0.85$  for non-terrestrial taxa;  $R = -0.04$ ,  $p = 0.06$  for terrestrial taxa).

### (d) Effect of environmental substrate on the relationship between species richness and subspecies richness

SR and mean SSR are positively correlated in mammals as a whole, but there is major scatter around this trend (Kendall's  $\tau = 0.15$ , two-sided  $p < 0.001$ ). Terrestrial mammals alone exhibit a weaker, but still positive, correlation ( $\tau = 0.11$ , two-sided  $p < 0.001$ ). In non-terrestrial mammals, the relationship is stronger: more speciose genera tend to have species with a higher number of mean subspecies ( $\tau = 0.31$ , two-sided  $p < 0.001$ ). ANOVAs showed the addition of the interaction term was significant ( $p < 0.05$ ) in all 50 phylogenetic regressions (see the electronic supplementary material, figure S2b): that is, a model with an interaction term explained more of the variance in average SSR than the two other models. In addition, the interaction term was significant ( $p < 0.05$ ) in the interaction model itself in all 50 phylogenetic regressions. To illustrate the difference in slopes between terrestrial and non-terrestrial groups, the phylogenetic regression on the first of our 50 trees is plotted in figure 1a.

### (e) The effect of latitude

No significant correlation between SR and mean SSR was found in temperate mammals. In tropical mammals, there is a significant positive correlation (Kendall's  $\tau = 0.11$ , two-sided  $p < 0.05$ ). Phylogenetic regressions in which the intercept or intercept and slope were allowed to vary based on latitude were non-significant. Models in which the relationship between SR and SSR was mediated by latitude and substrate together was not significant overall, and no two-way interaction terms were significant either.



## (f) Species ranges and mean subspecies richness

In all mammals, the number of subspecies per species increases with species range (km<sup>2</sup>) ( $\tau = 0.32$ , two-sided  $p < 0.001$ ). When compared by ecological substrate, the correlation is very similar between terrestrial mammals and bats. In terrestrial mammals,  $\tau = 0.32$  (two-sided  $p < 0.001$ ), while that in bats is very slightly weaker ( $\tau = 0.30$ , two-sided  $p < 0.001$ ).

The interaction term between median range size and ecological substrate was significant ( $p < 0.05$ ) in all phylogenetic regressions (see the electronic supplementary material, figure S2a). ANOVAs showed the addition of the interaction term was significant ( $p < 0.05$ ). The regression on the first of our 50 trees is plotted in figure 1b.

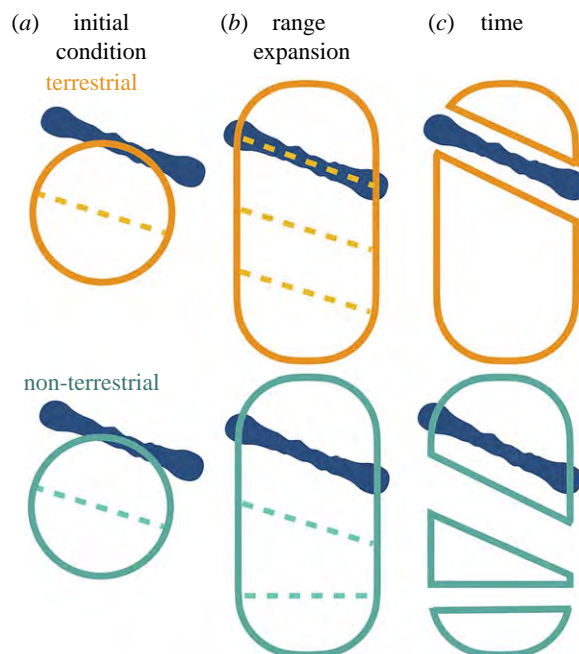
## (g) Testing for statistical artefacts

In our 10 000 randomized subsets, an increase in correlation coefficient to 0.31 (the observed Kendall's  $\tau$  for non-terrestrial taxa) did not occur, implying it is extremely unlikely that the increase we report occurred as a consequence of subsetting our dataset. All values of  $\tau$  above 0.09 yielded significant  $p$ -values ( $p < 0.05$ ). These results are shown in the electronic supplementary material, figure S3.

## 4. Discussion

We show that SR and SSR are positively related across all mammals, but that there is a significant difference in the strength of this relationship by environmental substrate, with a stronger correlation for non-terrestrial mammals. This result supports the hypothesis that if the permeable phenotypic or geographical boundaries between intraspecific populations are evolutionary faultlines along which speciation is generally more likely to occur, then intraspecific diversity and species diversity should be linked. Darwin's expectation that more speciose genera also comprise more subspecies on average is met, but there is considerable scatter around this trend.

The findings can be compared with those reported for birds; the relationship between SR and SSR in all mammals is much weaker at 0.15 than that reported for birds at 0.23 [7]. However, most birds are non-terrestrial, and when mammals are separated into terrestrial and non-terrestrial groups, non-terrestrial mammals have a substantially higher correlation coefficient (Kendall's  $\tau = 0.31$ ) than both terrestrial mammals (at 0.11) and birds. The difference in correlation strength between the two groups of mammals, and terrestrial mammals and birds, implies that the relationship between SR and SSR is mediated by terrestriality. To test whether an interaction between terrestriality and SSR is statistically significant, we compared three phylogenetic regressions (see Methods). The interaction term between the ecological substrate and SSR was significant ( $p < 0.05$ ) in all phylogenetic regressions, and ANOVAs confirmed the interaction model explained more of the variance in SR than the two other models. The interaction model is shown in figure 1a; the steeper slope of the non-terrestrial group compared to that of the terrestrial group illustrates that, for an equal increase in SSR, SR increases more in non-terrestrial habitats than terrestrial ones. It is possible that the conclusion that SR and SSR are more tightly linked in non-terrestrial taxa is specious if an increase in correlation strength from 0.15 to 0.31 is a likely consequence of subsetting a larger dataset, but such an increase did not occur in 10 000 random subsets of 270 genera (see the electronic supplementary material, figure S4). These results suggest that



**Figure 2.** Terrestrial habitats affect dynamics of diversification across two taxonomic levels more than non-terrestrial habitats. Both terrestrial and non-terrestrial species begin with two subspecies (a), the permeable boundary between which is indicated with a dotted line. Following range expansion (b) over the physical barrier (solid bar), the terrestrial species comprises four subspecies; subspecies formation is more tightly linked to local variation in landscape. By contrast, the non-terrestrial species now only comprises three subspecies, because it can maintain genetic unity over the physical barrier more easily. Given sufficient time (c), the barrier between the four subspecies in the terrestrial species impedes gene flow and the species gives rise to two new species. In the non-terrestrial species, the same physical barrier is not a barrier to gene flow, and the inherent boundaries between subspecies become the evolutionary faultlines along which new species are formed. (Online version in colour.)

ecological substrate mediates the relationship between SSR and SR in mammals—and more specifically, that terrestrial habitats might play a role in decoupling otherwise linked dynamics of diversity across the taxonomic hierarchy.

This is consistent with a scenario in which processes shaping terrestrial unit richness (that is, unit birth and death) are more influenced by physical barriers or ecological heterogeneity than those processes in non-terrestrial taxa. Alternatively, ecological features determine the dynamics of diversification in the same way in the two groups, and the pattern is the consequence of terrestrial habitats containing more physical barriers and ecological heterogeneity in the first place [16]. In both cases, two patterns should be evident: the relationship between SSR and range size should be stronger in terrestrial taxa than in non-terrestrial taxa, and SSR and SR should be more weakly correlated in terrestrial taxa than in non-terrestrial taxa. This model is illustrated in figure 2.

Indeed, the correlation between species range and SSR is somewhat stronger in terrestrial mammals (0.32) than those with powered flight (0.30). We compared three linear models, run on the 50 trees, in which average SSR was predicted by median species range size (see Methods). The interaction term was significant in the third model across all trees, and ANOVAs confirmed the inclusion of the interaction term explained significantly more of the variance in SSR. The prediction that species range should exert a stronger effect on SSR in terrestrial taxa is met: the differences in terrestrial versus

non-terrestrial slopes, illustrated in figure 2, indicate that SSR increases more with equal range expansions in terrestrial mammals than non-terrestrial mammals. These results are concordant with a model in which unit birth and death are more affected by ecological constraints in terrestrial habitats at two taxonomic levels. At smaller timescales, the formation of subspecies in terrestrial mammals is related more strongly to range size—in our model, this is explained by subspecies formation being more tightly linked to local variation in the landscape [4]. By contrast, either because of greater dispersal capacity or because they are not exposed to as many physical constraints within ranges in the first place [16], non-terrestrial taxa are able to maintain genetic unity over greater distances or (if present) over the same physical barrier; subspecies diversification is, consequently, less determined by physical constraints. Over greater timescales, a predominance of classic vicariant speciation [17], in which physical barriers impede gene flow and daughter species form along with these barriers, is consistent with a weaker relationship between SSR and SR in terrestrial taxa. In non-terrestrial species, equivalent physical barriers are less common or do not restrict gene flow to the same extent, and permeable boundaries between subspecies become the evolutionary faultlines along which new species are formed.

This model implies two key points, the first of which is a strong relationship between dispersal ability and diversification. Dispersal ability is a strong predictor of avian species diversification rates [18,19] and recently dispersed mammalian groups tend to contain more species than their sister clades in ancestral regions [20]. However, no work to date has directly compared terrestrial and non-terrestrial mammalian dispersal ability and its relationship to diversification: this remains to be explored.

A second implication is that subspecies might represent different evolutionary units in different mammalian taxa, and particularly depending on ecological substrate. In non-terrestrial taxa, as in birds, subspecies might be best conceived of as more often representing incipient species than in terrestrial taxa. Greater correlations between SR and SSR suggest speciation occurs along the phenotypic or geographical boundaries between subspecies more often in these clades. By contrast, terrestrial mammalian subspecies might more often be distinct but ephemeral populations and play a less pronounced role in speciation. It is important to emphasize that, even if ‘species’ in the biological species concept (BSC) sense and subspecies are stages in a continuum of genetic differentiation and the formation of subspecies is a stage through which species must pass (e.g. [6]), not every subspecies will inevitably become a species; subspecies can be re-absorbed into larger undifferentiated populations of individuals of the same species through persistent interbreeding.

It may be the case that correlated SR and SSR diversity is explained by heritable factors influencing their diversification [21], in which case phylogenetic signal in one or both should be high. Consistently significant and high values for Pagel’s  $\lambda$  in SSR, and the fact that SSR is not predicted by species’ branch lengths, suggest that this may be the case. If heritable diversification at both levels alone explains the correlation between SR and SSR, subspecies and species do not necessarily represent stages on an evolutionary continuum, as we suggest in our model. The heritability of diversification rate and subspecies diverging over time to become BSC species, however, are not mutually exclusive scenarios, and it seems reasonable to expect overlap between them. For example, subspeciation rate, or factors influencing subspeciation, might be heritable,

and lineages which inherited a high rate of subspeciation would, if subspecies represent incipient species, consequently comprise more species.

In terms of phylogenetic signal, the degree of asymmetry in the distributions of  $K$  and  $\lambda$  is unexpected, even if they measure different things ( $K$  being a measure of the partitioning of variance of the trait across clades and  $\lambda$  being a measure of covariance among species). Given the complex relationship between evolutionary process and phylogenetic signal [22], we are hesitant to use the asymmetry in these measures to infer information about the process of inheritance of subspeciation rate or factors influencing subspeciation in our sample. Instead, supplemental analyses (electronic supplementary material, Section B) show that the unusual distribution of trait values—right skewed with a heavy tail—can explain the consistent difference between  $K$  and  $\lambda$ .

Our analysis assumes a degree of equivalence between subspecies and ‘real’ evolutionary, genomically defined, units [6]. Mammalian subspecies, as a whole, have been subject to less intense scrutiny than birds. If mammalian subspecies do not consistently represent evolutionarily significant units, then relationships between intraspecific population processes and processes of diversification at the species level may be obscured, and stronger correlations between SR and the number of ‘real’ intraspecific populations than those we report here are to be expected.

A further major assumption in this paper is that taxonomists consistently define mammalian subspecies and species according to the same criteria across taxa and through time. Taxonomic practice is variable across taxa when it comes to species and subspecies ranks [23]. Biased departures from consistency confound our comparative analysis, because error resulting from species uncertainty will be non-randomly distributed through our dataset [23]. A shift from relatively universal use of the BSC towards the application of the phylogenetic species concept (PSC) in some groups further precludes accurate comparisons of SR and SSR between groups, because the PSC inflates subspecies under the BSC to full species status and consequently recognizes around 48% more species [23]. Taxonomic inflation should reduce SR–SSR correlations and could explain the reduction in correlation strength in terrestrial mammals if the PSC is more routinely applied in this group. We explored the degree to which these confounding variables would affect our results and conclusions and simulated the effect of different inflation regimes on correlation strength (see the electronic supplementary material). Reductions in correlation strength were only observed when taxonomic inflation was extreme—that is, all subspecies in most species in a genus were inflated to species status. In the contrasting regime, where only one subspecies was elevated to species status in a genus, correlation strength increased. The light inflation scenario is probably a more accurate approximation of the cumulative effect of no taxonomic inflation in some taxa and a moderate amount in others. If this is the case, the reduction in correlation strength in terrestrial mammals is unlikely to be the consequence of a greater degree of taxonomic inflation in this group than in the non-terrestrial group. Moreover, and in agreement with our simulation, correlation strength was reduced to  $\tau = 0.09$  (two-sided  $p < 0.001$ ) when we calculated the terrestrial Kendall correlation without Primates, the most heavily taxonomically inflated clade of the last two decades [23]. Finally, given that there

is potentially less consistent taxonomic practice or less frequent equivalence between phylogenetic structure and subspecies nomenclature in mammals than in birds, it is interesting that the correlation between SR and SSR is far stronger in non-terrestrial mammals than those in birds.

Ultimately, one of the biggest challenges in evolutionary biology is linking microevolutionary processes to macroevolutionary patterns. This paper sheds some new light on the factors mediating the relationship between population-level processes and speciation in mammals, and shows that the pathway from subspecies to species is environmentally contingent. In short, these results suggest that the dynamics of diversification of terrestrial mammals are more affected by physical barriers or ecological heterogeneity than those of non-terrestrial mammals, at two evolutionary levels. An implication of this model is that the evolutionary relationship between subspecies and species might, as a rule of thumb, differ between mammalian taxa, and be weakened in terrestrial habitats. This conclusion generates a number of testable hypotheses that should form the basis of future work. In particular, this study should encourage the exploration of the influence of substrate and other environmental parameters on speciation trajectories and probabilities, and consider these with more fine-grained taxonomic units and ecological categories.

## 5. Conclusion

Darwin proposed that lineages with higher diversification rates should evidence this capacity at both the species and

subspecies level, a view consistent with population-level processes being integral drivers of speciation. We show this hypothesis is weakly supported in mammals as a whole, but when taxa are separated by ecological substrate, non-terrestrial groups show much stronger correlations between taxonomic richness at the two levels than terrestrial mammals do. As these processes also appear more tightly coupled in birds, we suggested fundamental factors unique to terrestrial habitats, such as increased exposure to ecological or physical barriers, increasingly become the causal drivers of divergence at two levels. These results imply that the evolutionary relationship between subspecies and species might differ between mammalian taxa, and that this relationship is mediated significantly by ecological substrate.

**Data accessibility.** All data is submitted as an electronic supplementary Excel file; code is included in the electronic supplementary material.

**Authors' contributions.** L.v.H. conceived of the study, designed the analytical methods, collected the data and carried out the analysis; L.v.H. and R.A.F. interpreted the results and wrote the paper.

**Competing interests.** We declare we have no competing interests

**Funding.** This work was supported by a Cambridge Trust Vice-Chancellor's Scholarship and a St John's College Benefactor's Scholarship.

**Acknowledgements.** The manuscript was much improved by insightful and constructive feedback through the review process from Ally Phillimore, Loeske Kruuk and an anonymous reviewer, to whom we owe our sincerest thanks. We are also grateful to Jaap Saers and Keaghan Yaxley for their comments and support.

## References

- Harvey MG, Seeholzer GF, Smith BT, Rabosky DL, Cuervo AM, Brumfield RT. 2017 Positive association between population genetic differentiation and speciation rates in New World birds. *Proc. Natl Acad. Sci. USA* **114**, 6328–6333. (doi:10.1073/pnas.1617397114)
- Darwin C. 1959 *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life*, 502 p. London, UK: John Murray.
- Dynesius M, Jansson R. 2014 Persistence of within-species lineages: a neglected control of speciation rates. *Evolution* **68**, 923–934. (doi:10.1111/evo.12316)
- Patten MA. 2010 Chapter 3: Null expectations in subspecies diagnosis. *Ornithol. Monogr.* **67**, 35–41. (doi:10.1525/om.2010.67.1.35)
- Mayr E. 1940 Speciation phenomena in birds. *Am. Nat.* **74**, 249–278. (doi:10.1086/280892)
- Zink RM. 2004 The role of subspecies in obscuring avian biological diversity and misleading conservation policy. *Proc. R. Soc. Lond. B* **271**, 561–564. (doi:10.1098/rspb.2003.2617)
- Haskell DG, Adhikari A. 2009 Darwin's manufactory hypothesis is confirmed and predicts the extinction risk of extant birds. *PLoS ONE* **4**, e5460. (doi:10.1371/journal.pone.0005460)
- Wilson DE, Reeder DM (ed.) 2005 *Mammal species of the world. A taxonomic and geographic reference*, 2142 p, 3rd edn. Baltimore, MD: Johns Hopkins University Press.
- R Development Core Team. 2016 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Wickham H, François R, Henry L, Müller K. 2019 dplyr: a grammar of data manipulation. See <https://CRAN.R-project.org/package=dplyr>.
- Jones KE *et al.* 2009 PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology* **90**, 2648. (doi:10.1890/08-1494.1)
- Revell LJ. 2012 Phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223. (doi:10.1111/j.2041-210X.2011.00169.x)
- Upham NS, Esselstyn JA, Jetz W. 2019 Inferring the mammal tree: species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biol.* **17**, e3000494. (doi:10.1371/journal.pbio.3000494)
- Vertlife. 2020 See <https://vertlife.org/>.
- David O *et al.* 2018 caper: comparative analyses of phylogenetics and evolution in R. See <https://CRAN.R-project.org/package=caper>.
- Grosberg RK, Vermeij GJ, Wainwright PC. 2012 Biodiversity in water and on land. *Curr. Biol.* **22**, R900–R903. (doi:10.1016/j.cub.2012.09.050)
- Mayr E. 1982 Processes of animal speciation. In *Mechanisms of speciation* (ed. D Barigozzi), pp. 1–19. New York, NY: Alan R. Liss Inc.
- Phillimore AB, Freckleton RP, Orme CDL, Owens IPF. 2006 Ecology predicts large-scale patterns of phylogenetic diversification in birds. *Am. Nat.* **168**, 220–229. (doi:10.1086/505763)
- Claramunt S, Derryberry EP, Remsen JV, Brumfield RT. 2012 High dispersal ability inhibits speciation in a continental radiation of passerine birds. *Proc. R. Soc. B* **279**, 1567–1574. (doi:10.1098/rspb.2011.1922)
- Kisel Y, McInnes L, Toomey NH, Orme CDL. 2011 How diversification rates and diversity limits combine to create large-scale species-area relationships. *Phil. Trans. R. Soc. B* **366**, 2514–2525. (doi:10.1098/rstb.2011.0022)
- Phillimore AB. 2010 Subspecies origination and extinction in birds. In *Ornithological monographs* (eds K Winker, SM Haig), pp. 42–53. Washington, DC: The American Ornithologists Union.
- Revell LJ, Harmon LJ, Collar DC. 2008 Phylogenetic signal, evolutionary process, and rate. *Syst. Biol.* **57**, 591–601. (doi:10.1080/10635150802302427)
- Agapow PM, Bininda-Emonds ORP, Crandall KA, Gittleman JL, Mace GM, Marshall JC, Purvis A. 2004 The impact of species concept on biodiversity studies. *Q. Rev. Biol.* **79**, 161–179. (doi:10.1086/383542)