doi:10.1038/nature25794

SUPPLEMENTARY INFORMATION FOR "GLOBAL ELEVATIONAL DIVERSITY AND DIVERSIFICATION OF BIRDS"

Elevation biases in richness estimates: Swiss bird survey

We found a strong negative association of elevation and probability of detection, with a slope estimate of -0.6149 (95% Credible Interval (CI) = [-1.013, -0.229]). While both raw richness estimates and those corrected for detectability decrease toward higher elevation, their decrease differs (Extended Data Fig. 1). At higher elevations, for the same effort, a larger proportion of species remains undetected (Extended Data Fig. 1), thus impacting the interpretation and comparability of elevation analyses based on survey data.

Sensitivity of results to resolution of species elevation limit data

Similar to the gridding of polygonal occurrence information at ca. 110km resolution as balance between data accuracy and maximum analysis detail, the use of elevational range data required a choice of analysis resolution. For the main analysis we used 500m, after a careful consideration of the potential benefits and downsides of a yet finer resolution which would offer greater detail but also potentially increase false presences. In a set of sensitivity analyses we assessed the robustness of our results to this choice and used 300m as alternative grain. Thus, we divided all of the mountain regions into trapezoids of 110km lateral and 300m elevational extent, resulting in a total of 14,218 assemblages. For each trapezoid, we estimated species richness and average tip DR and related them to elevation using the same model selection procedure as with the 500m data for each mountain region (see Methods). Applying the same models used in the main analyses to the 300m solution data leads to only minimal differences in the posterior distributions of model parameters: out of 131 parameters assessed in the comparison of 500m and 300m resolution elevation data, in only four cases of the tip DR – elevation relationship and no single case of the SR - elevation relationship did directionality significantly change (from positive to negative, or vice versa; Supplementary Table 5). We found similarly negligible differences for the multilevel models of SR and tip DR along elevation. All the parameter medians estimated using a 300m resolution are within the 95% CI of the parameter estimates using the 500m resolution (Extended Data Figure 7). Extended Data Figure 7 highlights the similarity of the patterns inferred across both elevational resolutions. These findings confirm the robustness of the results to the methodological choice of 500m as resolution for the species elevational range data.

Elevation detection bias and assemblage richness comparison/validation

We predicted that local surveys underrepresent the number of species due to sampling biases, non-detections, small spatial extents, etc. (Rowe and Lidgard 2009). In contrast, the assemblage richness estimates for the same locations derived through elevationally-refined range maps are expected to be larger, as they are not directly affected by limited detection and integrate over longer times scale. They also represent larger areas than those surveyed. But above and beyond these effects, we expect differences in sampling effort to be a key driver of differences in the two richness estimates.

Our results are congruent with these expectations. We find consistent but varying lower richness estimates coming from local survey compared to elevationally-refined range maps (Extended Data Fig. 4). Differences in effort, scale, focal species subset, among others, all lead to strong biases when capturing the diversity gradient along elevation. For instance, most field surveys spent less than a few hours of sampling at a given location. Previous comparisons of methods for combining elevational assemblage data show that results from single locality samples tend to dramatically underestimate richness (Rowe and Lidgard 2009), as our results show. In contrast, elevational transects with more extensive field efforts better approximate our species richness estimates (e.g., Manu National Park and Papua New Guinea).

The results from the hierarchical model strongly suggest that sampling effort is mostly driving the differences between the richness estimates for each transect (Extended Data Fig. 4b,c). The effect of the natural logarithm of sampling effort on the slopes was 0.18 (95% CI = [0.13,0.22]). Concordant with our expectations, surveys with lower sampling effort had slopes under 1, some even exhibiting negative relationships, while transects with greater effort approximated a slope of 1 (Extended Data Fig. 4). Other factors might also influence the discrepancy between richness estimates, such as scale, bird group, environment, among others. But here we show that differences in effort alone contribute to dramatic underestimations and biased inferences of diversity patterns, necessitating the use of comparable samples as provided by our approach for large-scale analyses.

Bird diversification rates

Phylogenetic information and tip DR

Tip DR allocates the time-calibrated branches of the phylogenetic tree among the tips according to shared ancestry. Each branch in the tree is divided and apportioned to the

number of descending species that derive from it, and this is done iteratively. Species that result from highly diversified clades come from branches that are shared to a greater extent and thus receive a smaller portion of all the available branch lengths (Jetz et al. 2012). Therefore, tip DR have the same units as other diversification rate metrics: the net amount of species produced, which is always one, over time, which is the apportioned branch length. We note that this measure reflects the more recent variation in diversification rates because deep branches are divided to a greater extent than terminal ones, and thus contribute exponentially less to each tip (Jetz et al. 2012).

We derive the tip DR values from the posterior distribution of Jetz et al (2012) which captures the uncertainty in both genetic data (66% species) and the soft taxonomic constraints applied to species lacking this information (Thomas et al. 2013). While it would be ideal to have a 100% genetic coverage, it is evident that including non-genetically represented species is necessary to avoid biases introduced from spatially and environmentally non-even genetic sampling when estimating diversification rates. While limitations exist for the interpretation or use of Maximum Clade Credibility trees developed with this approach, recent work has confirmed the suitability of such bias-corrected phylogenies for standard diversification analyses (Rabosky 2015). The Thomas et al., (2014) methodology captures the uncertainty about the current knowledge in the placement of species conjointly with that due to available molecular data, and in our study we allow this to inform the analyses by summarizing tip DR from 1,000 phylogenetic trees of the posterior distribution of trees.

Comparison of tip DR with clade level rates

Tip DR is unique in that it allows estimates of diversification rates for non-monophyletic assemblages without the need to prune absent species that disregards crucial information on the phylogenetic tree. We explored how tip DR compare to traditional metrics addressing whole clades for differently sized species groups when using a birth-death process. Specifically, we contrasted metric results for the following geographically nested species groups (assemblages): i) Globe, South America, Andes, and a randomly selected locality in the Andes (to represent alpha richness); ii) Globe, Australia, Great Dividing Range, and a randomly selected locality in the Great Dividing Range. We first estimated the diversification and speciation rates for each of the above clades (by pruning species absent for each of the focal species groups, as in Price et al. 2014) according to a birth-death process in a Bayesian framework using RevBayes v1.0.0-beta (Höhna et al. 2014). We used 100 randomly selected phylogenetic trees from the pseudo-posterior distribution from Jetz et al., (2012), and

estimated posterior distributions for speciation and diversification rates. We then generated pseudo-posterior distributions for these parameters by conditioning on each of the 100 posterior trees as completely independent samples from a posterior distribution. Although this procedure is marginally incorrect when accounting for the full phylogenetic uncertainty (as it treats each tree as a full observation), it is substantially better than conditioning on only one tree.

Subsequently, we estimated tip DR for each of the focal assemblages as the harmonic mean of all constituent species for every 100 pseudo-posterior trees. Additionally, for each of the focal assemblages we used different subsets of tips from the global tree, resulting in different phylogenetic trees. For instance, when estimating the diversification rates of an assemblage in the Andes, we first used the global tree to estimate tip DR; secondly, we used the tree with only the South American species, then the one containing only the Andean species, and, finally, the tree containing only the assemblage species. This was done in the same manner with all the focal species groups. Clearly, these estimates can only be compared with the birth-death process estimates when the constituent species of the phylogenetic tree match the ones present in the spatially delimited assemblage.

Given that the Maximum Likelihood Estimate for the speciation rate under a Yule model corresponds to the harmonic mean of the tip DR (Jetz et al. 2012), the results confirmed the expectation that the harmonic average of the tip DR should be larger than the diversification rate of a birth-death process, but smaller than its corresponding speciation rate (Extended Data Fig. 6). This is explained because of the difference between a Yule and Birth-Death process; the speciation rate in the Birth-Death process must be equal or higher than that of a pure-birth to accommodate the unobserved extinctions. Pruning the phylogenetic tree to only contain the species present within some spatial boundaries dramatically biases the estimates (usually yielding lower diversification estimates) disallowing any sensible comparison with estimates from other regions. The diversification rates estimates across different sites become incompatible with each other as the number of species present influences them. For example, South America has harboured several recent bird radiations, with evidence suggesting overall higher diversification rates (Ricklefs 2006). However, their higher diversification rates in comparison with the average global rate across all bird species can only be inferred when using the whole avian phylogenetic tree instead of pruning out species to obtain an artificial monophyletic clade (Extended Data Fig. 6). These results demonstrate the advantage of using the species-level metric to compare diversification rates

across space since they allow measuring rates for non-monophyletic assemblages while maintain comparable units of diversification.

Comparison of tip DR with BAMM tip rates

Finally, we also compared the species-level metric from the tip DR to the tip rates resulting from a BAMM analysis (Rabosky 2014; Rabosky et al. 2015). Because we wanted to make a species-level comparison for all species, we decided to run BAMM with all species (i.e., 9993 species) instead of only including genetically represented species trees and setting a sampling fraction <1. We emphasise that the goal of these analyses was to compare how the species-level diversification rate metrics fare against each other. Since a BAMM analysis can only be performed on a single phylogenetic tree at a time, we ran BAMM v2.5.0 for 50 randomly chosen bird trees from the pseudo-posterior distribution to better capture phylogenetic uncertainty. Each run consisted of 50 million generations of RJMCMC, logging every 1000th generation, where we set the *expectedNumberofShifts* prior to 100 –to achieve faster convergence– and followed the output of the 'BAMMtools' v2.1 package for R (Rabosky et al. 2014) to set the other priors. We monitored convergence by plotting the log-likelihood against the RJMCMC generations and discarded the first 20 million generations as 'burn-in' for convergence was achieved since in all 50 runs. The effective sizes of the log-likelihood and number of shifts where over 200 in all runs. The average, minimum and maximum effective sizes among runs for the log-likelihood were 603.03, 268.75 and 1016.74, respectively, while for the number of shifts were 875.05, 562.47 and 1182.17, respectively, as given by the 'coda' package v0.17-1 for R (Plummer et al. 2006).

To contrast tip DR with the BAMM tip-rate metric, we estimated the harmonic mean across the 50 runs of BAMM for each of the 9993 species, yielding an average tip rate for each of the species. Similarly, we calculated tip DR for each species based on the same 50 phylogenetic trees. Extended Data Figure 6a shows the relationship between BAMM speciation tip-rates and diversification tip-rates vs tip DR. As expected, tip DR are usually lower than speciation rates (intercept = 0.0202, slope = 0.944) but higher than diversification rates (intercept = 0.00506, slope = 1.14; Extended Data Fig. 5a). Although the slopes are close to a 1:1 relationship, there is considerable unexplained variance (r^2=0.665 for speciation rates and r^2=0.669 for diversification rates). This is not surprising given that BAMM assigns some optimal number and configuration of rate shifts along the tree and it is very unlikely that a rate shift is detected within smaller clades (i.e., lower taxonomic levels such as genus). Thus, for closely related species, if the model does not have enough power to detect a shift,

then all species end up with the same rate (even when marginalizing over the posterior run). This contrasts with tip DR since only sister species are guaranteed to have the same rate. Consequently, tip DR discerns rate differences among every species (except sister-species) and thus has higher variance for a given clade than BAMM (note the vertically aligned points in Extended Data Fig. 5a). To illustrate, while the tip DR of the genus *Acanthiza*, consisting of 13 species, range from 0.0786 to 0.153, with an average of 0.117, BAMM diversification rates for the same genus range from 0.0915 to 0.0922 with an average of 0.0919 (this is not due to the interplay of speciation and extinction: BAMM speciation rates ranged from 0.0956 to 0.0958).

We then compared how assemblage average tip DR compared with the BAMM rates. We estimated these rates for all of our 21,655 assemblages across the world by taking the harmonic mean among the species present. Extended Data Fig. 5b shows the relationship between assemblage averages of tip DR and those of BAMM tip speciation and diversification rates. tip DR were highly similar to BAMM stip peciation rates with an intercept of -0.0182, a slope of 1.08 and an r^2 of 0.767, while an intercept of -0.0315, a slope of 1.29 and an r^2 of 0.716 characterized their relationship with BAMM diversification rates. Finally, we analysed how tip DR and BAMM speciation and diversification rates vary along elevation across the globe using each mountain system as a mixed effect using INLA (using the framework in Eq. 4 in Methods). Extended Data Fig. 5c-f shows the results for the three separate multilevel regressions. Overall, the scatter of the points and the coefficients were only marginally different between using BAMM tip rates or tip DR.

Together, these results highlight the advantages of using a species level metric to analyse gradients of diversification and support the usage of tip DR as a valid metric in comparison with a more model-based approach such as BAMM.

Bias exploration when controlling for overrepresentation

We explored any systematic bias for a relationship between sSR and tip sDR using simulations. First, we characterized the frequency distribution of range sizes (i.e., the number of sampling units species breeding ranges intersect with) and diversification rates amongst all bird species. Then, we produced artificial species with randomly drawn range sizes (i.e., number of locations it occurs) and diversification rates from the above-parameterized distributions. We allocated species to the available locations of occurrence according to their range sizes and estimated for each location their species richness, diversification rates, sSR

and tip sDR. Finally, we regressed species richness against diversification rates and sSR versus tip sDR and registered their slopes and r^2's of the relationship. We used 10, 50, 100, 200, 500 and 1000 total number of available locations and 20, 50, 100, 200, 500 and 1000 total number of species for a total of 36 different simulated scenarios. Each scenario was simulated 1000 times. The results showed no systematic bias in our subsampling approach for a relationship between sSR and tip sDR.

Explaining the variation in elevational richness gradients

Hypotheses and Analysis

We first assessed the covariance structure among different predictors (Extended Data Fig. 10). Elevation range, total area, temperature seasonality and latitude, and elevational range and regional richness are all strongly positively associated with the area, temperature seasonality and latitude of mountain regions. The opposite is the case for average temperature. Given the universal species-area relationship, we expect higher species richness in mountains with greater areal extent (Storch et al. 2012). Similarly, a wider elevational range should allow for the packing of more species, everything else being equal. Area and elevational range could also explain differences in regional diversification rates (Kisel et al. 2011). Productivity can also influence both species richness and diversification rates, primarily by setting a 'carrying capacity' on the number of species through density dependent processes (Brown and Maurer 1987). Similarly, temperature might also influence diversification dynamics by, for instance, accelerating metabolic rates and thereby increasing genetic variability either through shorter generation times or cellular division or both (Gillooly and Allen 2007). Additionally, temperature is a measure of available energy, and thus might influence the number of species in a given region. Relatedly, species living under lower temperature seasonality are expected to have narrower elevational ranges and lower dispersal rates (Janzen 1967), thus potentially affecting species richness and diversification rate patterns. Given that local assemblage richness is a subset of the mountain region species pool, we assessed the effect of regional species richness (Ricklefs 1987). We, also explored if mountain age had an effect on richness following the time-for-speciation effect (Stephens and Wiens 2003).

Finally, we tested for the effect of past rates of climate change on diversity and diversification rates. Some regions were particularly impacted by periods of iterative cooling and warming during the last millions of years (Kennett 1995; Hewitt 2000; Wallis et al. 2016). For instance, during glaciations, newly formed ice sheets covered much of North

America and Europe (Wallis et al. 2016), which resulted in increased faunal turnover rates and migration (Opdyke 1995). In contrast, tropical mountain regions were particularly suited as refuges for new species to survive during these damaging climatic fluctuations (Hewitt 2000). Recently, differential turnover rates along latitude have also been found at the intraspecific level using a phylogeographic approach across the Americas (Smith et al. 2017).

Limitations of geological and climatic datasets

Information on mountain orogeny varies greatly in coverage and quality and its accuracy is often debated by the geological community. First, for a given mountain system, there are usually several different uplift pulses of varying intensity that are non-regular along time. Second, given that our delimitation of mountain systems was biologically driven rather than determined by "geological independence", some systems have spatial variation in uplift episodes. Third, dating of the uplift pulses is often controversial, and a single mountain system may see several hypotheses put forward, sometimes differing by tens of millions years. Finally, given the geographic heterogeneity of published research in this area and the global nature of the study, the available information for some of the mountain systems is very scant. While we attempted to bring together the best information currently available, it is still obvious from Supplementary Table S3 that great variability in the estimated dates of the orogeny for the mountain systems exist. For each mountain system, we used the oldest pulse of uplift at which significant height was attained. This decision follows the time-for-speciation hypothesis, in which the newly colonisable area becomes available. Similarly, we expect lower elevations to be older (in terms of time available for species to colonise) than higher elevations resulting for mountain-building. Likewise, colonisation can only occur when the lineage is present; thus, we decided to use a prominently proposed date of the emergence of modern birds, ca. 72.9 Myrs (Prum et al. 2015) as maximum, when the uplift date of the mountain system was older.

The climatic dataset used for the present, CRU CL 2.0, consists of an interpolation of available weather stations at a 10' longitude/latitude resolution (New et al. 2002). Spatial climatic interpolation is necessary given the spatial unevenness and incompleteness of weather stations across the globe. Consequently, small-scale spatial variation will not be captured evenly across the globe; mountain regions are particularly affected in areas where weather stations are sparse (New et al. 2002; Soria-Auza et al. 2010). Nevertheless, we expect these issues to only have minimal effects on the estimates and results presented here, as the

relatively large spatial extent of the mountain system (>100km) for which we aggregated climatic data (Daly 2006). Future work addressing climatic factors at the assemblage rather than aggregate mountain system level will benefit from additional, e.g. remote sensing supported, climate characterizations and more field station data.

We also recognize several limitations in the reconstructions of past climatic information based on Global Climate Models (GCMs) from CIMP5 simulations, with the specific modelling decisions from the different institutions yielding ample variability (Taylor et al. 2011). In order to reduce systematic biases in these reconstructions and derive a consensus estimate, we averaged over the different and, to some extent, independent models available, produced by independent institutions.

Interpreting group-level predictors in quadratic regression

We assessed the effect of mountain system variables on the coefficients of the quadratic regressions for elevation and species richness, elevation and diversification rates, and diversification rates and species richness using multilevel models (See Eq. 6 in Methods). Here, we allow the intercept, the linear coefficient, and the quadratic coefficient for each mountain system to be influenced by a mountain system variable, such as mountain-wide average temperature. Let $\boldsymbol{\gamma} = \{\gamma_0, \gamma_1, \gamma_2\}$ be the linear effect of mountain-level variable x_m on the intercept, the linear and the quadratic coefficient of the quadratic regression. We note that for ease of interpretation and optimization and computational purposes, we always standardized the independent variable (i.e., the 'x' axis has mean of 0 and standard deviation of 1). Then, the intercepts of the quadratic regressions represent the predicted dependent variable at 0 (i.e., the height of the curve). Since we standardized the 'x' axis, '0' corresponds to the average of either elevation or diversification rates among our sampling points. Thus, a positive effect of x_m on the intercept ($\gamma_0 > 0$) would indicate that, at 0, mountain systems with larger values of x_m will have higher predicted values in the y axis.

The linear coefficient of the quadratic regression corresponds to the instantaneous slope of the curve at 0 (i.e., the derivative at 0). Then, a positive effect of x_m on the linear coefficient ($\gamma_1 > 0$) would indicate that, at 0, the larger rates of change are associated with larger values of x_m. Finally, the quadratic coefficient represents the degree of curvature of the regression. Thus, if $\gamma_2 > 0$, then higher values of x_m are associated with more closed curves.

Past rates of temperature change, annual temperature seasonality, average annual temperature, and latitude.

We examined and predicted specific directional relationships for 'Past rates of temperature change' on the basis of mechanisms and *a priori* hypotheses. This predictor emerged as the mountain-level covariate that most strongly matched our predictions, and we therefore place emphasis on it accordingly (Fig. 3). Nevertheless, we acknowledge that significant collinearity exists between past rates of temperature change and temperature seasonality, average temperature and latitude (Extended Data Figure 10). This is evinced in the similarity of the mountain-level covariable effects on SR and sSR and tip DR and tip sDR along elevation and sSR vs tip sDR (Extended Data Figure 8, 9). For the relationship of sSR with elevation and with tip sDR, past rates of temperature change are not the strongest predictor, but is very similar qualitatively and quantitatively to these other collinear variables. Still, past rates of climate change have the strongest effect on the overall mountain range tip sDR, and on how tip sDR varies along elevation (Fig. 3 & 4, Extended Data Figure 9). Collinearities represent a challenge in many none-experimental hypothesis testing setting, and we follow common practice and guidance (Dormann et al. 2013) by careful preselecting and emphasis on mechanism and hypothesis relevance in the interpretation of results.

Simulations addressing potential impact of phylogenetic non-independence of assemblages

When assessing predictors of species-level traits, including diversification rate metrics such as tip DR, effective sample size or even parameter estimates may be affected by phylogenetic non-independence (Felsenstein 1985; Freckleton et al. 2002). When analysis units correspond to lineages that form tips in a phylogenetic tree, common approaches, such as Phylogenetic Generalized Least Squares (PGLS), can be employed to account for phylogenetic relatedness by estimating an appropriate covariance matrix among the tips (Martins and Hansen 1997; Mazel et al. 2016). For instance, PGLS has been used to test for associations between population differentiation and tip metrics of diversification/speciation rates (including tip DR; Harvey et al. 2017).

Conceivably, a bias and/or error from phylogenetic non-independence may extend to relationships of metrics aggregated at the assemblage level, such as the association between average assemblage tip DR and elevation analyzed in our study. To our knowledge, this issue has not yet been examined or dealt with in the literature when assessing geographic difference in diversification rates (Kennedy et al. 2014; Belmaker and Jetz 2015), and no methodology is

available to formally correct for shared evolutionary history in this type of assemblage analyses. Current comparative methods are not designed to deal with the complexities of assemblage-level data, where each data point corresponds to an aggregate metric (i.e., sum, median, average, etc.) of a group of lineages that almost never would constitute a clade, i.e. be monophyletic.

In order to nevertheless explore this issue in the context of our analyses and assess the potential sensitivity of our results we conducted simulations. Specifically, we studied the possible bias and type I error that arises for assemblage-level relationships when using tip DR and an evolved trait on a phylogenetic tree. Specifically, we first simulated phylogenetic trees using a birth-death process, with speciation rate of 1.0 and extinction rate of 0.2, until arrival at 1000 species (if the process went extinct before arrival at 1000 species, we started anew) using the 'phytools' R package (Revell 2013; R Core Team 2015) and estimated tip DR. We then simulated tip trait data on these trees following three different models of trait evolution: white noise (WN; no phylogenetic signal), Brownian motion (BM) and Ornstein-Uhlenbeck (OU). To resemble the elevational range variable, we considered the evolved trait values at the tips to represent the species' elevational midpoint and then randomly sampled a symmetric trait range to characterize its elevational extent (e.g., if a species had an evolved mean trait of 0.5 and we sample a range of 0.2, we calculate the trait range as 0.4-0.6). All simulations were performed using R (R Core Team 2015).

We then randomly sampled 100 normally distributed points to mimic assemblages. Firstly, we considered as an assemblage all species whose trait range overlapped a given sampling point. Nonetheless, it would be unrealistic to assume that all species that share an elevation range are 'sympatric', so we first sampled the assemblage size randomly and uniformly between 1 and the total number of species overlapping a given sampling point. Then we selected a uniformly drawn random subset of species from this sampled assemblage of overlapping species.

We performed the same subsampled approach applied to the empirical data and described in the Methods (to control for the overrepresentation of species that occur in multiple assemblages). As in our main analyses, during each subsampling iteration, we aggregated estimated assemblage-level tip sDR using harmonic averages. And, as with the empirical data, we performed a random-effect linear regression using tip sDR as response and the assemblage value as the predictor, with assemblage as a random effect, in a Bayesian setting using INLA (Rue et al. 2009). We used 1000 iterations for each of the trait evolution

models.

We found that, following this simulation, there is no bias in the parameter estimates of the tip sDR – assemblage values relationship (i.e., the expected value was 0). This confirms that the trends fitted in Figure 2 should be immune to effects of phylogenetic non-independence pertaining to the assemblage level. However, we did detect a somewhat inflated type I error: the 95% Credible Interval (CI) did not overlap '0' to in 13.5-22.5% of the simulations for each of the trait evolution scenarios (WN = 13.5%, BM = 22.1%, OU = 22.5%). Raising the CI to 99.5% reduces the type I error (WN = 3.1%, BM = 8.7% & OU = 6.5%) and raising the CI to 99.8% brings it back in line with that expected for a 95% CI level (WN = 1.6%, BM = 5.9% & OU = 4.9%). Overall, this points to an inflated type I error when using tip DR as response for a single mountain system, where a 99.8% CI should be used to represent the commonly used "significance level" of 95%.

In our study, we conservatively did not assess significances for relationships in single mountain systems. Instead our main analyses evaluate relationships with multiple mountain systems as (phylogenetically quite separate) replicates, which we expect to decrease type I errors as observed for single mountain systems. In order to formally assess this, we conducted further simulations. Specifically, we repeated similar analyses as described above, but using instead a phylogenetic tree of 10000 species and 30 different groups of 50 assemblages, comparable to having 30 different mountain systems, each with 50 assemblages. We assigned species into the different groups such that they shared about 10% of the species among them (similar to the average percentage of species shared between our mountain systems). We then performed a three-level hierarchical regression using assemblage and group as random effects. Simultaneously, we explored the type I error on the effect of a group-level variable in explaining the variance among groups (in the intercept and in the slope) by randomly sampling 30 values from a normally distributed variable and performing a hierarchical regression with high-level predictors. All analyses were conducted in a Bayesian framework using INLA. Given the computational burden, we used 100 iterations per trait evolution model.

We find that using separate groups as replicates in a three-level hierarchical model, analogous to our global results, results in levels of type I error that agree with those expected from using a 95% CI. This confirms the power of using separate mountain systems as biological replicates. The percentage of false positives in the global estimates using a 95% CI ranged between 1%-8% (WN = 1%, BM = 8%, OU = 4%). Similarly, we find that the type I error is low for the effect of a high-level predictor on the intercept (i.e., γ_0 in Equation 6, at

the 95% CI: WN = 8%, BM = 7%, OU = 9% and 0% for all three models at the 99% CI) and on the slope (i.e., γ_1 in Equation 6, WN = 5%, BM = 6%, OU = 5% % and 0% for all three models at the 99% CI).

In summary, while we find an inflated false positive rate in individual mountain systems, the type I error in our global results does not show such an inflation. This applies to both the global relationship and high-level predictors (although it is possibly marginally inflated for γ_0), and to the Credible Intervals of parameter estimates which accurately represent their uncertainty. Accordingly, we find that our results hold when assessing the effect of mountain-level predictors as shown in Figure 3 and Extended Data Figs. 8 & 9. Finally, we note that the marginal distribution of the parameters describing the effect of 'Past rates of climate change' in explaining across mountain variation (Figure 4) do not include '0' at levels higher than 99.999% CI (99.999% CI = 0.00896-0.0449) for the intercept (i.e., γ_0), and at levels higher than 99% CI level (99% CI = 0.00084 – 0.0302) for the slope (i.e., γ_1).

Supplementary References

Belmaker J., Jetz W. 2015. Relative roles of ecological and energetic constraints, diversification rates and region history on global species richness gradients. Ecol. Lett. 18:563–571.

Brown J.H., Maurer B.A. 1987. Evolution of species assemblages: effects of energetic constraints and species dynamics on the diversification of the North American avifauna. Am. Nat. 130:1–17.

Daly C. 2006. Guidelines for assessing the suitability of spatial climate data sets. Int. J. Climatol. 26:707–721.

Dormann C.F., Elith J., Bacher S., Buchmann C., Carl G., Carré G., Marquéz J.R.G., Gruber B., Lafourcade B., Leitão P.J., Münkemüller T., McClean C., Osborne P.E., Reineking B., Schröder B., Skidmore A.K., Zurell D., Lautenbach S. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography. 36:27–46.

Felsenstein J. 1985. Phylogenies and the Comparative Method. Am. Nat. 125:1–15.

Freckleton R.P., Harvey P.H., Pagel M. 2002. Phylogenetic analysis and comparative data: A test and review of evidence. Am. Nat. 160:712–726.

Gillooly J.F., Allen A.P. 2007. Linking Global Patterns in Biodiversity to Evolutionary Dynamics Using Metabolic Theory. Ecology. 88:1890–1894.

Harvey M.G., Seeholzer G.F., Smith B.T., Rabosky D.L., Cuervo A.M., Brumfield R.T. 2017.

Positive association between population genetic differentiation and speciation rates in New World birds. Proc. Natl. Acad. Sci. 114:6328–6333.

Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. Nature. 405:907–913.

Höhna S., Heath T.A., Boussau B., Landis M.J., Ronquist F., Huelsenbeck J.P. 2014. Probabilistic Graphical Model Representation in Phylogenetics. Syst. Biol. 63:753–771.

Janzen D.H. 1967. Why mountain passes are higher in the tropics. Am. Nat. 101:233–249.

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

Kennedy J.D., Wang Z., Weir J.T., Rahbek C., Fjeldså J., Price T.D. 2014. Into and out of the tropics: the generation of the latitudinal gradient among New World passerine birds. J. Biogeogr. 41:1746–1757.

Kennett J.P. 1995. A review of polar climatic evolution during the neogene, based on the marine sediment record. In: Vrba E.S., Denton G.H., Partridge T.C., Burckle, editors. Paleoclimate and evolution with emphasis in human origins. Chelsea, Michigan, USA.: Yale University Press.

Kisel Y., McInnes L., Toomey N.H., Orme C.D.L. 2011. How diversification rates and diversity limits combine to create large-scale species–area relationships. Philos. Trans. R. Soc. Lond. B Biol. Sci. 366:2514–2525.

Martins E.P., Hansen T.F. 1997. Phylogenies and the Comparative Method: A General Approach to Incorporating Phylogenetic Information into the Analysis of Interspecific Data. Am. Nat. 149:646–667.

Mazel F., Davies T.J., Georges D., Lavergne S., Thuiller W., Peres-Neto P.R. 2016. Improving phylogenetic regression under complex evolutionary models. Ecology. 97:286–293.

New M., Lister D., Hulme M., Makin I. 2002. A high-resolution data set of surface climate over global land areas. Clim. Res. 21:1–25.

Opdyke N.D. 1995. Mammalian migration and climate over the last seven million years. In: Vrba E.S., Denton G.H., Partridge T.C., Burckle, editors. Paleoclimate and evolution with emphasis in human origins. Chelsea, Michigan, USA.: Yale University Press.

Plummer M., Best N., Cowles K., Vines K. 2006. CODA: convergence diagnosis and output analysis for MCMC. R News. 6:7–11.

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

225.

Prum R.O., Berv J.S., Dornburg A., Field D.J., Townsend J.P., Lemmon E.M., Lemmon A.R. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature. 526:569–573.

R Core Team. 2015. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rabosky D.L. 2014. Automatic Detection of Key Innovations, Rate Shifts, and Diversity-Dependence on Phylogenetic Trees. PLoS ONE. 9:e89543.

Rabosky D.L. 2015. No substitute for real data: A cautionary note on the use of phylogenies from birth–death polytomy resolvers for downstream comparative analyses. Evolution. 69:3207–3216.

Rabosky D.L., Grundler M., Anderson C., Title P., Shi J.J., Brown J.W., Huang H., Larson J.G. 2014. BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. Methods Ecol. Evol. 5:701–707.

Rabosky D.L., Title P.O., Huang H. 2015. Minimal effects of latitude on present-day speciation rates in New World birds. Proc R Soc B. 282:20142889.

Revell L.J. 2013. phytools: Phylogenetic Tools for comparative biology (and other things). .

Ricklefs R.E. 1987. Community Diversity: Relative Roles of Local and Regional Processes. Science. 235:167–171.

Ricklefs R.E. 2006. Global variation in the diversification rate of passerine birds. Ecology. 87:2468–2478.

Rowe R.J., Lidgard S. 2009. Elevational gradients and species richness: do methods change pattern perception? Glob. Ecol. Biogeogr. 18:163–177.

Rue H., Martino S., Chopin N. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. Ser. B Stat. Methodol. 71:319–392.

Smith B.T., Seeholzer G.F., Harvey M.G., Cuervo A.M., Brumfield R.T. 2017. A latitudinal phylogeographic diversity gradient in birds. PLOS Biol. 15:e2001073.

Soria-Auza R.W., Kessler M., Bach K., Barajas-Barbosa P.M., Lehnert M., Herzog S.K., Böhner J. 2010. Impact of the quality of climate models for modelling species occurrences in countries with poor climatic documentation: a case study from Bolivia. Ecol. Model. 221:1221–1229.

Stephens P.R., Wiens J.J. 2003. Explaining species richness from continents to communities: the time-for-speciation effect in Emydid turtles. Am. Nat. 161:112–128.

Storch D., Keil P., Jetz W. 2012. Universal species-area and endemics-area relationships at continental scales. Nature. 488:78–81.

Taylor K.E., Stouffer R.J., Meehl G.A. 2011. An Overview of CMIP5 and the Experiment Design. Bull. Am. Meteorol. Soc. 93:485–498.

Thomas G.H., Hartmann K., Jetz W., Joy J.B., Mimoto A., Mooers A.O. 2013. PASTIS: an R package to facilitate phylogenetic assembly with soft taxonomic inferences. Methods Ecol. Evol. 4:1011–1017.

Wallis G.P., Waters J.M., Upton P., Craw D. 2016. Transverse alpine speciation driven by glaciation. Trends Ecol. Evol. 31:916–926.

SUPPLEMENTARY FIGURES

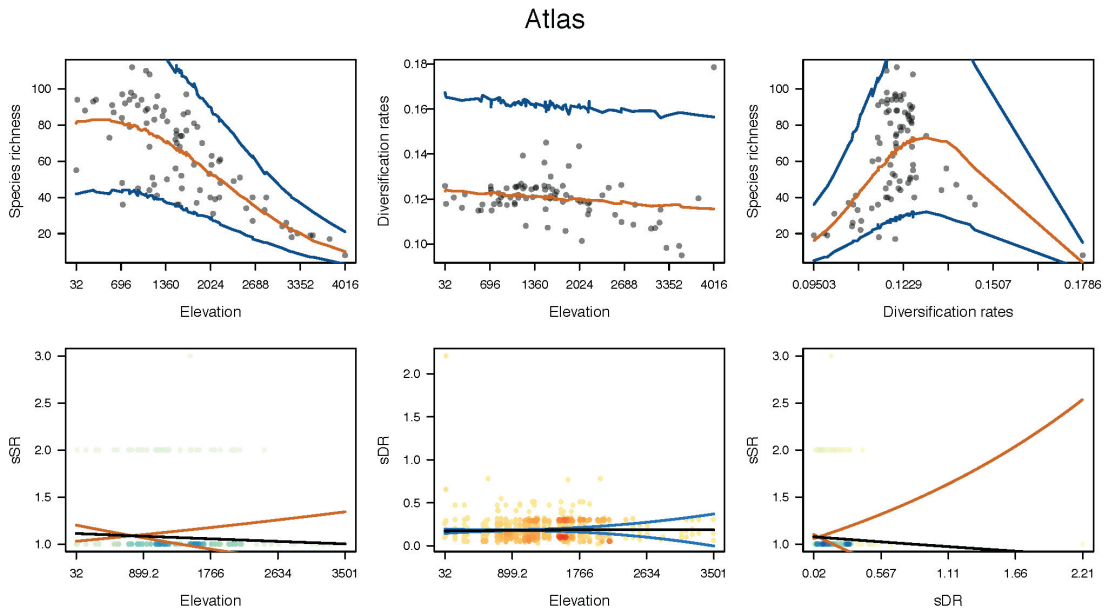

Figure S 1. Species richness and diversification patterns for the Atlas Mountains. Top left: species richness (SR) along elevation. Top middle: tip DR along elevation. Top right: effect of tip DR on SR. Bottom left: subsampled species richness (sSR) along elevation. Bottom middle: subsampled diversification rates (tip sDR) along elevation. Bottom right: effect of tip sDR on sSR. For the top row, orange lines correspond to the posterior expectation while purple lines correspond to the posterior 95% predictive quantiles (n = 86 assemblages). For the bottom row, black lines correspond to the posterior mean of regression with their respective 95% posterior quantiles in orange (left, right panels) and blue (middle panel; n = 955 subsampled assemblages).

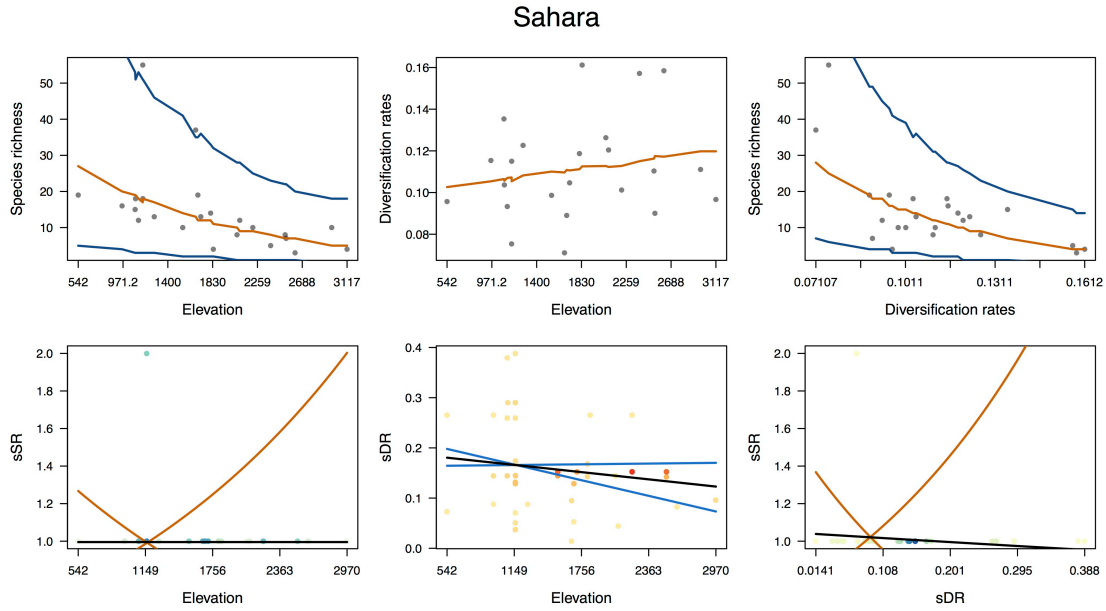

Figure S 2. Species richness and diversification patterns for the Sahara Mountains. Description as in Figure S1. Top row n = 23 assemblages and bottom row n = 89 subsampled assemblages.

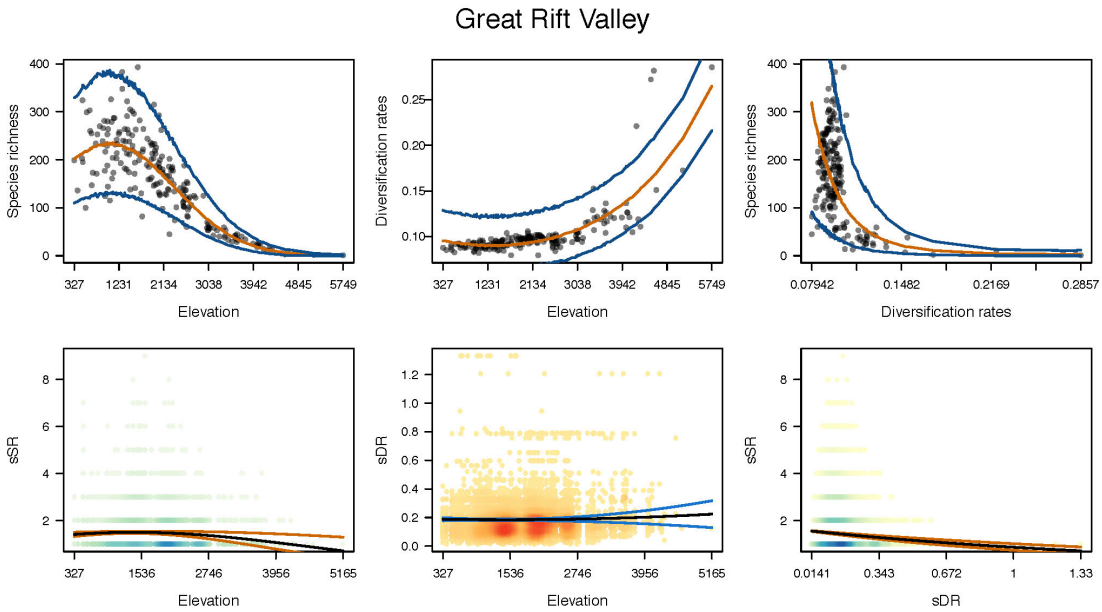

Figure S 3. Species richness and diversification patterns for the Great Rift Valley Mountains. Description as in Figure S1. Top row n = 204 assemblages and bottom row n = 9,153 subsampled assemblages.

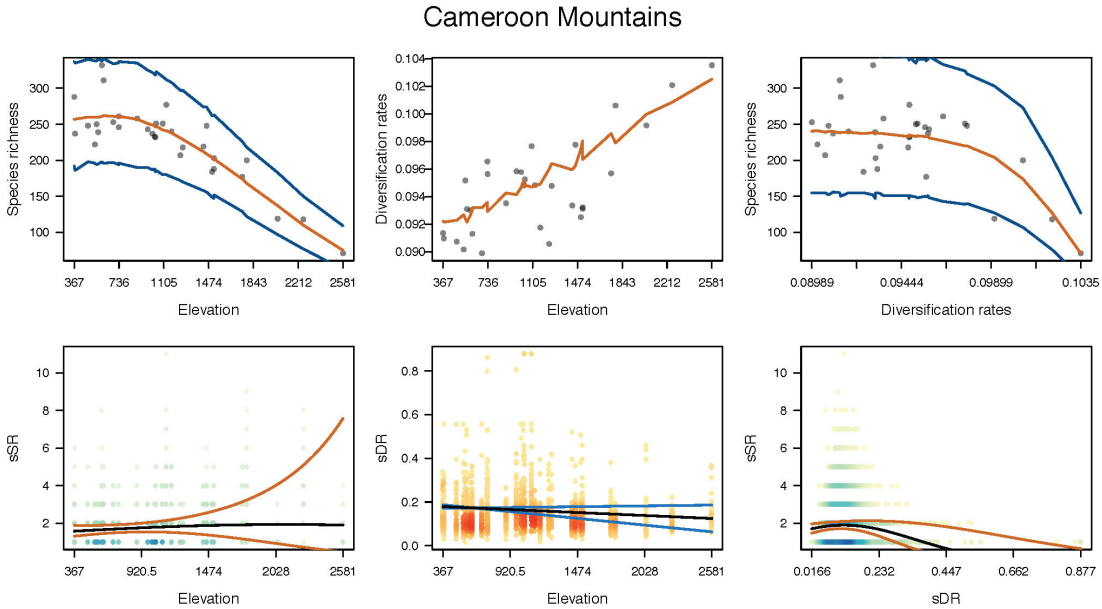

Figure S 4. Species richness and diversification patterns for the Cameroon Mountains. Description as in Figure S1. Top row n = 32 assemblages and bottom row n = 2,099 subsampled assemblages.

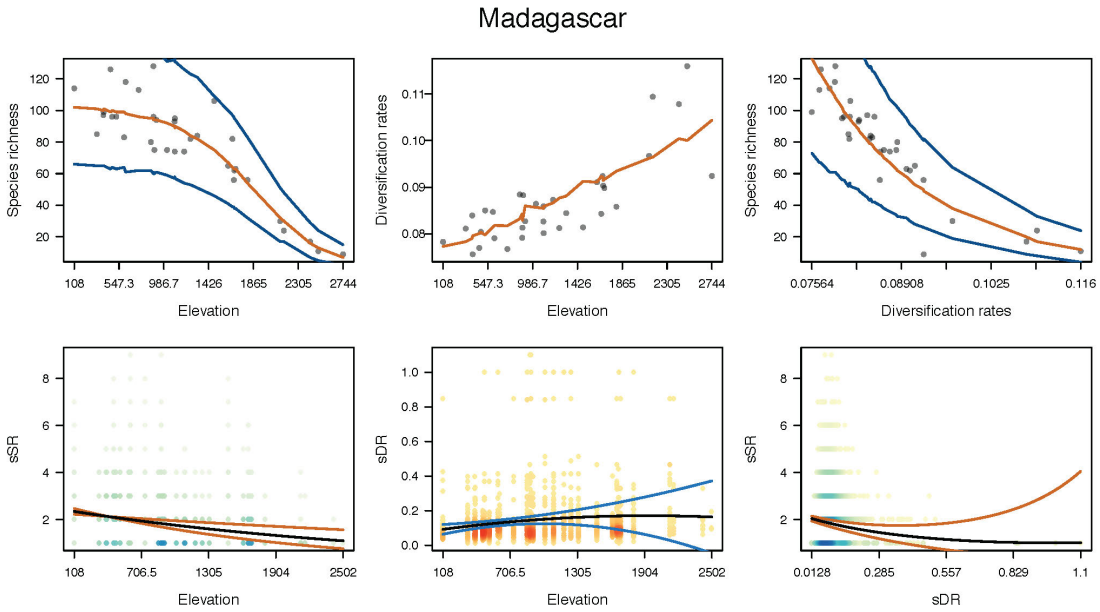

Figure S 5. Species richness and diversification patterns for the mountains in Madagascar. Description as in Figure S1. Top row n = 34 assemblages and bottom row n = 2,062 subsampled assemblages.

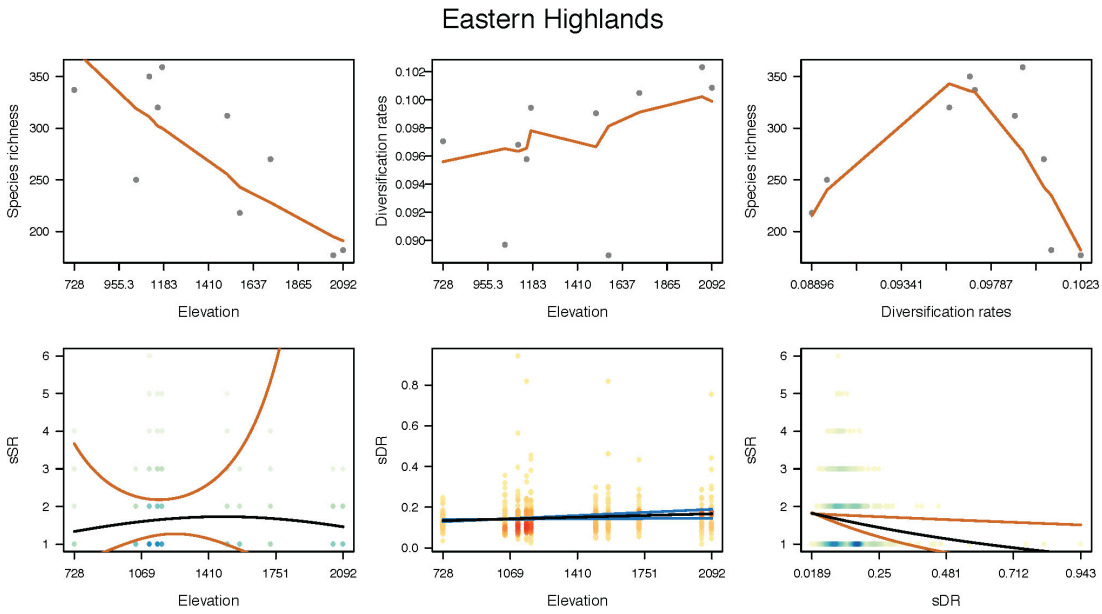

Figure S 6. Species richness and diversification patterns for the Eastern Highlands. Description as in Figure S1. Top row n = 10 assemblages and bottom row n = 656 subsampled assemblages.

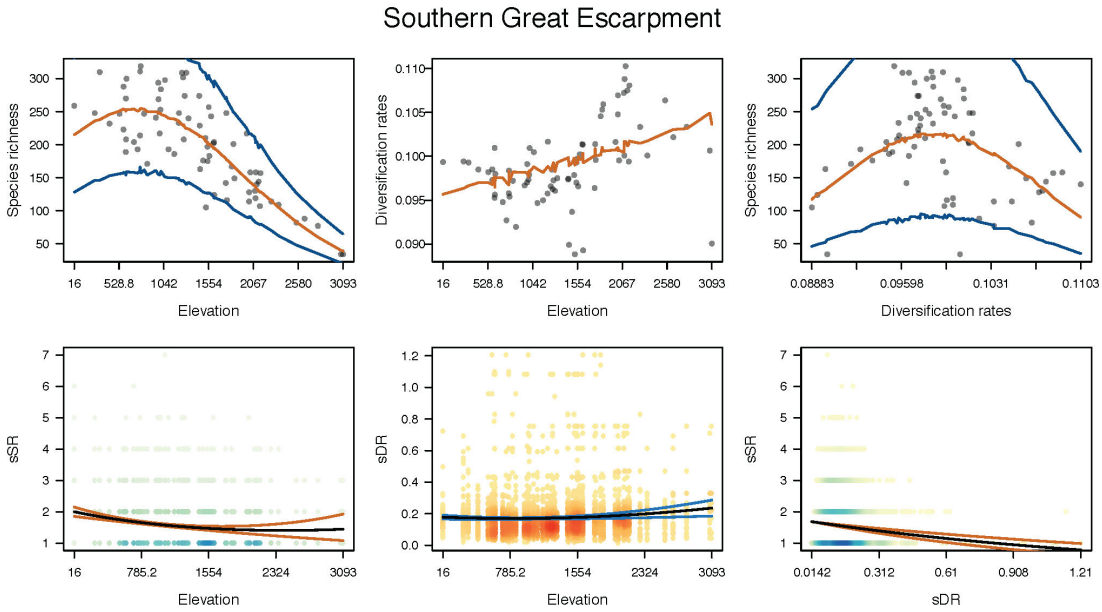

Figure S 7. Species richness and diversification patterns for the Southern Great Escarpment. Description as in Figure S1. Top row n = 73 assemblages and bottom row n = 4,355 subsampled assemblages.

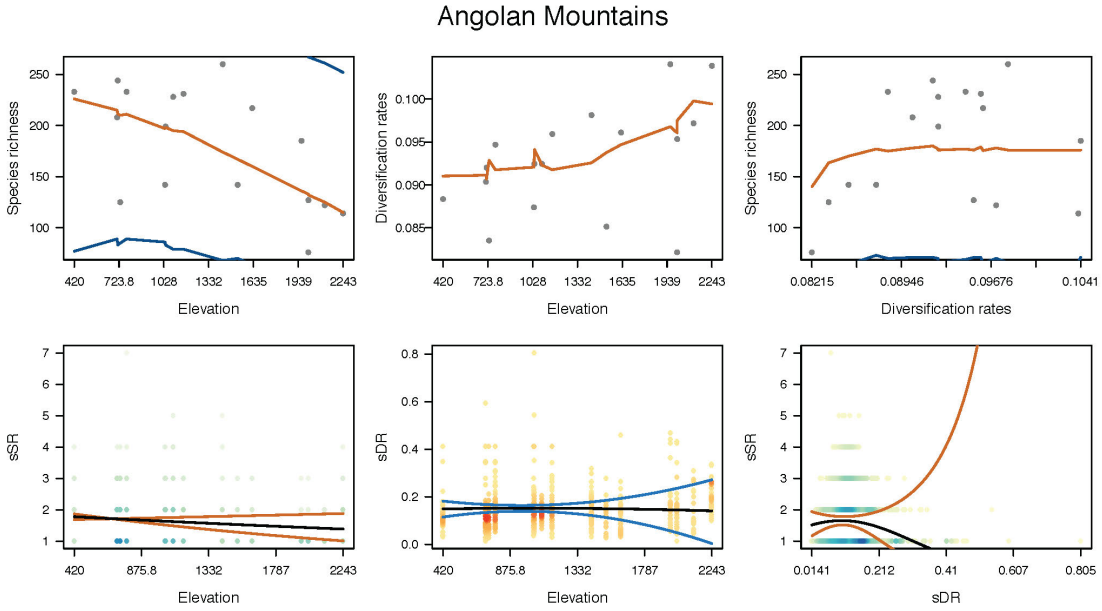

Figure S 8. Species richness and diversification patterns for the Angolan Mountains. Description as in Figure S1. Top row n = 17 assemblages and bottom row n = 995 subsampled assemblages.

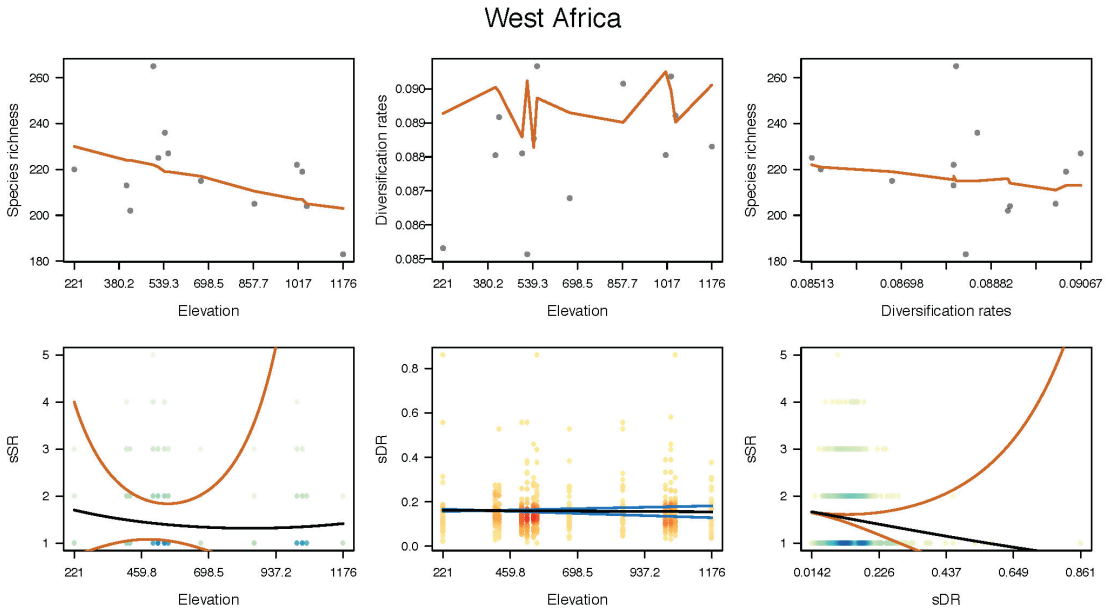

Figure S 9. Species richness and diversification patterns for the West African Mountains. Description as in Figure S1. Top row n = 13 assemblages and bottom row n = 657 subsampled assemblages.

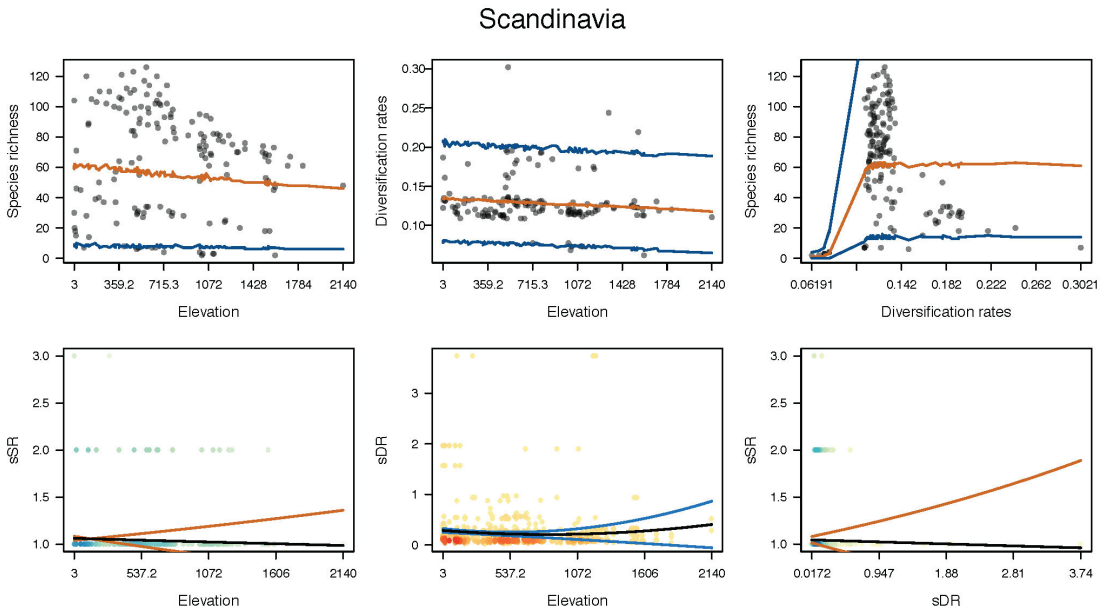

Figure S 10. Species richness and diversification patterns for the Scandinavian Mountains. Description as in Figure S1. Top row n = 152 assemblages and bottom row n = 742 subsampled assemblages.

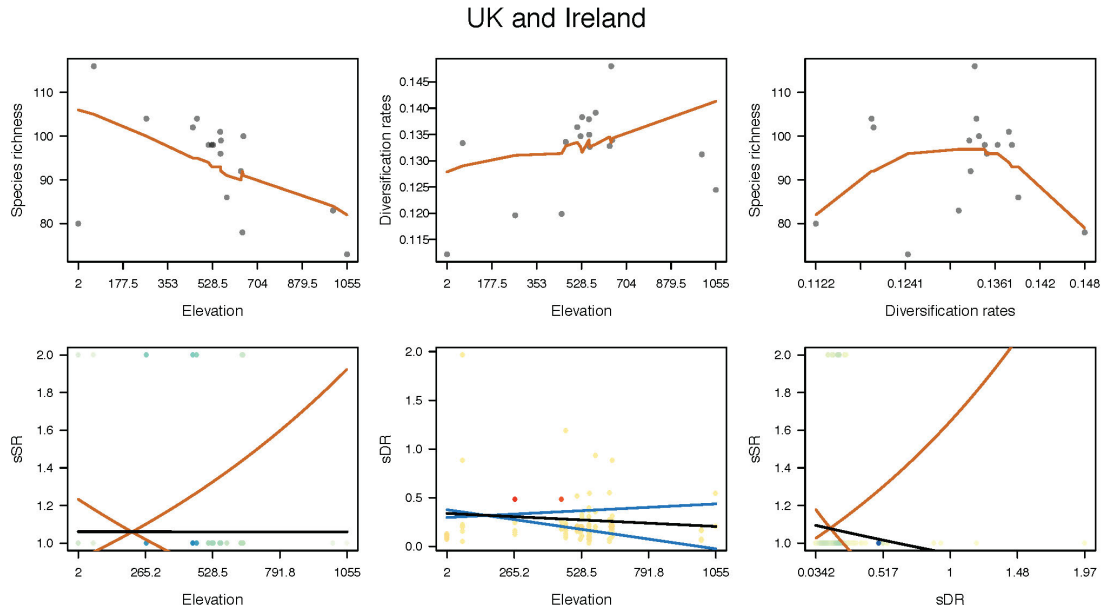

Figure S 11. Species richness and diversification patterns for the mountains in UK and Ireland. Description as in Figure S1. Top row n = 17 assemblages and bottom row n = 201 subsampled assemblages.

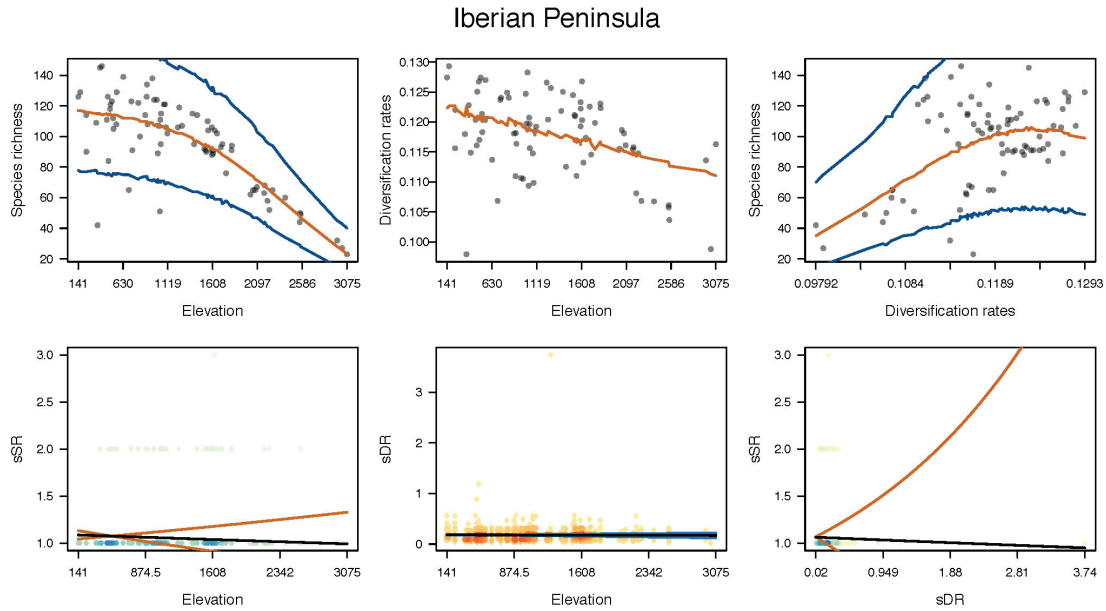

Figure S 12. Species richness and diversification patterns for the mountains in the Iberian Peninsula. Description as in Figure S1. Top row n = 80 assemblages and bottom row n = 794 subsampled assemblages.

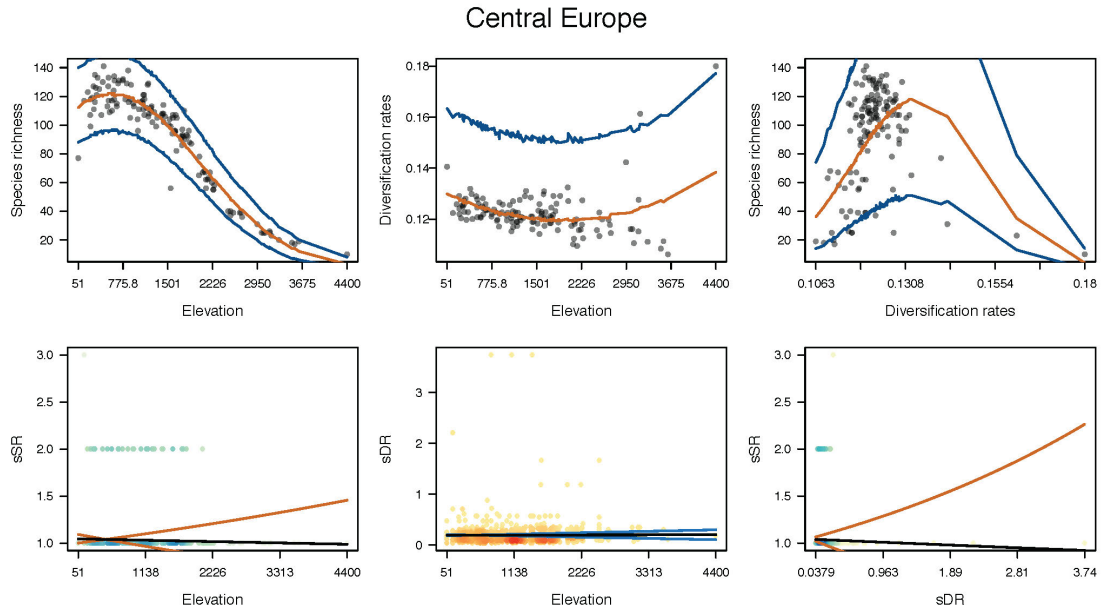

Figure S 13. Species richness and diversification patterns for the mountains in central Europe. Description as in Figure S1. Top row n = 132 assemblages and bottom row n = 1,005 subsampled assemblages.

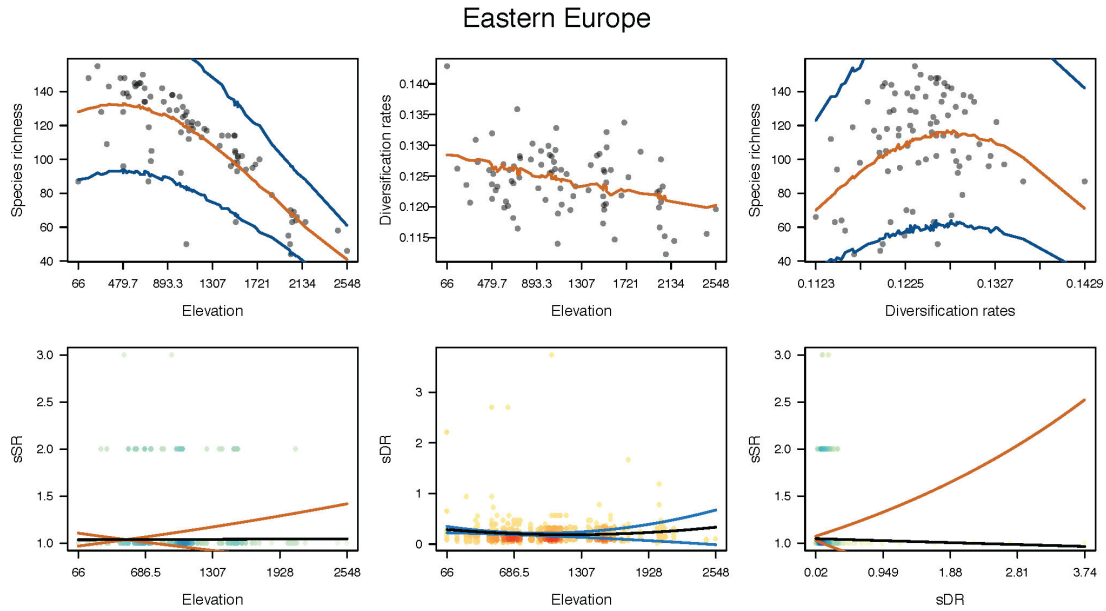

Figure S 14. Species richness and diversification patterns for the mountains in Eastern Europe. Description as in Figure S1. Top row n = 83 assemblages and bottom row n = 714 subsampled assemblages.

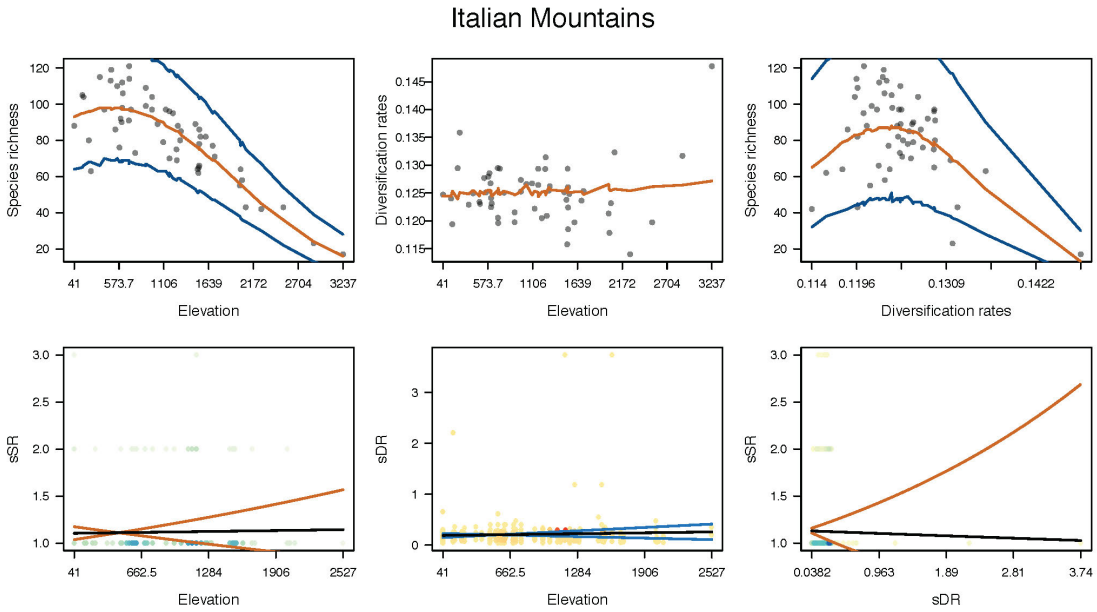

Figure S 15. Species richness and diversification patterns for the Italian Mountains. Description as in Figure S1. Top row n = 56 assemblages and bottom row n = 642 subsampled assemblages.

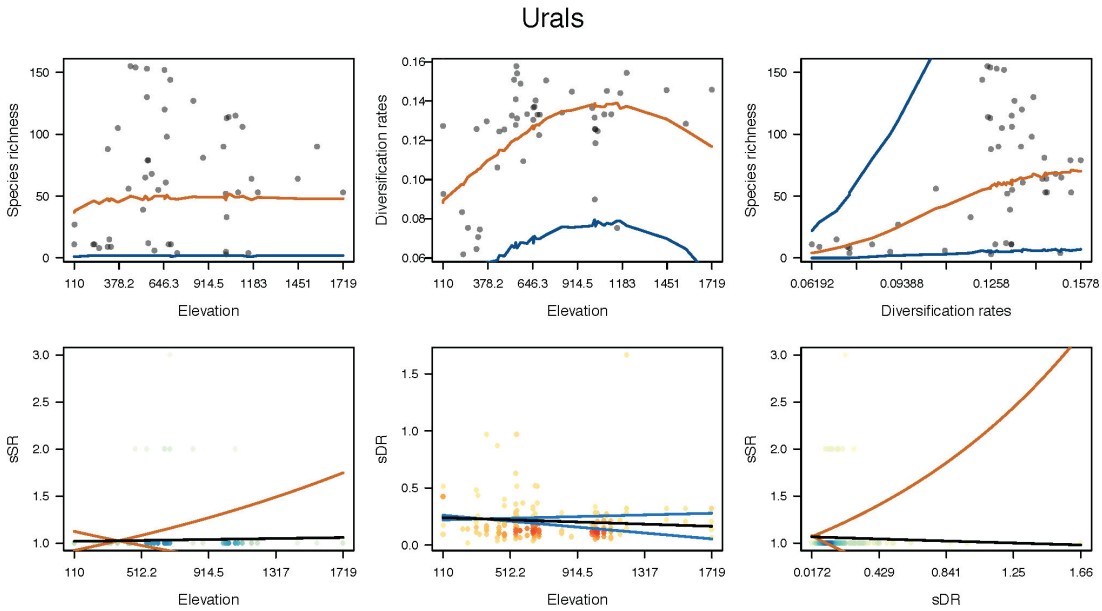

Figure S 16. Species richness and diversification patterns for the Urals. Description as in Figure S1. Top row n = 49 assemblages and bottom row n = 206 subsampled assemblages.

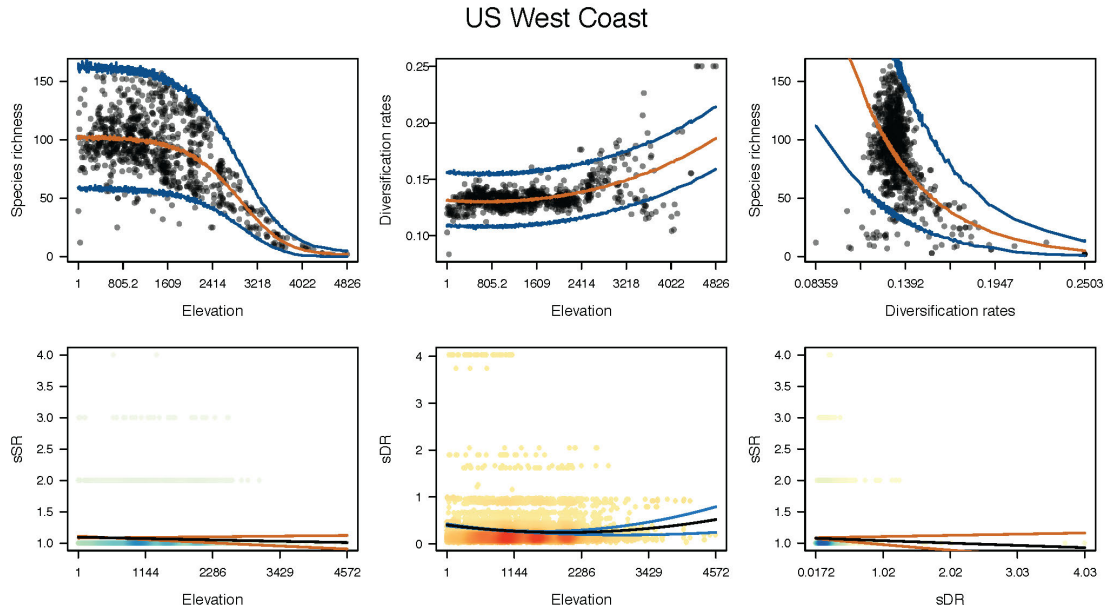

Figure S 17. Species richness and diversification patterns for the US West Coast Mountains. Description as in Figure S1. Top row n = 893 assemblages and bottom row n = 8,249 subsampled assemblages.

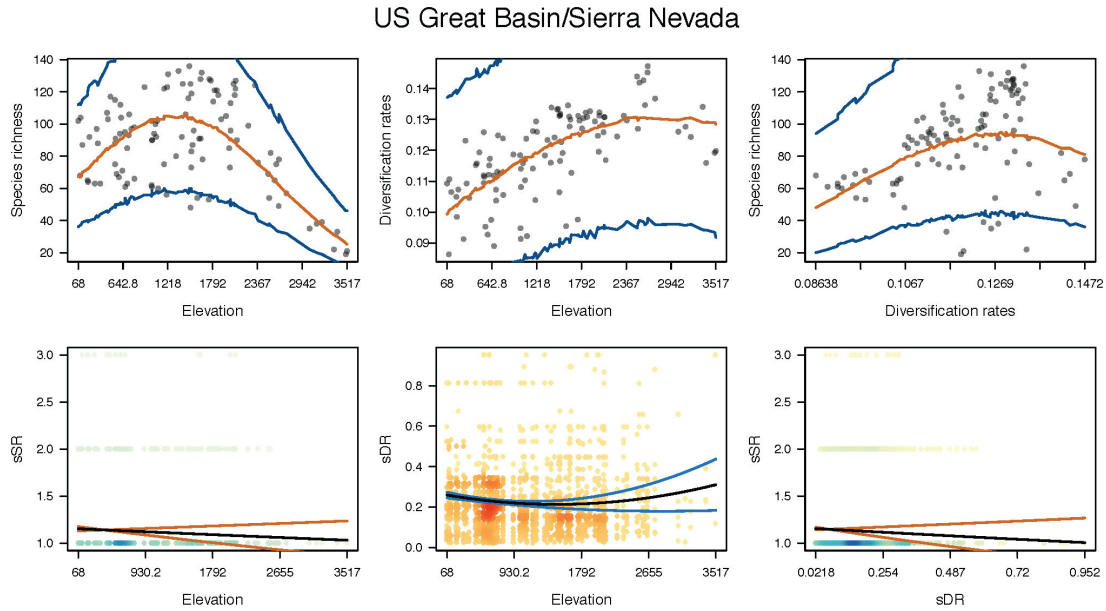

Figure S 18. Species richness and diversification patterns for the US Great Basin/Sierra Nevada Mountains. Description as in Figure S1. Top row n = 105 assemblages and bottom row n = 1,942 subsampled assemblages.

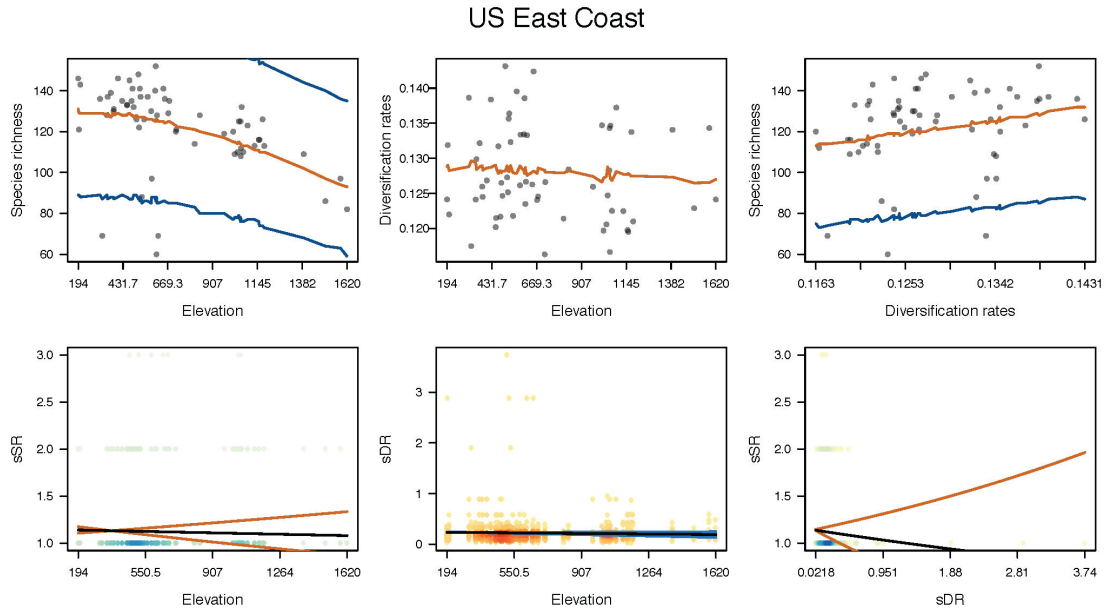

Figure S 19. Species richness and diversification patterns for the US East Coast Mountains. Description as in Figure S1. Top row n = 61 assemblages and bottom row n = 1,094 subsampled assemblages.

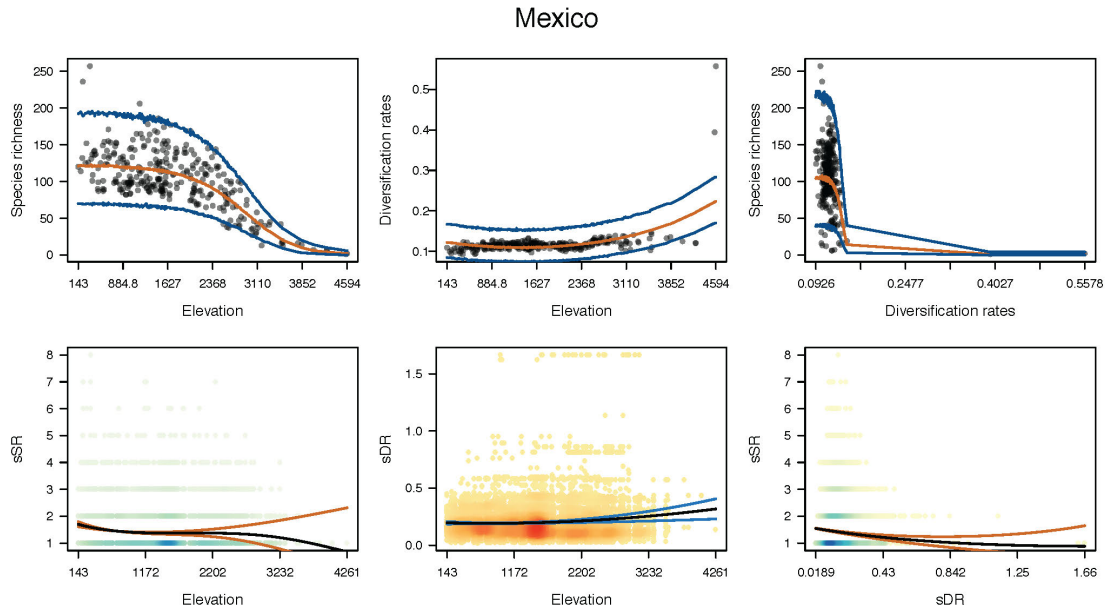

Figure S 20. Species richness and diversification patterns for the Mexico Mountains. Description as in Figure S1. Top row n = 301 assemblages and bottom row n = 12,316 subsampled assemblages.

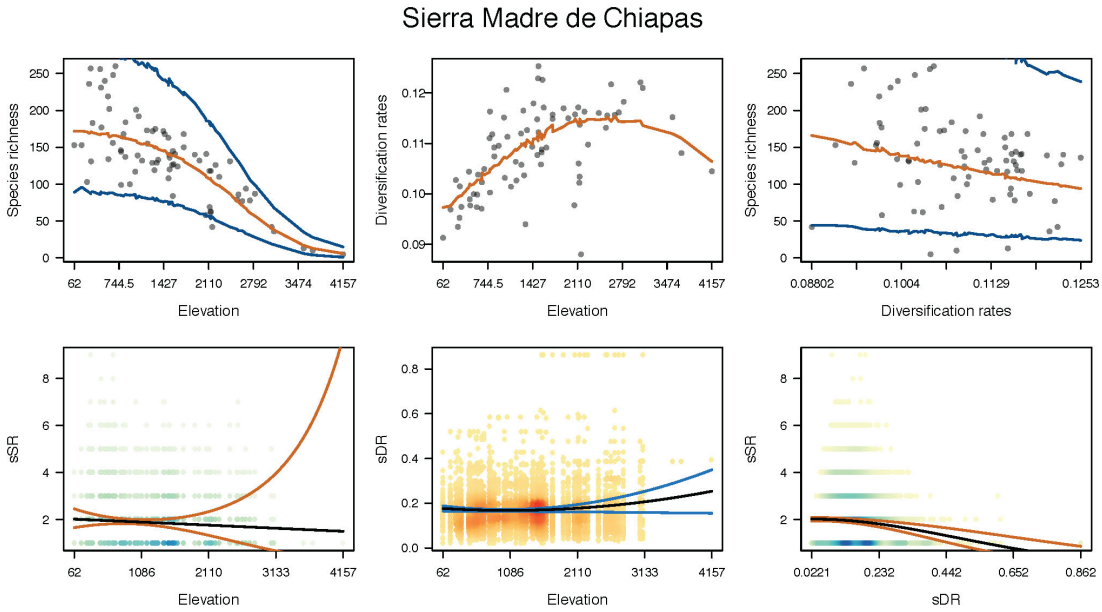

Figure S 21. Species richness and diversification patterns for the Sierra Madre de Chiapas. Description as in Figure S1. Top row n = 76 assemblages and bottom row n = 5,344 subsampled assemblages.

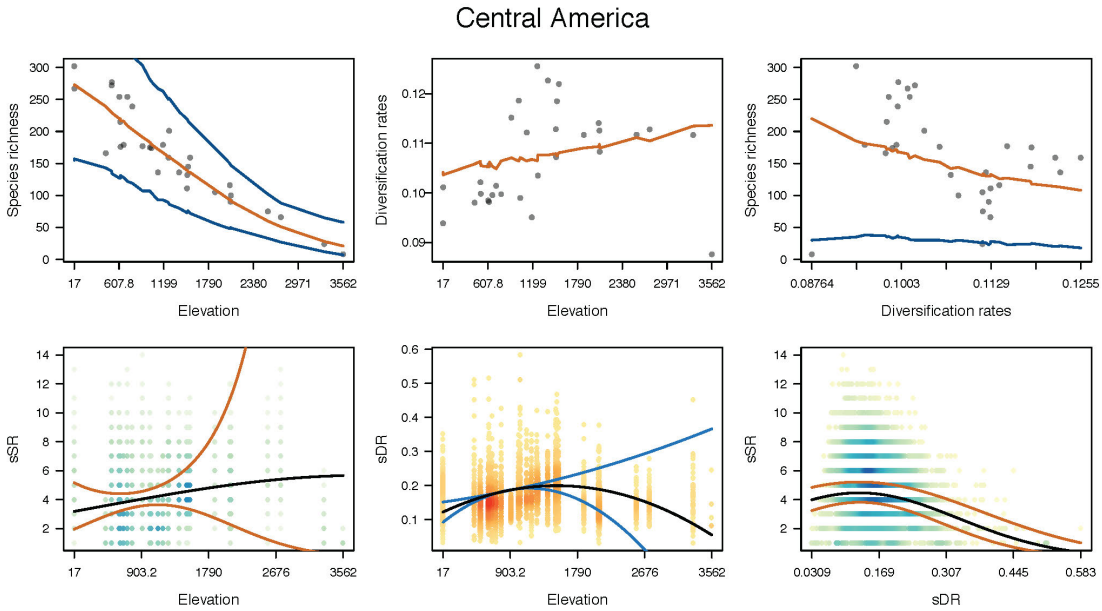

Figure S 22. Species richness and diversification patterns for the mountains in Central America. Description as in Figure S1. Top row n = 31 assemblages and bottom row n = 2,929 subsampled assemblages.

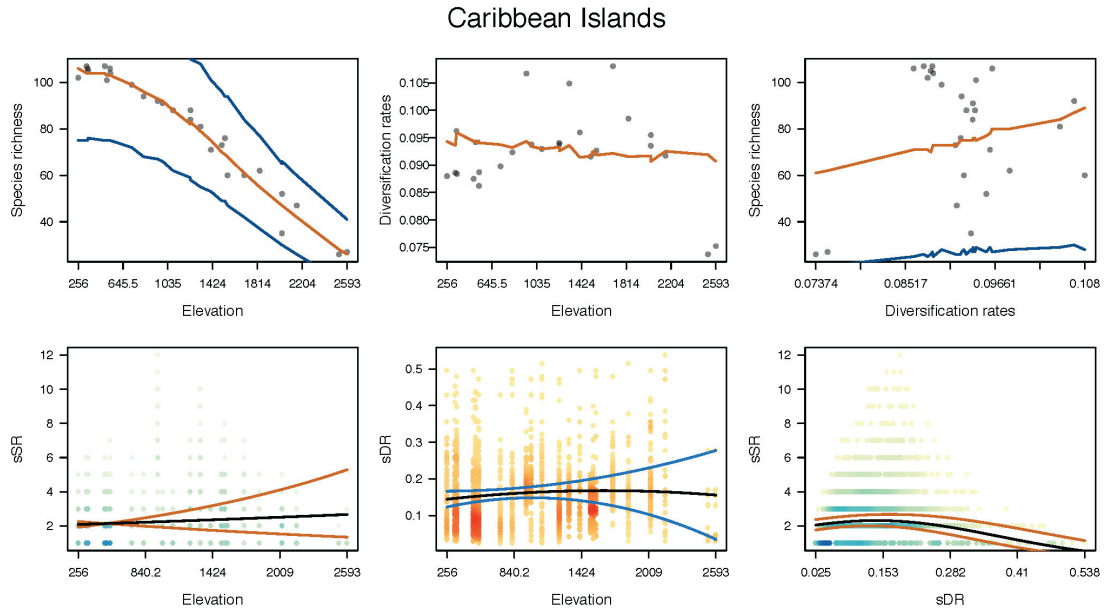

Figure S 23. Species richness and diversification patterns for the Caribbean Islands Mountains. Description as in Figure S1. Top row n = 27 assemblages and bottom row n = 2,106 subsampled assemblages.

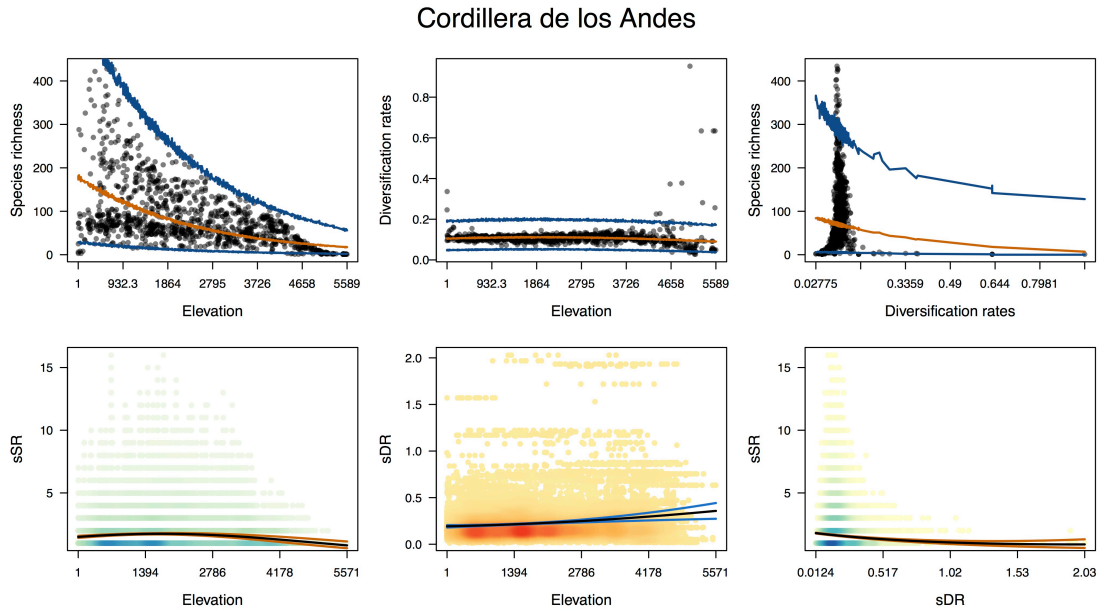

Figure S 24. Species richness and diversification patterns for the Cordillera de los Andes. Description as in Figure S1. Top row n = 1,025 assemblages and bottom row n = 48,297 subsampled assemblages.

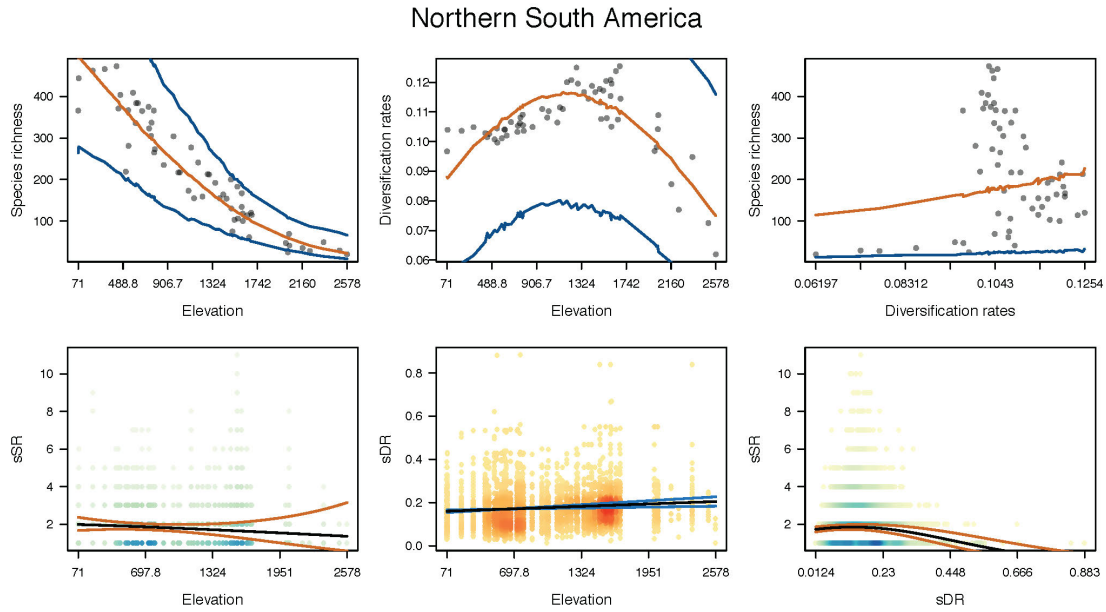

Figure S 25. Species richness and diversification patterns for Northern South America Mountains. Description as in Figure S1. Top row n = 60 assemblages and bottom row n = 3,718 subsampled assemblages.

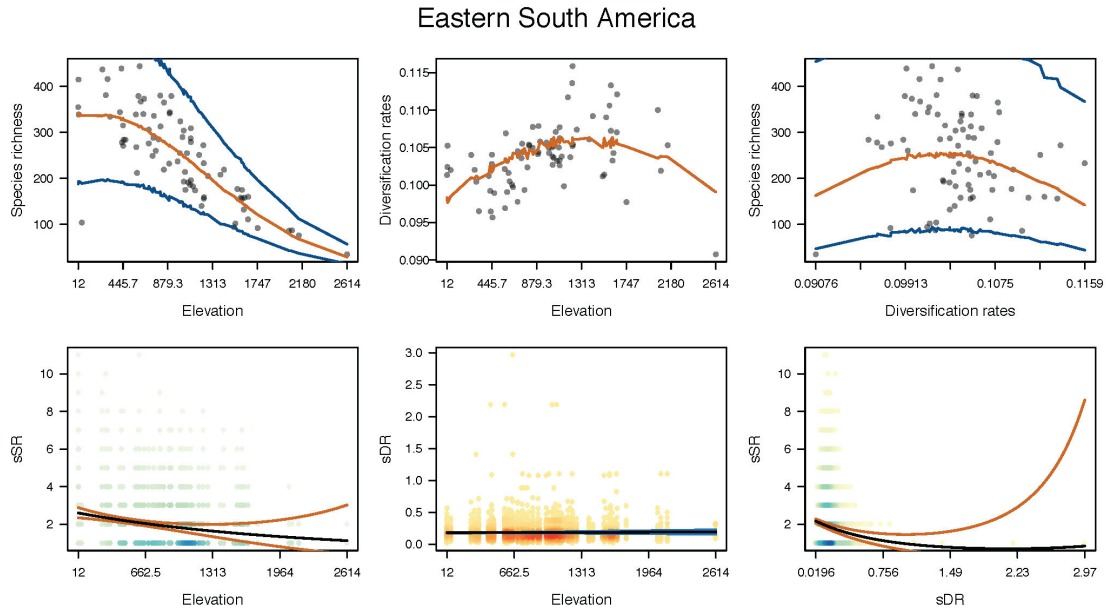

Figure S 26. Species richness and diversification patterns for the Eastern South America Mountains. Description as in Figure S1. Top row n = 79 assemblages and bottom row n = 5,292 subsampled assemblages.

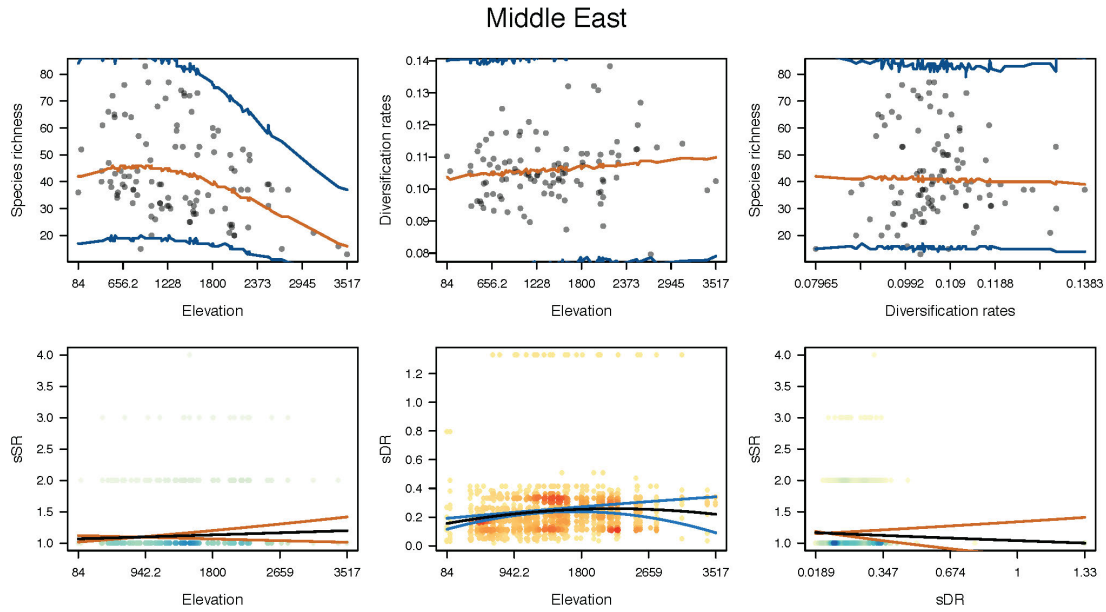

Figure S 27. Species richness and diversification patterns for the Middle East Mountains. Description as in Figure S1. Top row n = 107 assemblages and bottom row n = 1,633 subsampled assemblages.

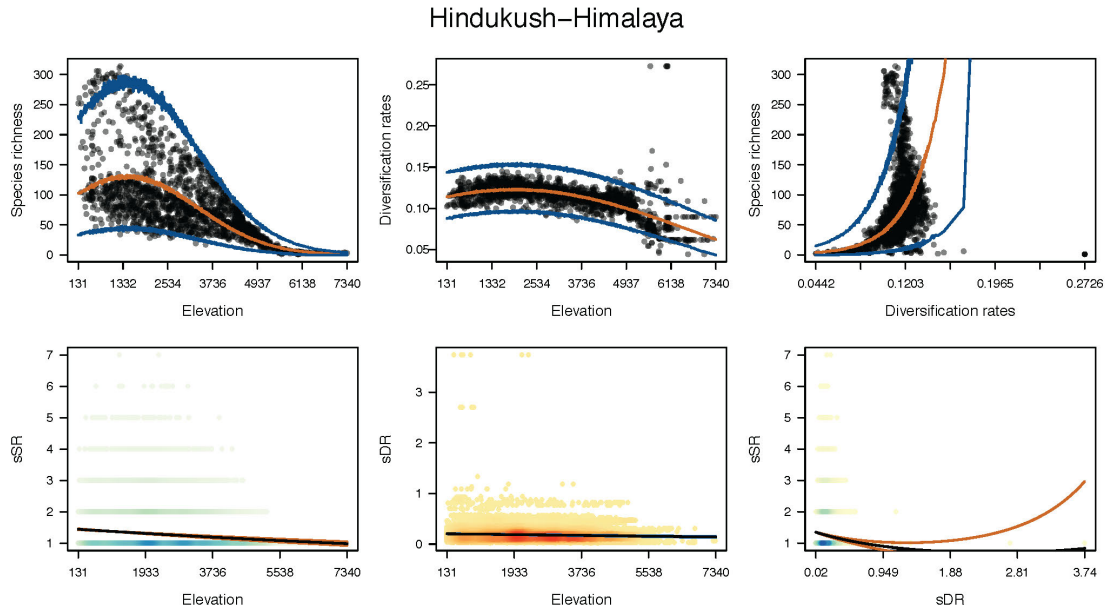

Figure S 28. Species richness and diversification patterns for the Himalayas. Description as in Figure S1. Top row n = 1,816 assemblages and bottom row n = 31,780 subsampled assemblages.

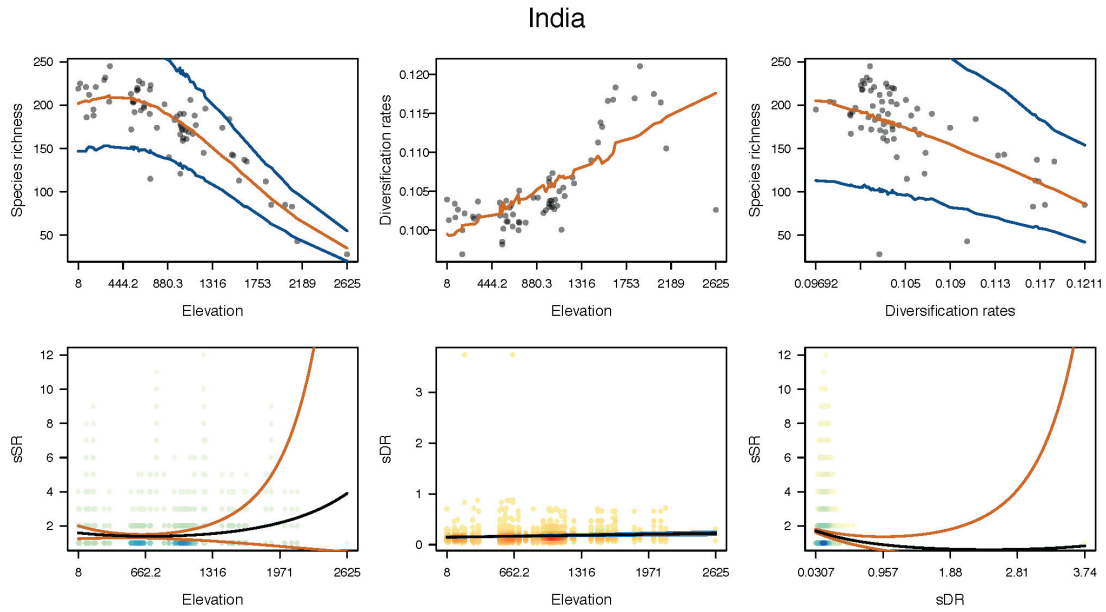

Figure S 29. Species richness and diversification patterns for southern India mountains. Description as in Figure S1. Top row n = 67 assemblages and bottom row n = 3,616 subsampled assemblages.

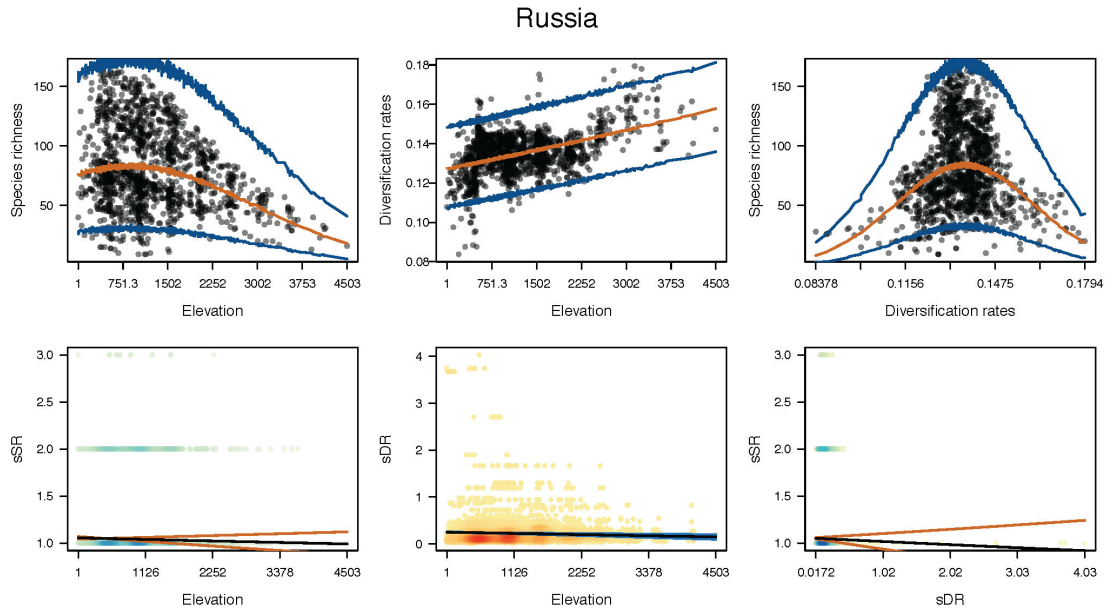

Figure S 30. Species richness and diversification patterns for the Russian Mountains. Description as in Figure S1. Top row n = 1,377 assemblages and bottom row n = 7,686 subsampled assemblages.

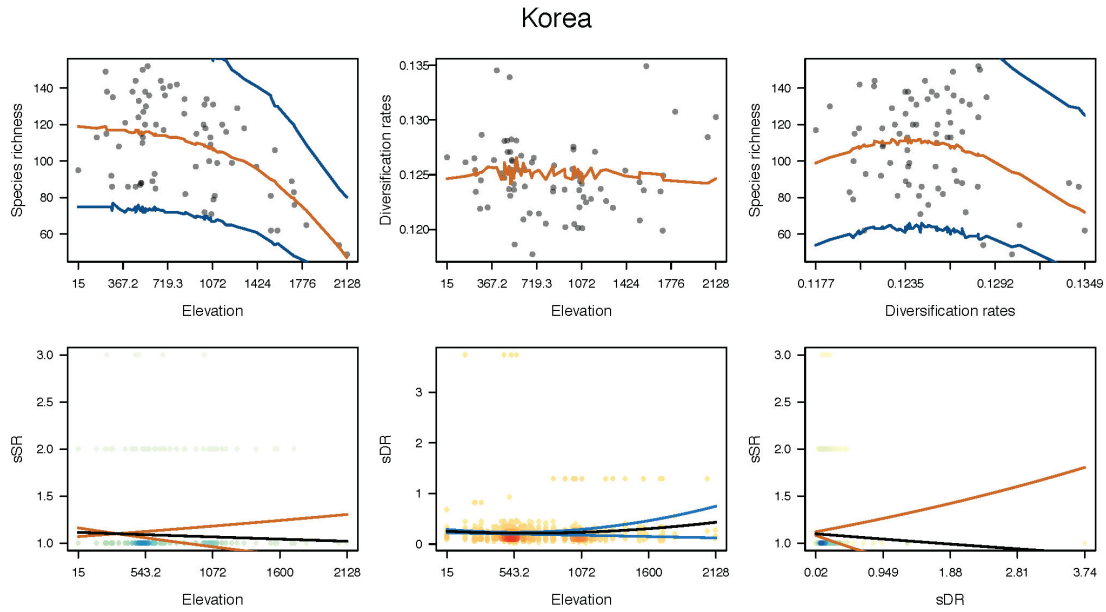

Figure S 31. Species richness and diversification patterns for the Korean Mountains. Description as in Figure S1. Top row n = 73 assemblages and bottom row n = 1,075 subsampled assemblages.

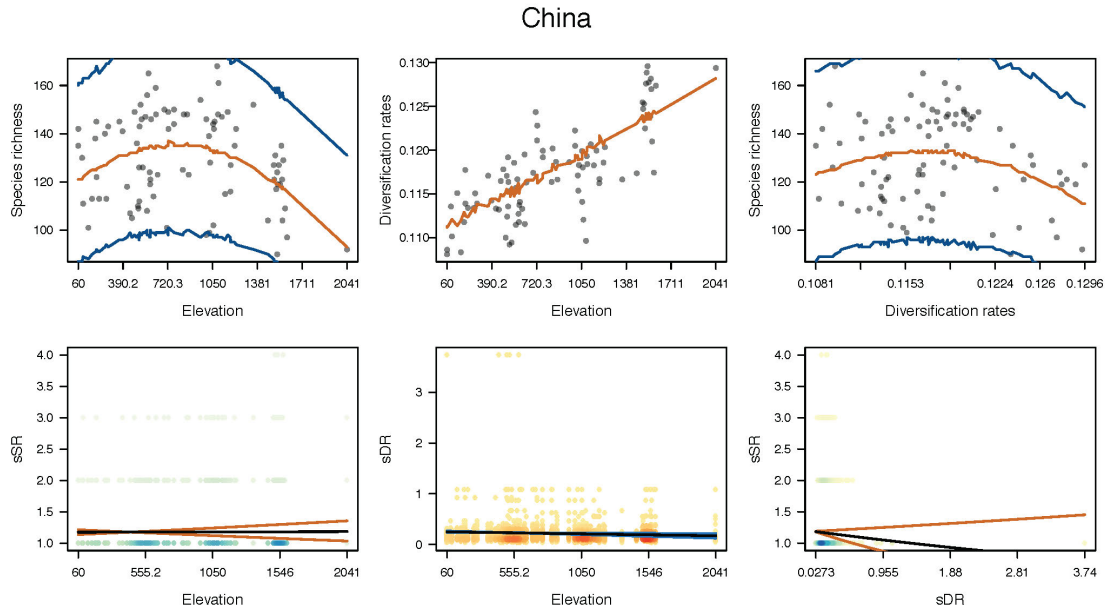

Figure S 32. Species richness and diversification patterns for mountains in China. Description as in Figure S1. Top row n = 88 assemblages and bottom row n = 2,204 subsampled assemblages.

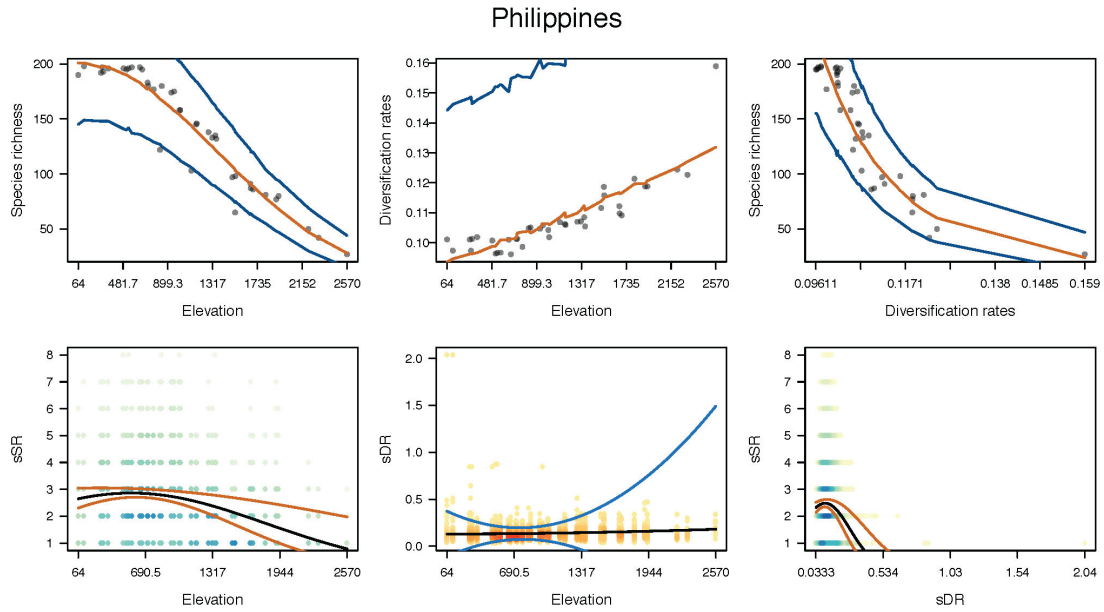

Figure S 33. Species richness and diversification patterns for the Philippines Mountains. Description as in Figure S1. Top row n = 40 assemblages and bottom row n = 3,406 subsampled assemblages.

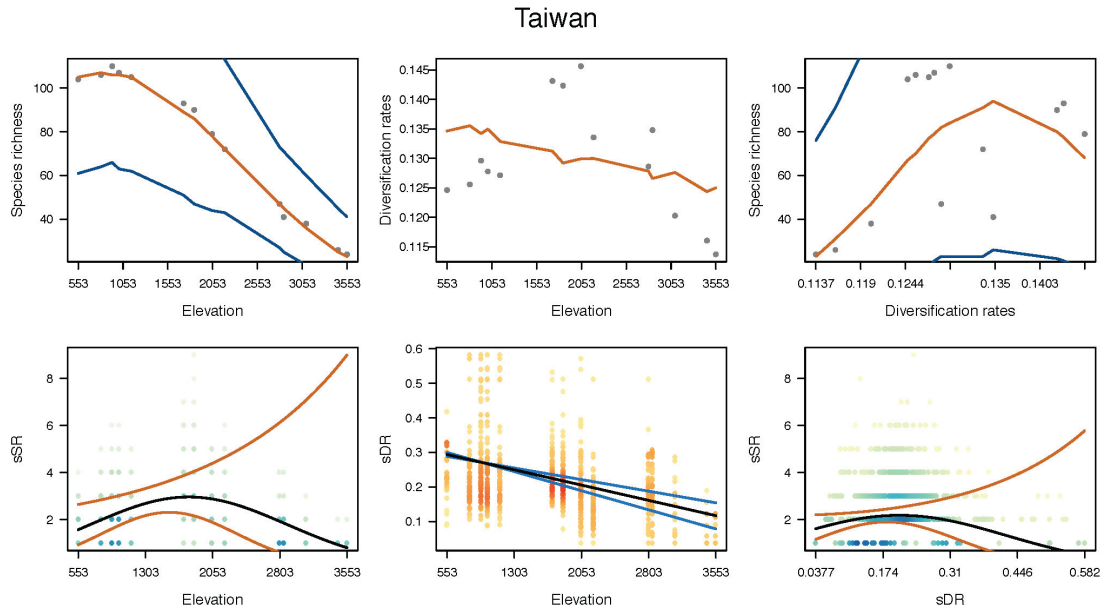

Figure S 34. Species richness and diversification patterns for Taiwan Mountains. Description as in Figure S1. Top row n = 14 assemblages and bottom row n = 1,121 subsampled assemblages.

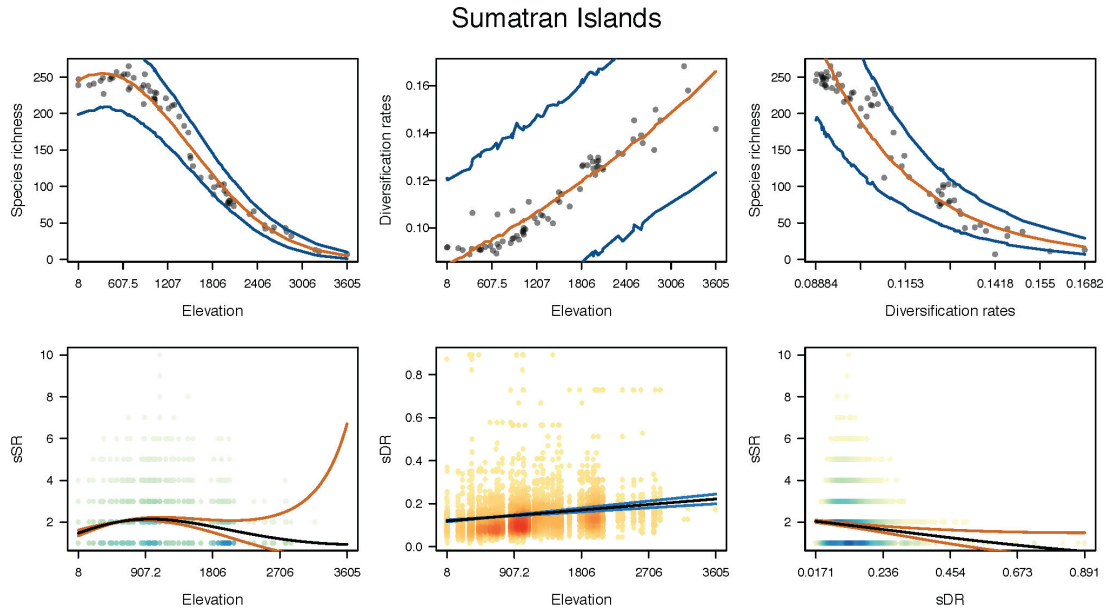

Figure S 35. Species richness and diversification patterns for the Sumatran Island Mountains. Description as in Figure S1. Top row n = 69 assemblages and bottom row n = 4,689 subsampled assemblages.

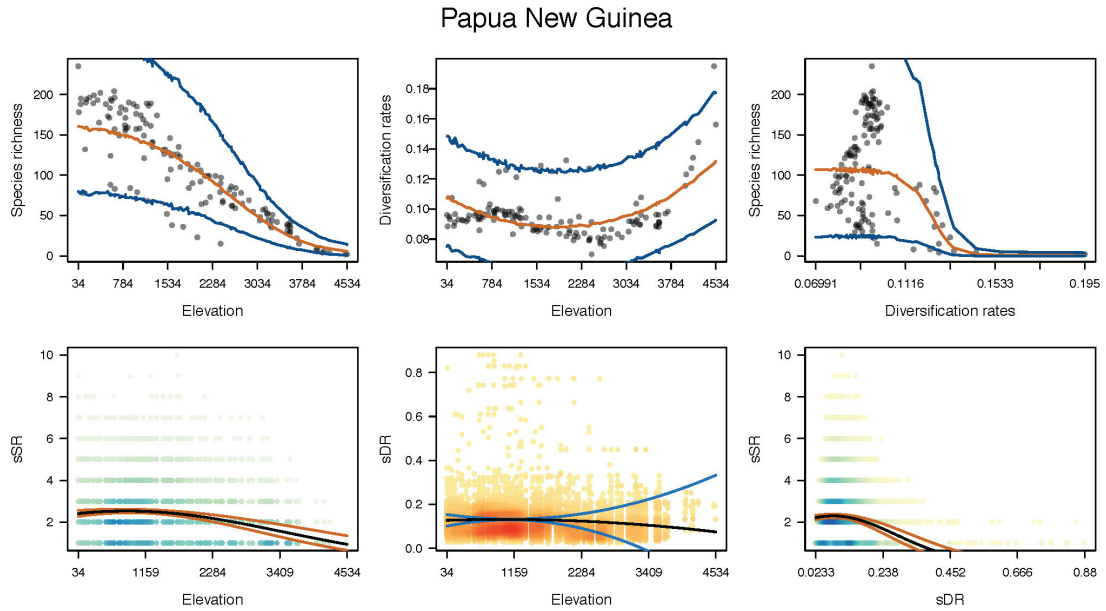

Figure S 36. Species richness and diversification patterns for the mountains in Papua New Guinea. Description as in Figure S1. Top row n = 134 assemblages and bottom row n = 10,778 subsampled assemblages.

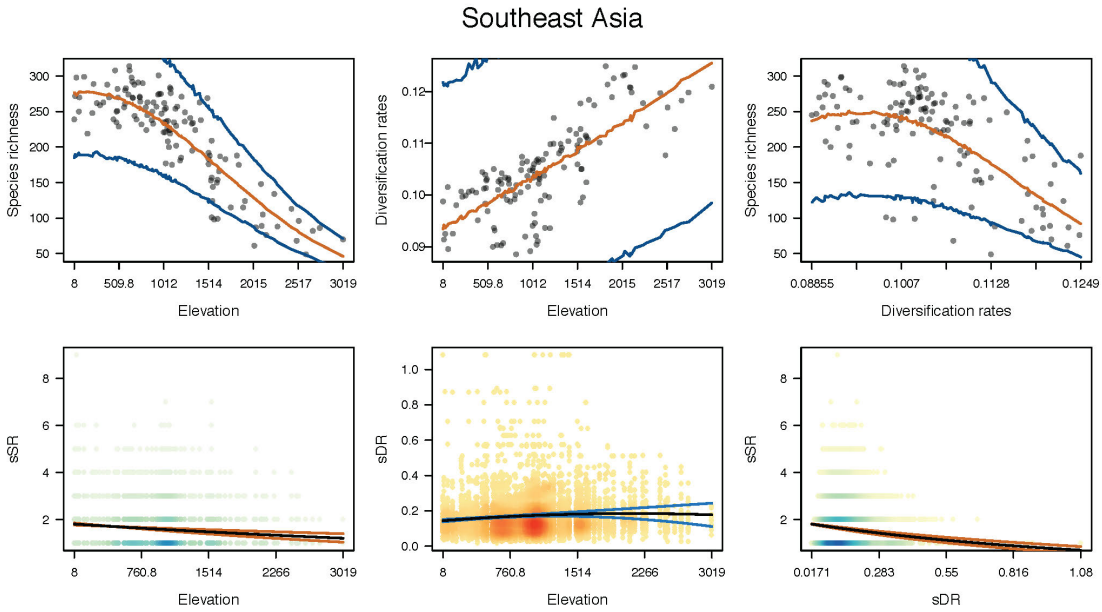

Figure S 37. Species richness and diversification patterns for the Southeast Asian Mountains. Description as in Figure S1. Top row n = 130 assemblages and bottom row n = 7,691 subsampled assemblages.

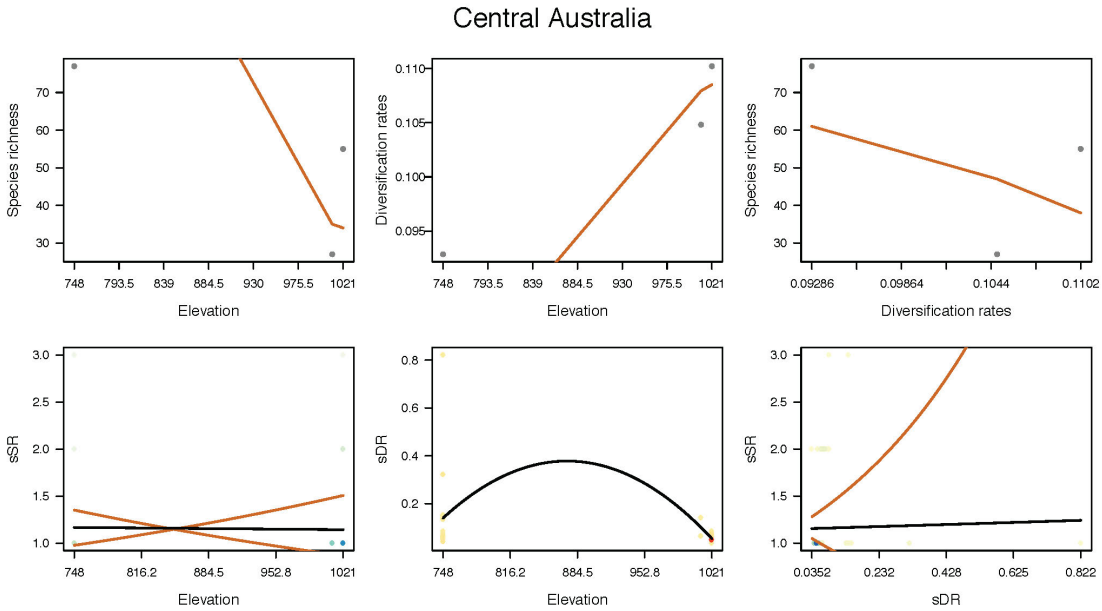

Figure S 38. Species richness and diversification patterns for mountains in Central Australia. Description as in Figure S1. Top row n = 3 assemblages and bottom row n = 130 subsampled assemblages.

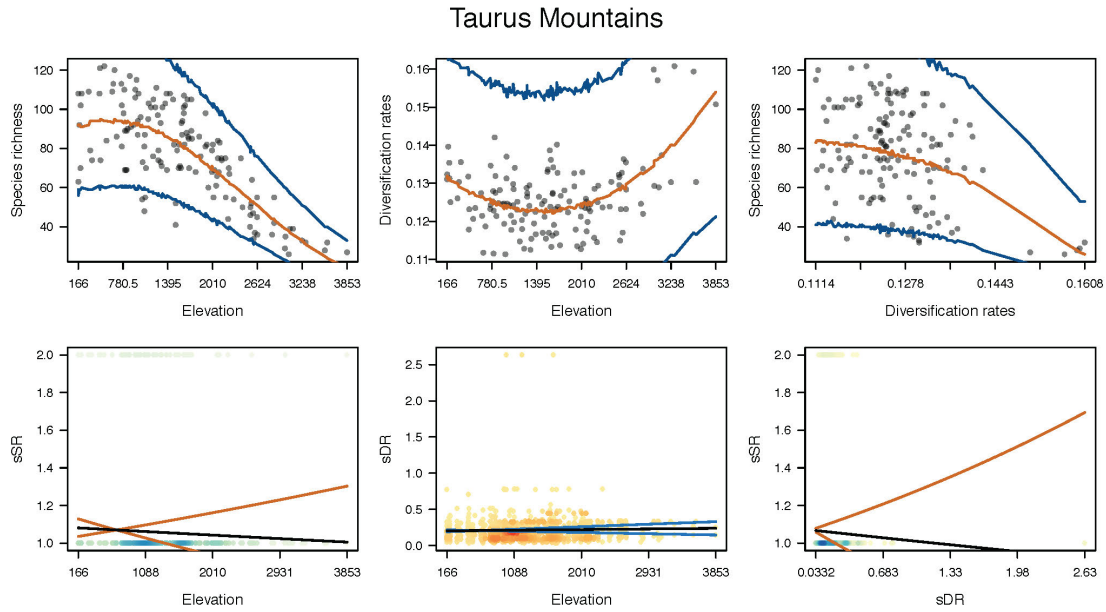

Figure S 39. Species richness and diversification patterns for the Taurus Mountains. Description as in Figure S1. Top row n = 141 assemblages and bottom row n = 1,316 subsampled assemblages.

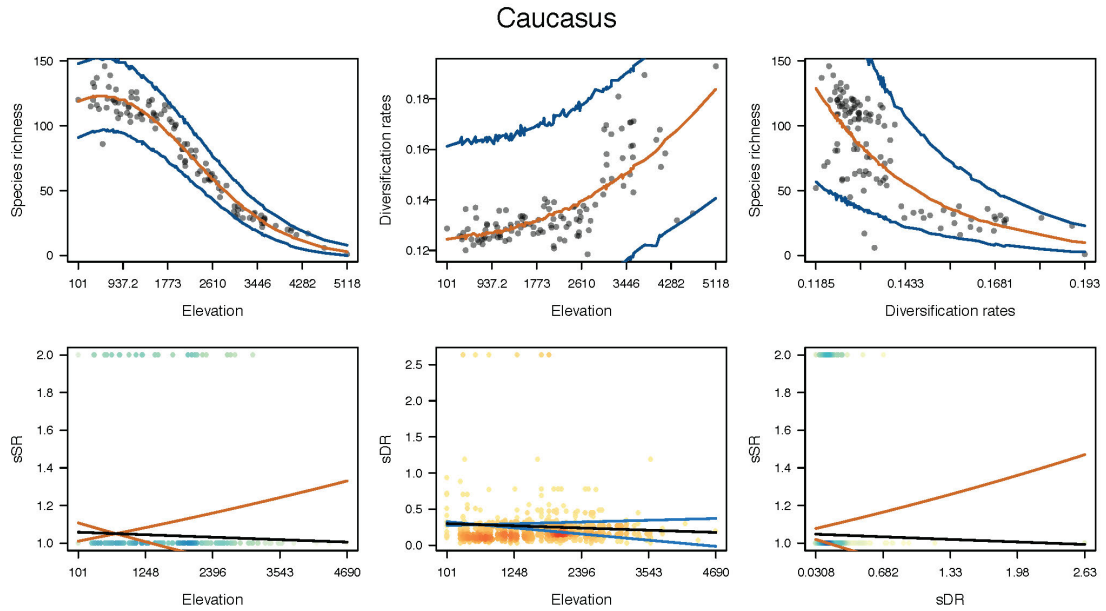

Figure S 40. Species richness and diversification patterns for the Caucasus. Description as in Figure S1. Top row n = 113 assemblages and bottom row n = 920 subsampled assemblages.

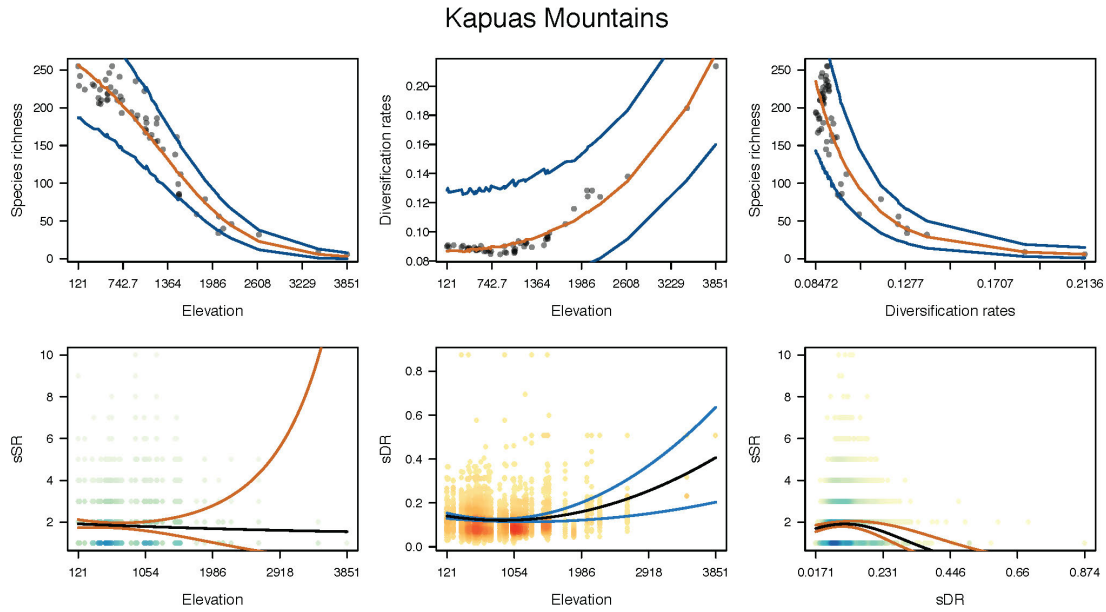

Figure S 41. Species richness and diversification patterns for the Kapuas Mountains. Description as in Figure S1. Top row n = 58 assemblages and bottom row n = 4,063 subsampled assemblages.

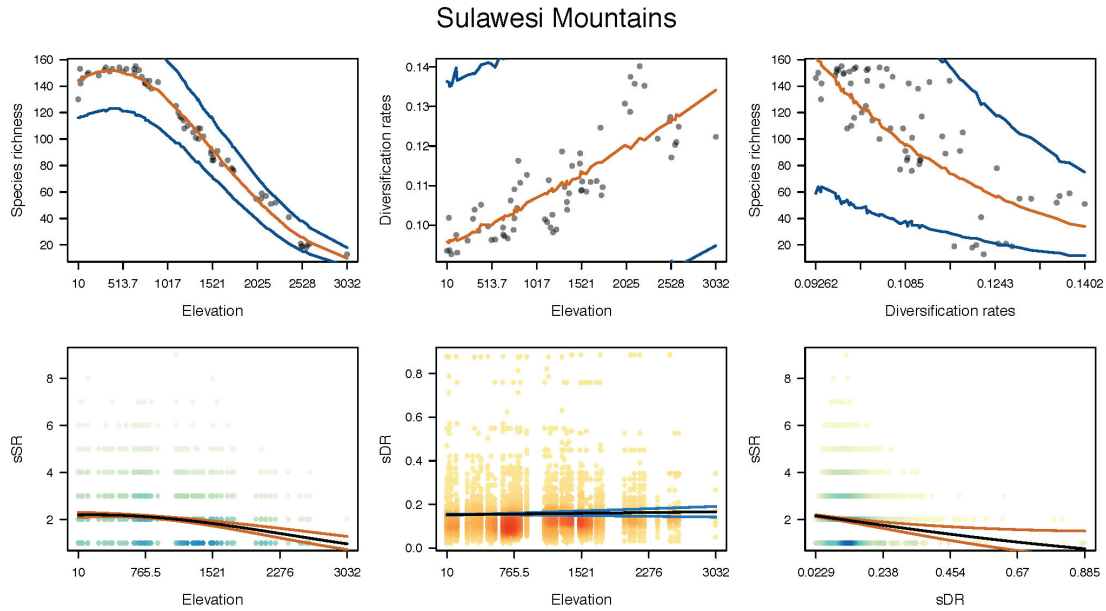

Figure S 42. Species richness and diversification patterns for the Sulawesi Mountains. Description as in Figure S1. Top row n = 64 assemblages and bottom row n = 4,704 subsampled assemblages.

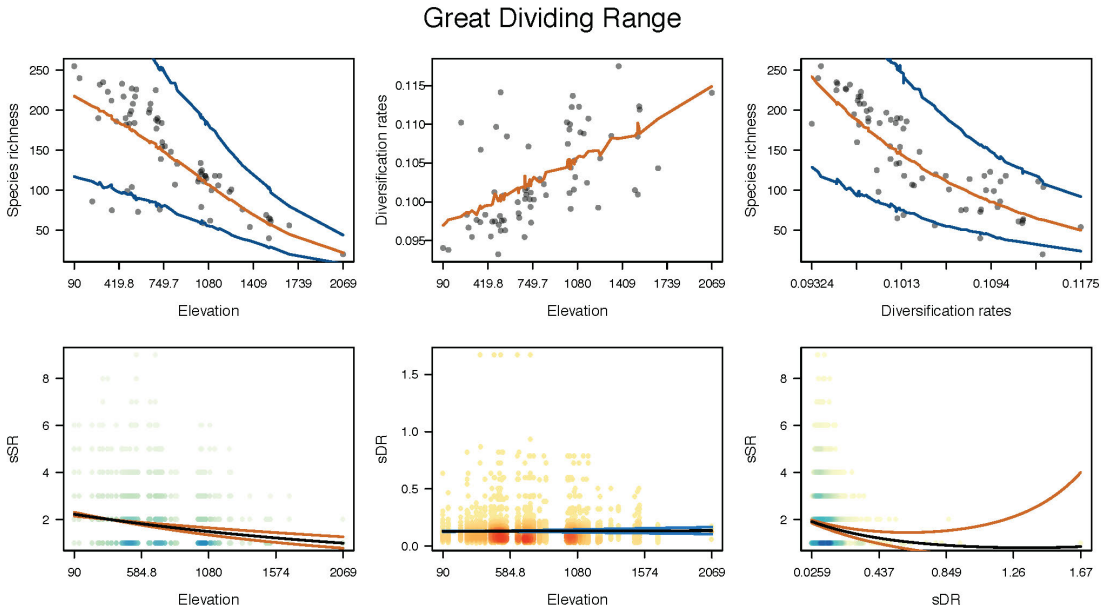

Figure S 43. Species richness and diversification patterns for the Great Dividing Range. Description as in Figure S1. Top row n = 70 assemblages and bottom row n = 4,588 subsampled assemblages.

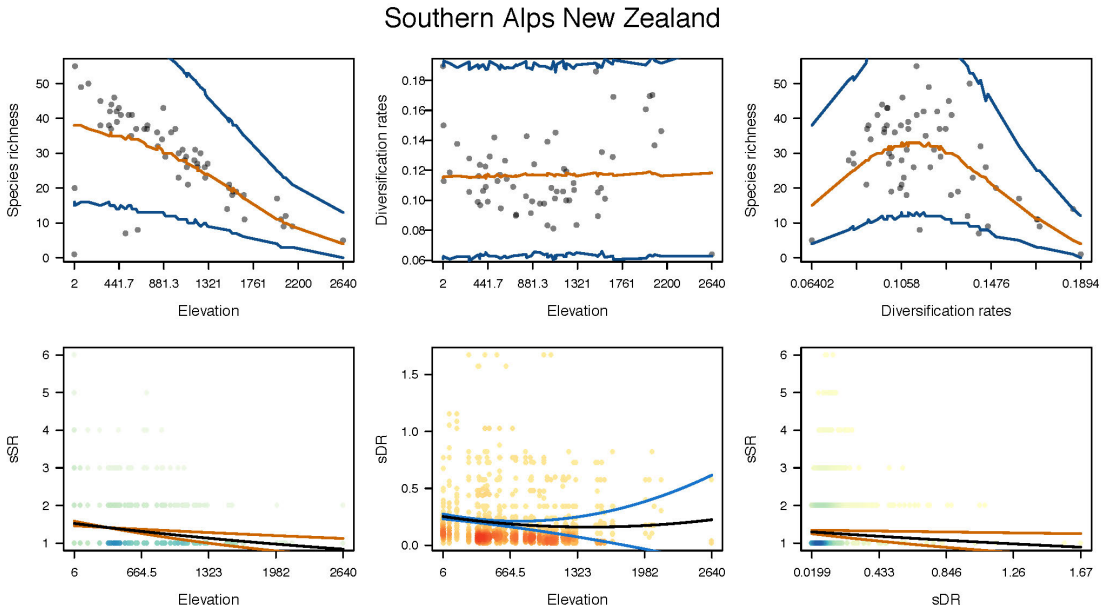

Figure S 44. Species richness and diversification patterns for the Southern Alps in New Zealand. Description as in Figure S1. Top row n = 61 assemblages and bottom row n = 1,983 subsampled assemblages.

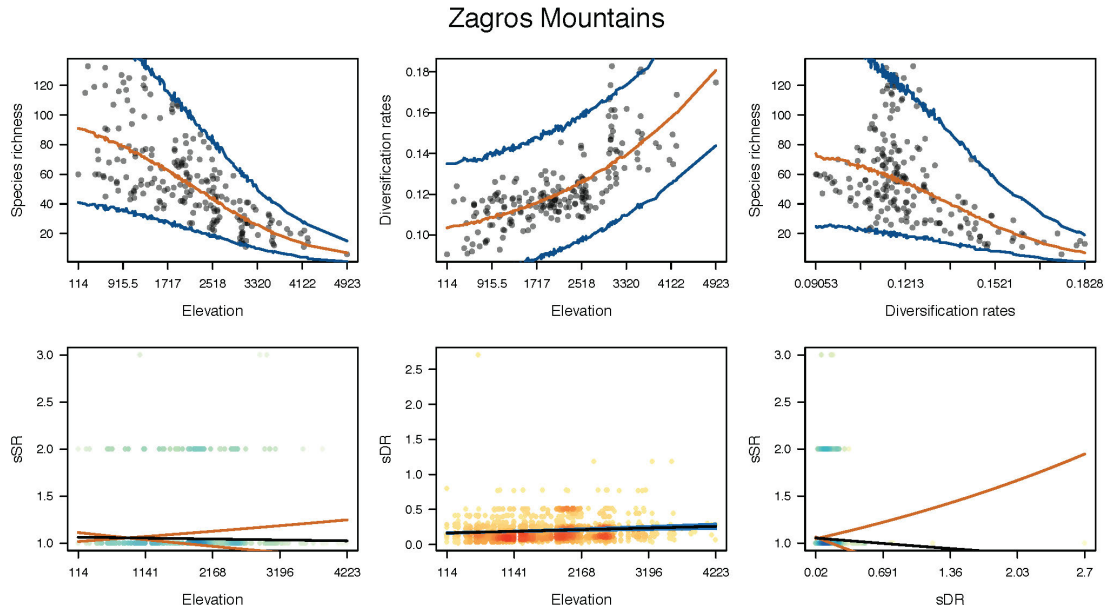

Figure S 45. Species richness and diversification patterns for the Zagros Mountains. Description as in Figure S1. Top row n = 201 assemblages and bottom row n = 1,536 subsampled assemblages.

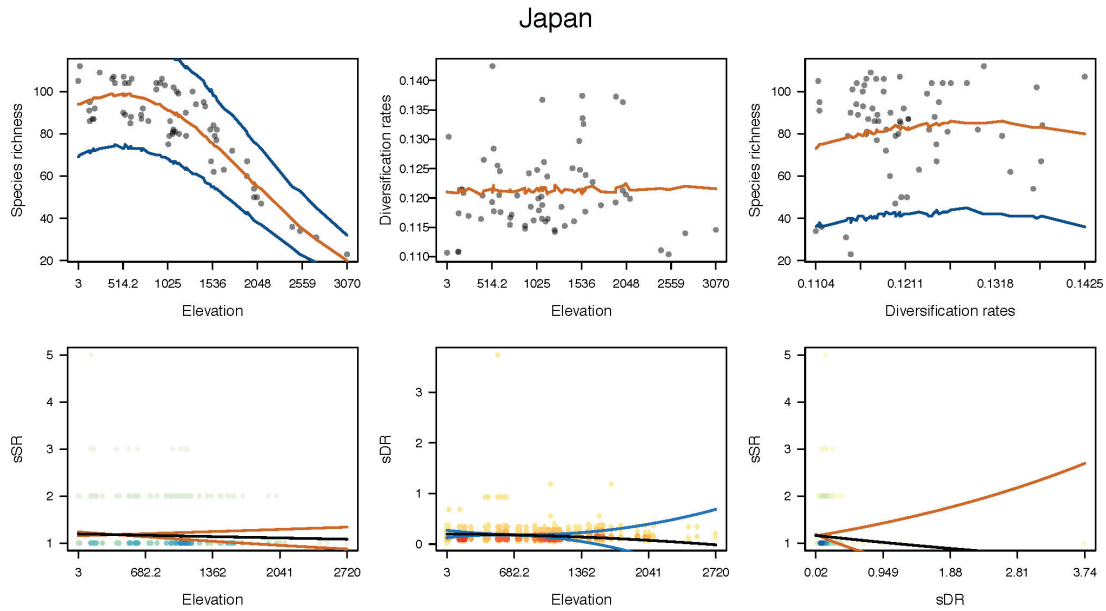

Figure S 46. Species richness and diversification patterns for the mountains in Japan. Description as in Figure S1. Top row n = 65 assemblages and bottom row n = 1,450 subsampled assemblages.