Strapline: Correspondence

Does daily climate variation have an effect on species' elevational range size?

Hong Qian[1,2,*†], Richard Field[3†], Jian Zhang[4], Yangjian Zhang[5]

[1]College of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou, Zhejiang 310036, China.

[2]Research and Collections Center, Illinois State Museum, 1011 East Ash Street, Springfield, IL 62703, USA.

[3]School of Geography, University of Nottingham, Nottingham, UK. ORCID ID: 0000-0003-2613-2688

[4]School of Ecological and Environmental Science, East China Normal University, Shanghai 200241, China.

[5]Key Laboratory of Ecosystem Network Observation and Modelling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China.

*To whom correspondence should be addressed. E-mail: hqian@museum.state.il.us

†These authors contributed equally to this work.

Keywords: Diurnal temperature range, Gilchrist's hypothesis, seasonal temperature range, species elevational range, vertebrates.

Running head: Climate variation and elevational range of species

Numbers: Number of words (abstract = 150; main body = 2172 excluding references and Figures), number of cited references = 19; number of figures = 3.

Abstract

In their recent paper published in *Science* (2016, **351**, 1437–1439), Chan *et al.* analyzed 137 montane gradients, concluding that they found a novel pattern—a negative relationship between mean elevational range size of species and daily temperature variation, which was claimed as empirical evidence for a novel macrophysiological principle (Gilchrist's hypothesis). This intriguing possibility was their key conceptual contribution. Unfortunately, as we show, the empirical evidence was flawed because of errors in the analyses and substantial sampling bias in the data. First, we re-ran their analyses using their data, finding that their model should have been rejected. Second, we performed two additional re-analyses of their data, addressing biases and pseudoreplication in different ways, both times again rejecting the evidence claimed to support Gilchrist's hypothesis. These results overturn the key empirical findings of Chan *et al.*'s study. Therefore, the 'macrophysiological principle' should be regarded as currently remaining unsupported by empirical evidence.

Species' distributional ranges determine broad-scale species richness patterns, and assessing the mechanisms driving species' distributional ranges is central to ecology. Because a disproportionately large amount of biodiversity occurs in mountainous regions (Heywood, 1995), understanding how species' elevational range sizes (i.e. the range of elevations occupied by each species) are driven by environmental factors can provide insights into the mechanisms driving global patterns of range size and species richness. One important body of theory (the 'climatic variability hypothesis') proposes that temperature variability through time drives elevational range sizes of species (Janzen, 1967; Stevens, 1992; McCain, 2009), with larger elevational range sizes resulting from greater variability. The reasoning is that species that can tolerate changes in temperature in one place can also tolerate equivalent changes in temperature associated with higher or lower elevation. This theory has been tested almost exclusively with respect to seasonal temperature variability. However, it has been suggested that shorter-term temperature variability may select for thermal specialists, and thus smaller elevational ranges (Gilchrist, 1995). Gilchrist (1995) explained this reversal, to negative elevational range size–temperature variability relationships at shorter temporal scales of temperature variability, by distinguishing between within-generation and between-generation temperature variation.

Chan *et al*. (2016) used a global-extent dataset (though lacking latitudes poleward of 40°S or N) of 137 montane gradients to relate mean elevational range size of species to their measures of seasonal temperature range and diurnal temperature range simultaneously. They claimed that they found a novel pattern in their study: diurnal temperature range negatively affects mean elevational range size (Fig. 1B). They considered this pattern their most important finding and interpreted it as supporting their extension of Gilchrist's (1995) model —that between-generation temperature variation favours thermal generalists but within-generation temperature variation favours thermal specialists. This conclusion is interesting and represents the key conceptual advance of their paper. Unfortunately, as we show, the empirical pattern on which it is based results from flaws in their analyses, and sampling bias. The ‘best model’ of Chan *et al.* (2016), on which their empirical conclusions were based, should have been rejected by any standard criteria, and by their own criteria. We now explain in more detail.

Chan *et al*. (2016) analyzed 137 montane gradients obtained from McCain (2009). In the dataset, the diurnal temperature range and mean elevational range size variables are not correlated with each other ($r = -0.039$, $P = 0.651$; Fig. 1A). Chan *et al*. constructed 29 path models, selecting as ‘best’ one that generates a weak ($R^2 = 0.06$; $P = 0.012$) direct effect of diurnal temperature range on mean elevational range size (Fig. 1C); note that the ‘$R = -0.25$’ they state on p.1437 is the standardized path coefficient within their structural equation model (SEM), which is a partial correlation coefficient, controlling for effects of both seasonal temperature range and precipitation on mean elevational range size. They based their conclusions on this ‘best model’, but when we used their data to rerun their model, we found several errors in their reported results, as follows.

Crucially, while the key result of Chan *et al.*’s analysis (a negative diurnal temperature range→mean elevational range size effect) was significant within their ‘best model’ (Fig. 1B), this model should have been rejected. Their stated procedure was to first reject any of their 29 SEMs that failed to meet all of the following criteria for model-fit statistics: root mean square error of approximation (RMSEA) < 0.08, comparative fit index (CFI) > 0.95 and standardized root mean square residual (SRMR) < 0.1. For models meeting these criteria they then selected the model with the lowest SRMR (even though SRMR does not penalize model complexity; Hooper *et al*., 2008). According to their Table S2, sixteen of their 29 SEMs meet their criteria, including their ‘best model’ (Fig. 1B; model 28 in their

Table S2). However in the case of the 'best model' the RMSEA value was incorrectly reported as 0.062 when actually RMSEA = 0.178 (Fig. 1B,C), which makes their 'best model' unacceptable by their criterion (note also that the 90% confidence interval for the RMSEA does not include 0.08). The actual value is also far in excess of other commonly used RMSEA thresholds for model acceptability (e.g. 0.10, 0.06, 0.05; Browne & Cudeck, 1993; Hu & Bentler, 1999; Shipley, 2000).

For their 'best model' only, Chan *et al*. (2016) also reported the result of a χ^2 test (testing discrepancy between the data and the model), a standard test of acceptability of an SEM. Models for which the data and the model are significantly different ($P < 0.05$) should be rejected before considering model-fit statistics such as RMSEA or SRMR (Shipley, 2000; Grace, 2006). Very importantly, Chan *et al*. (2016) reported $P = 0.157$ for their χ^2 test, but the correct *P*-value is 0.005 (Fig. 1C), indicating an unacceptable model. Thus, Chan *et al*.'s key conclusion (empirical evidence for a 'novel macrophysiological principle') was based on incorrectly reported results and misinterpreted significance; the model should have been rejected twice over.

We further note how weak the model is, even if we ignore the fact that it should have been rejected. It is normal in macroecology to report the R^2 (proportion of variation in the dependent variable(s) accounted for) when using the SEM approach (e.g. Hawkins & Porter, 2003; Hawkins *et al*., 2007; Jetz *et al*., 2009; Oberle *et al*., 2009; Spitale *et al*., 2009; Jonsson *et al*., 2011). Chan *et al*. (2016) did not do so, thus failing to report that only 11% of the variation in mean elevational range size was accounted for by their 'best model' (Fig. 1C). The direct effect of diurnal temperature range on mean elevational range size within their model (the key result of their study) accounted for only 6% of the variation. Further, this key direct effect was not significant ($P > 0.05$) for any of the taxa when analyzed separately (Fig.

S11 of Chan *et al.*, 2016). Thus, even taking the model at face value, the evidence for Chan *et al.*'s main conclusion is tenuous.

We are unable to meaningfully improve on the analysis of this dataset that was published by McCain (2009), so we do not attempt to provide a new ‘best model’. We do note, however, that of the remaining 15 SEMs reported by Chan *et al.* (2016; their Table S2) as meeting their criteria of RMSEA < 0.08, CFI > 0.95 and SRMR < 0.1, the model that their selection criteria would choose as ‘best’ is model 3 (SRMR = 0.0416). This SEM only includes latitude and precipitation, and therefore does not include diurnal temperature range. Thus their reported results and selection criteria suggest a model that rejects their own findings. However, we hesitate to conclude much here because we cannot replicate the results reported for model 3 in Chan et al. (2016), nor those for many of the other models reported in their Table S2.

Another key criticism of Chan *et al.*'s (2016) analysis is that it suffers from bias and pseudoreplication, with respect to taxon sampling and geographical distribution of samples. Unlike McCain (2009), they did not attempt to reduce these problems before analyzing the data. The first bias problem is that montane gradients in dry climates are substantially over-represented in the data set. Only ~30% of the world's land surface outside the Antarctic/polar deserts is under arid climates (Hess & McKnight, 2013), but 47% of the 137 montane gradients used in Chan *et al*. were classified as ‘dry’ mountains by McCain (2009), who used an unbiased criterion to assign each montane gradient into dry (humidity index < 0.50) or humid (humidity index > 0.50) class. Importantly, the dry mountains in the data are geographically biased (particularly in south-western USA and northern Africa, in latitudes higher than most other montane gradients used).

We re-ran Chan *et al*.'s model after attempting to address the over-representation of dry montane gradients in their data. Specifically, we first divided the 137 montane gradients into two subsets: ‘dry’ or ‘arid’ according to McCain (2009; N = 64), and the remaining samples (‘humid mountains’; N = 73). Next, we re-ran Chan *et al.*’s ‘best model’ on each subset, finding a diurnal temperature range effect on mean elevational range size only for dry mountains (Fig. 2A,B) and only a weak one (Fig. 2B). Then, we addressed the problem of over-representation of arid montane gradients in the data set by reducing the proportion of arid mountains from 47% to 30% (by randomly sampling 32 of the 64 arid mountain data-points), to match the proportion of arid land in the world. Following a suggestion by Chan *et al.* (responding to an early version of this paper), we performed this procedure 100 times, each time combining the 32 randomly sampled dry gradients with all 73 humid ones and running their ‘best model’ on each resulting dataset. The average value of the 100 standardized diurnal temperature range→mean elevational range size path coefficients was -0.209, and the mean of the 100 associated *P*-values was 0.121 (Fig. 2C), which is substantially larger than the standard significance threshold ($P < 0.05$, used by Chan *et al.*). Thus, removing just one of the biases in the data overturns the key empirical conclusion of Chan *et al.* (2016). Note that the reduction in sample size from 137 to 105 would play little part in the ‘loss’ of significance; 105 remains quite a large sample and is much bigger than the 64 dry samples in Fig. 2B.

Problems of bias and pseudoreplication in the analysis of Chan *et al.* (2016) go further. The 137 montane gradients in the full dataset were located in only 82 sites, some more localized than others, with the same site appearing up to 6 times for different taxa. These 82 sites were primarily in four clusters (western New World, Mediterranean region, south-eastern Africa plus Madagascar, south-eastern Asia; Fig. S1 of Chan *et al.*, 2016). Thus, the montane gradients in the dataset are strongly biased geographically. The dataset is also biased with respect to taxon sampling. On average, each site has only 1.7 gradients (137/82), each taxon only 20 gradients (range 12–33), and taxon samples are substantially biased geographically. For example, lizard gradients occurred only in two of the four sample clusters (none in the south-eastern Asian and south-eastern African clusters; Fig. S1 of Chan *et al.*, 2016). In the paper from which the data are taken (McCain, 2009), Christy McCain discussed the biases and because of them she did not perform an analysis of the ‘vertebrates’ group as a whole. In contrast, the analysis of Chan *et al.* (2016) was of this composite group. McCain (pers. comm.) strongly cautions against this, arguing in particular that including the rodent data in the analysis is inappropriate because rodents have the opposite elevational range size trend to the other vertebrate groups.

Following this advice, our final reanalysis started by excluding the rodent data and shows that removing a single, pseudoreplicated data-point again overturns the key empirical conclusion of Chan *et al*. (2016). Eyeballing Fig. 3 (equivalent to Fig. 1D of Chan *et al*., which itself suggests the same issue) suggests that the two data-points in the bottom-right corner of the scatterplot are highly influential in the analysis of the link between daily temperature variation and mean elevational range size. In fact, each point is pivotal: excluding either point from the analysis changes the key diurnal temperature range→mean elevational range size path in the SEM from significant to non-significant ($P = 0.024$ to $P = 0.087$ and 0.065, respectively). Examination of influential points should routinely be done, and here it is particularly apposite: these two pivotal points are pseudoreplicates. Both are from Martin (1961), both are reptile groups (one snakes and the other lizards) from the same study site (the Chiricahua Mountains in Arizona, USA), and both have exactly the same data for all the environmental variables. Note that many of the other data-points have similar pseudoreplication problems, including others that are influential in pulling the regression line in a negative direction (e.g. the two left-most points in Fig. 3, with the lowest diurnal temperature range, are both from a site in the Calabria region in Italy, and both are amphibian taxa [frogs and salamanders] taken from the same study).

Finally, we note in passing that, using their ‘best model’, Chan *et al*. (2016) reported a significant, negative effect of mean annual precipitation on seasonal temperature range (their Fig. 1A), which they claimed as a ‘novel pattern found in [their] study’ (their Fig. 1C, which states ‘$P < 0.05$’). However, our reanalysis showed that this path was non-significant (Fig. 1C), using the established biological alpha of 0.05, which was used in Chan *et al*. to determine statistical significance (as shown in their Fig. 1A).

In sum, re-analyzing McCain’s (2009) data provides no meaningful advance on the conclusions she originally published. Although Chan *et al*.’s theoretical extension of Gilchrist’s model is interesting, the patterns predicted by this ‘macrophysiological principle’ have not yet been empirically supported: there is currently no reliable evidence that daily climate variation affects species’ elevational range sizes.

ACKNOWLEDGEMENTS: We are grateful to Christy McCain and anonymous reviewers for helpful comments, and to Sheng-Feng Shen for allowing us to access data used in the original paper on which we have commented.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Code and results of SEM.

REFERENCES

Browne, M.W. & Cudeck, R. (1993) Alternative ways of assessing model fit. *Testing structural equation models* (ed. by K.A. Bollen and J.S. Long), pp.136–162. Sage, Newbury Park, CA.

Chan, W.-P., Chen, I.-C., Colwell, R.K., Liu, W.-C., Huang, C.-Y., & Shen, S.-F. (2016) Seasonal and daily climate variation have opposite effects on species elevational range size. *Science*, **351**, 1437–1439.

Gilchrist, G.W. (1995) Specialists and Generalists in Changing Environments. I. Fitness Landscapes of Thermal Sensitivity. *The American Naturalist*, **146**, 252–270.

Grace, J.B. (2006) *Structural equation modeling and natural systems*. Cambridge University Press, Cambridge, UK.

Hawkins, B.A. & Porter, E.E. (2003) Does herbivore diversity depend on plant diversity? the case of California butterflies. *The American Naturalist*, **161**, 40–49.

Hawkins, B.A., Diniz-Filho, J.A.F., Jaramillo, C.A., & Soeller, S.A. (2007) Climate, niche conservatism, and the global bird diversity gradient. *The American Naturalist*, **170**, S16–S27.

Hess, D. & McKnight, T.L. (2013) *McKnight's physical geography: a landscape appreciation*. 11th edn. Pearson Prentice Hall, Upper Saddle River, NJ.

Heywood, V.H. (ed.) (1995) *Global biodiversity assessment*. Cambridge University Press, Cambridge.

Hooper, D., Coughlan, J. & Mullen, M. (2008) *Structural equation modelling: guidelines for determining model fit*. Articles, 2.

Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, **6**, 1–55.

Janzen, D.H. (1967) Why mountain passes are higher in the tropics. *The American Naturalist*, **101**, 233–249.

Jetz, W., Kreft, H., Ceballos, G., & Mutke, J. (2009) Global associations between terrestrial producer and vertebrate consumer diversity. *Proceedings of the Royal Society of London B*, **276**, 269–278.

Jonsson, M., Englund, G., & Wardle, D.A. (2011) Direct and indirect effects of area, energy and habitat heterogeneity on breeding bird communities. *Journal of Biogeography*, **38**, 1186–1196.

Martin, P.S. (1961) Southwestern animal communities in the late Pleistocene. In: Bioecology of the arid and semiarid lands of the Southwest. (eds. Shields, LM & Gardner, LJ). New Mexico Highlands University Bulletin Las Vegas, pp. 56-66.

McCain, C.M. (2009) Vertebrate range sizes indicate that mountains may be 'higher' in the tropics. *Ecology Letters*, **12**, 550–560.

Oberle, B., Grace, J.B., & Chase, J.M. (2009) Beneath the veil: plant growth form influences the strength of species richness–productivity relationships in forests. *Global Ecology and Biogeography*, **18**, 416–425.

Shipley, B. (2000) *Cause and correlation in biology: a user's guide to path analysis, structural equations and causal inference*. Cambridge University Press, Cambridge, UK.

Spitale, D., Petraglia, A., & Tomaselli, M. (2009) Structural equation modelling detects unexpected differences between bryophyte and vascular plant richness along multiple environmental gradients. *Journal of Biogeography*, **36**, 745–755.

Stevens, G.C. (1992) The elevational gradient in altitudinal range, an extension of Rapoport's latitudinal rule to altitude. *The American Naturalist*, **140**, 893–911.

Editor: Richard Pearson

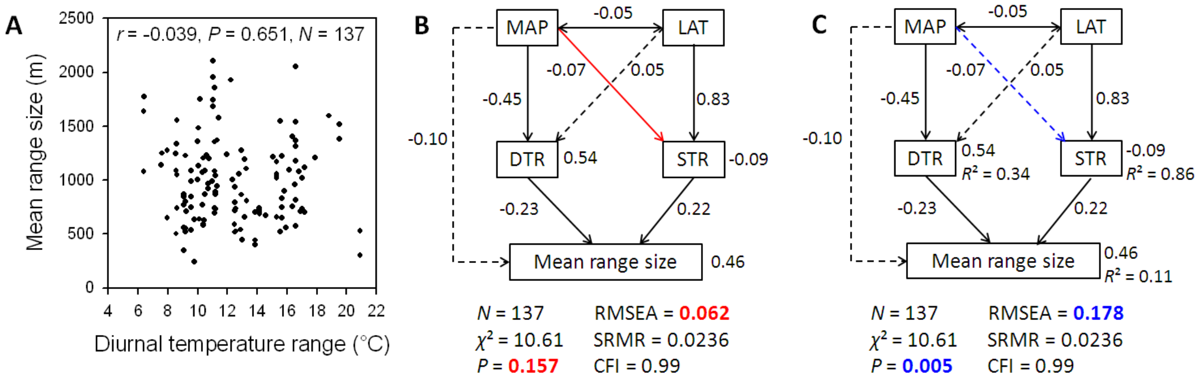

Figure 1 **(A)** Scatterplot of mean elevational range size against diurnal temperature range (DTR) for terrestrial vertebrate species in the 137 elevational gradients used in Chan *et al*. (2016) (Pearson's correlation results are shown). **(B)** The 'best' structural equation model (SEM) of Chan *et al*., showing relationships among mean annual precipitation (MAP), absolute latitude (LAT), diurnal temperature range (DTR), and seasonal temperature range (STR) in explaining variation in mean elevational range size. The statistics are as reported by them, and three key errors are highlighted in red. **(C)** Values were taken from (B) except for R^2 values and the corrected values (highlighted in blue) which resulted from when we ran the same SEM using the same data, model and software as in Chan *et al*. We also used LAVAAN package to repeat the analysis [see Appendix S1 in Supporting Information for code and results when repeating the analyses using the LAVAAN package (version 0.5-20) in R (cran.r-project.org/web/packages/lavaan)]. In (**B**) and (**C**), solid arrows are significant ($P < 0.05$) and dashed arrows are not significant. Numbers next to arrows and boxes are unstandardized slopes and intercepts, respectively, unless otherwise indicated. RMSEA is root mean square error of approximation; SRMR is standardized root mean square residual; CFI is comparative fit index.

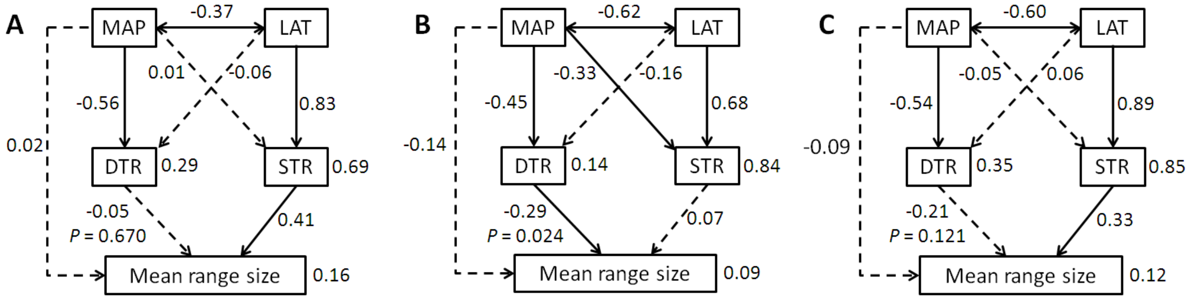

Figure 2 Structural equation models (SEM) showing relationships among mean annual precipitation (MAP), absolute latitude (LAT), diurnal temperature range (DTR) and seasonal temperature range (STR) in accounting for variation in mean elevational range size of terrestrial vertebrate species. Numbers next to arrows are standardized path coefficients; numbers next to boxes are R^2 values. Solid arrows are significant ($P < 0.05$); dashed arrows are not significant ($P > 0.05$). *P*-values are shown only for the focal path (see text). **(A)** SEM using the subset of 73 mountain gradients that were not indicated as ‘dry’ or ‘arid’ by McCain (2009). **(B)** SEM using the subset of 64 mountain gradients that were indicated as ‘dry’ or ‘arid’ by McCain (2009). **(C)** SEM using the 73 non-dry mountain gradients and 32 of the 64 dry mountain gradients, showing mean values for 100 randomizations.

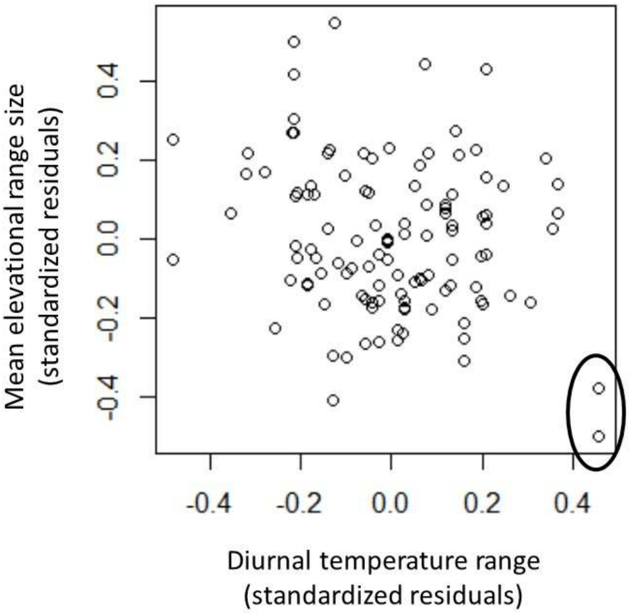

Figure 3 Partial residual plot of the modelled relationship between diurnal temperature range and mean elevational range size. This is the equivalent of Fig. 1D in Chan *et al.* (2016), but here using a dataset that excludes rodents (see text for explanation). The two influential points discussed in the text are indicated. Both represent reptile groups from the same study in the same study site, with identical values for all the environmental variables. Removing either makes the negative relationship non-significant.