**Supplemental information to accompany the manuscript “Interactions between plants and primates shape community diversity in a rainforest in Madagascar” by James P. Herrera**

**Appendix S1. Supplementary Information on lemur field surveys**

*Survey methodology:*

Survey personnel: I trained six Malagasy research assistants in the methods outlined below during the first three months of fieldwork and I conducted routine inter-observer reliability assessments with personnel throughout the study period to avoid observer drift. We conducted tests of distance and height estimates made by eye compared to measured values to ensure precision and accuracy in these data (all estimates +/-1m). I rotated all assistants to minimized systematic bias due to which observer recorded data.

Field data collection: *Lemur sampling*: I conducted lemur surveys during overlapping months among localities to reduce variation in abundance estimates due to the seasonality. Each locality was sampled during 2-8 months of each the wet and dry seasons (Table S1). Three nocturnal species, *Microcebus rufus, Cheirogaleus crossleyi* and *C. sibreei* are less active or entirely dormant during the dry season, using daily torpor or hibernation to survive periods of food scarcity (Dausman et al. 2004, Blanco and Rahalinarivo 2010). Nocturnal sampling was decreased between May and July and the total survey efforts were adjusted for those species. For *Cheirogaleus*, survey effort included all surveys conducted between the first sighting of the year (between August and September) and the last sighting (May). For *Microcebus*, separate survey efforts were calculated for dry (May – September) and wet seasons. With my team of research assistants, I conducted diurnal and nocturnal surveys on transects using distance sampling techniques (Buckland et al. 2010), a standardized methodology for estimating primate abundance (e.g., Johnson and Overdorff 1999, Irwin et al. 2005, Lehman et al. 2007, Herrera et al. 2011). We surveyed transects repeatedly to detect all species present in the area and gain accurate estimates of encounter rates (diurnal mean = 10 repetitions, range=5-30, nocturnal mean=5, range=2 - 30). We alternated the start position of transects with each repetition to avoid any systematic bias due to the time of the survey. Surveys were conducted during most weather conditions, except if conditions such as rain and fog decreased visibility to less than 25 meters. If weather conditions changed during the survey such that visibility decreased < 25m, the survey was terminated and the reduced survey effort was recorded. We walked slowly along transects in the morning (7:00-12:00hrs) and night (18:00-23:00) looking and listening for lemurs. During nocturnal surveys, we searched the forest with dim headlamps which reflect the ‘eye-shine’ of the lemurs due to the reflective *tapetum lucidum* in their eyes. Once a lemur was detected from eye-shine, we used powerful flashlights to accurately identify species. When lemurs were encountered, we collected the following standard survey data: time and location of sighting on the transect, species, distance to the first lemur sighted (estimated by eye), bearing from the transect (measured with a compass), demographics of encountered groups of lemurs, group spread (furthest distance among individuals in the encountered group), and geographic coordinates with a Garmin 62 hand-held GPS unit (Garmin International, Inc. Olathe KS USA). All distance and group spread measures were estimated by eye, and researchers were trained and tested in distance estimates by comparing to measured distances monthly to avoid drift such that all observers’ estimates were +/- 1m precision at distances < 50m. The distance from the observer to the first animal sighted was converted to the perpendicular distance using the bearing from the transect. If multiple individuals were encountered, I added half the distance of the group spread to the perpendicular distance to account for under-estimating distances because the first individual sighted is most likely at the periphery of the group and thus the distance to the center of the group is half the spread (Buckland et al. 2010). These distances were small, and did not dramatically affect density estimates compared to those made from perpendicular sighting distances alone (results not shown).

References

Blanco, M. B. and V. Rahalinarivo (2010). "First direct evidence of hibernation in an eastern dwarf lemur species (*Cheirogaleus crossleyi*) from the high-altitude forest of Tsinjoarivo, central-eastern Madagascar." Naturwissenschaften 97(10): 945-950.

Dausmann, K. H., J. Glos, J. U. Ganzhorn and G. Heldmaier (2004). "Physiology: hibernation in a tropical primate." Nature 429(6994): 825-826.

**Table S1.** Summary of survey time table and survey efforts. The number of months in the wet and dry seasons, the years sampled, and the total survey effort (# kilometers of transects surveyed) are given for each locality. The latitude and longitude of the base camps for each locality are also given.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Locality** | **Latitude** | **Longitude** | **# dry months sampled** | **# wet months sampled** | **Years sampled** | **Diurnal survey effort (km)** | **Nocturnal survey effort (km)** |
| Ampatsona | -21.009 | 47.397 | 4 | 3 | 2013-2014 | 296.33 | 67.25 |
| Maharira | -21.338 | 47.408 | 0 | 4 | 2012-2013 | 86 | 25.7 |
| Miaranony | -21.148 | 47.533 | 8 | 7 | 2011-2013 | 211.25 | 125.5 |
| Valohoaka | -21.297 | 47.438 | 2 | 8 | 2011-2013 | 209.575 | 146.7 (98.7) |
| Vohiparara | -21.235 | 47.396 | 6 | 2 | 2013 | 198 | 116 (60) |

**Appendix S2 Supplemental Information on habitat sampling**

*Environmental gradients*: I recorded elevation (meters above sea level) on transects every 100m using a GPS unit. Elevation ranged from 674 – 1347 m asl. I gathered data on habitat disturbance every 100m by recording any instances of the following disturbance types: evidence of cattle in the forest, cut trees and tree stumps from logging, evidence of local artisanal gold mining including dams and pools dug in rivers as well as mines on river banks, man-made traps set for animals, forest encampments, agricultural activity, and finally human use and occupation history as told by local residents native to the area, land owners, and forestry records at the Centre ValBio research station (CVB). The human habitat disturbance could be broadly classified into three levels: 0=near pristine: negligible signs of active human resource extraction, no record of past commercial use or human habitation; 1=light disturbance: few signs of active human resource extraction, past history of selective hardwood exploitation and/or agricultural use 30+ years ago or human habitation 70+ years ago; 2= heavy disturbance: active human disturbance including lumbar extraction, agriculture, gold-mining, temporary settlements, past human habitation 50+ years ago.

 I measured the botanical composition and structure on transects using the point-centered-quarter method (Ganzhorn 2011). At 100m intervals along the transects, a point sample was taken. The point was divided into quadrats and within each quadrat, the tree closest to the center of the point was sampled for one tree in each of four size classes based on diameter at breast height (DBH): 0-5cm, 5-10cm, 10-25cm, and 25+cm. If there was no tree in a size class within 10m radius from the point in any quadrat, that size class was recorded as lacking a tree. Trees were identified to the lowest taxonomic level possible by a team of four Malagasy research assistants trained in botany by CVB, the Missouri Botanical Garden, previous international research teams, and Malagasy forestry departments. Trees were identified based on leaf patterns, venation, and color, bark pattern, sap/latex production, and fruit and flower form when available. Trees in the field were recorded by Malagasy name and later translated to taxonomic Family, Genus and, where possible, species based on the extensive botanical database available at CVB. DBH was recorded with a measuring tape, while height, crown height and crown diameter were estimated by eye. Crown height was defined as the height from the first branches of the crown to the top of the tree. Crown diameter was defined as the maximum diameter of the minimum ellipse spanning the branches of the tree. Observers were trained in height and diameter estimation by comparing estimates made by eye to those measured either with a measuring tape or a clinometer until observers were within a precision of +/-1m.

I used principal components analysis (PCA) to reduce the dendrometric variables into a single composite score (the first principal component) to represent forest structure. PCA was performed using the *prcomp* function in the base package of R on scaled variables without subsequent rotation of the axes. The first PC summarized 70% of the variance and was retained for subsequent analysis (Table S2).

 The relative abundance of lemur food trees was calculated by tabulating the number of individual trees in the large size class that were known lemur food resources, based on a review of the literature on the diets of these species at RNP (Table S3). The number of food trees on each transect was divided by the total number of trees (food and non-food trees combined) to obtain the relative frequency of food trees per transect.

**Table S2.** Loadings of original dendrometrics on the first principal component, which summarizes 70% of the variation in the forest structure variables.

|  |  |
| --- | --- |
| **Variable** | **PC1** |
| Diameter at breast height | 0.37 |
| Tree height | 0.53 |
| Crown height | 0.51 |
| Crown diameter | 0.57 |

**Table S3. List of tree species used as food resources by lemurs at RNP.** Tree species identified on study transects that were classified as lemur food trees based on published studies on the diets of most species in the community.

|  |
| --- |
| **Binomial name** |
| *Abrahamia ditimena* |
| *Abrahamia sp* |
| *Abrahamia thouvenotii* |
| *Abrahamia turkii* |
| *Ambavia capuronii* |
| *Canarium sp* |
| *Chrysophyllum boivinianum* |
| *Cryptocarya acuminata* |
| *Cryptocarya ovalifolia* |
| *Cryptocarya parareolata* |
| *Cryptocarya sp* |
| *Cryptocarya sp2* |
| *Cryptocarya sp3* |
| *Cryptocarya thouvenotii* |
| *Erythroxylum sp* |
| *Ficus botryoides* |
| *Ficus pachyclada* |
| *Ficus politoria* |
| *Ficus rubra* |
| *Ficus sp* |
| *Ficus tilifolia* |
| *Haroungana madagascariensis* |
| *Mammea angustifolia* |
| *Mammea bongo* |
| *Mycronychia macrophylla* |
| *Mycronychia madagascariensis* |
| *Noronhia incurvifolius* |
| *Noronhia introversa* |
| *Noronhia ovalifolia* |
| *Noronhia sp* |
| *Noronhia sp1* |
| *Ocotea auriculiformis* |
| *Ocotea grayi* |
| *Ocotea racemosa* |
| *Ocotea sp* |
| *Oncostemum botryoides* |
| *Oncostemum leprosum* |
| *Oncostemum nervosum* |
| *Plagioscyphus louvelli* |
| *Scolopia madagascariensis* |
| *Sideroxylon betsimisarakum* |
| *Streblus dimepate* |
| *Streblus mauritanus* |
| *Syzygium condensata* |
| *Syzygium emirnense* |
| *Syzygium parkeri* |
| *Syzygium sp* |
| *Vitex sp* |
| *Albizia gummifera* |
| *Allophylus cobbe* |
| *Anthocleista amplexicolis* |
| *Beilschmeidia velutina* |
| *Cabucala erytrocarpa* |
| *Canarium madagascariense* |
| *Carissa edulis* |
| *Cinnamosma madagascariensis* |
| *Clerodendrum petunioides* |
| *Dombeya acerifolia* |
| *Dombeya angustipetala* |
| *Dombeya antsihanakensis* |
| *Dombeya hilsembergii* |
| *Dombeya laurifolia* |
| *Dombeya sp* |
| *Garcinia aphanophlebia* |
| *Garcinia ceracifer* |
| *Garcinia goudotiana* |
| *Garcinia megaphylla* |
| *Grewia sp* |
| *Mussaenda erectiloba* |
| *Pauridiantha sp* |
| *Polyscias fraxinifolia* |
| *Polyscias ornifolia* |
| *Polyscias sp* |
| *Psidium cattleianum* |
| *Psychotria mandrarensis* |
| *Psychotria reducta* |
| *Ravenea robustior* |
| *Schefflera longipedicellata* |
| *Schefflera mirianata* |
| *Symphonia gymnoclada* |
| *Treculia africana* |
| *Anthocleista madagascariensis* |
| *Antirhea borbonica* |
| *Aphloia theiformis* |
| *Craspidospermun verticillatum* |
| *Diospyros gricilipes* |
| *Dypsis nodifera* |
| *Dypsis sp* |
| *Ephippiandra madagascariensis* |
| *Eugenia louvelii* |
| *Gaertnera brevipedicellata* |
| *Gaertnera phyllostachya* |
| *Gaertnera sp* |
| *Gyrostipula foveolata* |
| *Melanophylla crenata* |
| *Pandanus sp* |
| *Potameia rubra* |
| *Premna corymbosa* |
| *Psorospermum sp* |
| *Saldinia sp* |
| *Tambourissa perrieri* |
| *Vepris ampody* |
| *Vepris cauliflora* |
| *Vepris sp* |
| *Weinmannia bojeriana* |
| *Weinmannia rutenbergii* |

Tree species identified as food resources for lemurs in: Atsalis 2007, Arrigo-Nelson 2006; Balko 1998; Faulkner and Lehman 2006; Grassi 2001; Razafindratsima et al. 2014; Wright et al. 2011

**Appendix S3 Supplemental Information on the functional traits of species in the study communities**

*Functional traits:* The following functional traits were selected because they capture variation in fundamental niche axes that have been identified as adaptive traits for tolerating comparatively harsh environments (Wright 1999): (1) social group size, (2) % leaves in diet, (3) life-history (age at first reproduction), (4) body mass (ln transformed), (5) activity pattern. Data on these traits were taken from the primary literature and obtained during field work (Table S4).

**Table S4.** Functional traits of lemur species in the regional species pool. % leaves = percentage of leaves in the diet; AFR = age at first reproduction, in years; group size = mean number of individuals observed together in social units either from long-term studies or data collected during this study, ln mass = natural log of body mass, activity = primary activity pattern, nocturnal, diurnal or cathemeral. Data were taken from the literature or collected in this study, as indicated in the table.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Binomial name** | **% leaves** | **AFR** | **group size\*** | **ln mass1** | **Activity16,\*** |
| *Daubentonia madagascariensis* | 02 | 43 | 1.00 | 7.89 | nocturnal |
| *Varecia variegata* | 54 | 1.63 | 2.75 | 8.161 | diurnal |
| *Hapalemur simus* | 905 | 3 | 5.00† | 7.601 | diurnal† |
| *Hapalemur aureus* | 905 | 3 | 2.00 | 7.313 | diurnal |
| *Hapalemur griseus* | 855 | 1.5 | 2.33 | 6.908 | diurnal |
| *Eulemur rufifrons* | 256 | 1.54 | 7.66 | 7.601 | cathemeral |
| *Eulemur rubriventer* | 106 | 1.77 | 3.06 | 7.601 | cathemeral |
| *Lepilemur microdon* | 907 | 1.638 | 1.00 | 6.802 | nocturnal |
| *Microcebus rufus* | 09 | 19 | 1.00 | 3.807 | nocturnal |
| *Cheirogaleus sibreei* | 010 | 2 | 1.14 | 5.49311 | nocturnal |
| *Cheirogaleus crossleyi* | 010 | 212 | 1.14 | 5.90011 | nocturnal |
| *Avahi peyrierasi* | 9513 | 2.588 | 1.65 | 6.908 | nocturnal |
| *Propithecus edwardsi* | 5014 | 415 | 3.62 | 8.412 | diurnal |

\*this study, estimated from survey data

† Note that *H. simus* group size was known from long-term study group (Centre ValBio unpublished data). Also, Tan (2000) suggested *H. simus* was cathemeral based on five nights of nocturnal observations. This species was never observed during nocturnal surveys, however, and was coded as diurnal here. The species was not observed and was excluded from further analyses.

1 Kamilar, J. M., Muldoon, K. M., Lehman, S. M. & Herrera, J. P. (2012). Testing Bergmann's rule and the resource seasonality hypothesis in Malagasy primates using GIS‐based climate data. *Am. J. Phys. Anthropol.* **147**, 401-408.

2 Sterling, E. J. (1993). *Behavioral ecology of the aye-aye (Daubentonia madagascariensis) on Nosy Mangabe, Madagascar* Ph.D. dissertation thesis, Yale University.

3 Zehr, S. M. *et al.* (2014). Life history profiles for 27 strepsirrhine primate taxa generated using captive data from the Duke Lemur Center. *Scientific Data* **1**.

4 Balko, E. A. (1998). *A behaviorally plastic response to forest composition and logging disturbance by Varecia variegata variegata in Ranomafana National Park, Madagascar.* Ph.D. Dissertation thesis, State University of New York.

5 Tan, C. L. (2000). *Behavior and ecology of three sympatric bamboo lemur species (genus Hapalemur) in Ranomafana National Park, Madagascar* Ph.D. Dissertation thesis, State University of New York.

6 Overdorff, D. J. (1991). *Ecological correlates to social structure in two prosimian primates: Eulemur fulvus rufus and Eulemur rubriventer in Madagascar* Ph.D. Dissertation thesis, Duke University.

7 Ganzhorn, J. U. (1992). Leaf chemistry and the biomass of folivorous primates in tropical forests. *Oecologia* **91**, 540-547, doi:10.1007/bf00650329.

8 Godfrey, L. R., Samonds, K. E., Jungers, W. L. & Sutherland, M. R. (2001). Teeth, brains, and primate life histories. *Am. J. Phys. Anthropol.* **114**, 192-214.

9 Atsalis, S. (2007). *A Natural History of the Brown Mouse Lemur*. Pearson Prentice Hall.

10 Fietz, J. & Ganzhorn, J. U. (1999). Feeding ecology of the hibernating primate *Cheirogaleus medius*: how does it get so fat? *Oecologia* **121**, 157-164.

11 Blanco, M. B. *et al.* (2009). Discovery of sympatric dwarf lemur species in the high-altitude rain forest of Tsinjoarivo, eastern Madagascar: implications for biogeography and conservation. *Folia Primatol.* **80**, 1-17.

12 Blanco, M. B. (2010). *Reproductive biology of mouse and dwarf lemurs of eastern Madagascar, with an emphasis on brown mouse lemurs (Microcebus rufus) at Ranomafana National Park, a southeastern rainforest* Ph.D. thesis, University of Massachusettes Amherst.

13 Faulkner, A. L. & Lehman, S. M. (2006). Feeding patterns in a small-bodied nocturnal folivore (*Avahi laniger*) and the influence of leaf chemistry: a preliminary study. *Folia Primatol.* **77**, 218-227.

14 Hemingway, C. A. (1996). Morphology and phenology of seeds and whole fruit eaten by Milne-Edwards’ sifaka, Propithecus diadema edwardsi, in Ranomafana National Park, Madagascar. *Int. J. Primatol.* **17**, 637-659.

15 Pochron, S. T., Tucker, W. T. & Wright, P. C. (2004). Demography, life history, and social structure in Propithecus diadema edwardsi from 1986–2000 in Ranomafana National Park, Madagascar. *Am. J. Phys. Anthropol.* **125**, 61-72, doi:10.1002/ajpa.10266.

16 Donati G, and Borgognini-Tarli SM. (2006). From darkness to daylight: cathemeral activity in primates. Journal of Anthropological Sciences 84(1):7-32.

**Appendix S4 Supplementary Information on data processing**

**Density analyses:**

I used the program DISTANCE to estimate the detection probability from the perpendicular sighting distances. The program DISTANCE allows the user to estimate the width of the survey transect (effective strip width) which, when combined with the survey effort on the transect, is used to calculate the sampled area. To estimate the width of the survey area, I compared the fit of the following functions to the perpendicular distance data: uniform, half-normal, and hazard-rate, each with cosine and simple polynomial terms. The best function was selected using the minimum Akaike information criterion corrected for small samples (AICc). Model adequacy was assessed using quantile-quantile plots and goodness of fit tests as implemented in DISTANCE. To ensure a good fit of the data to the function, if goodness-of-fit tests suggested the data significantly departed from the function, I re-analyzed the data truncating the extreme values (usually truncated at 25-40m), since those extreme values can lead to significant deviations from the model. The models used, detection probabilities and related data are given in Table S5. The mean observed adult group size was used to estimate individual density from cluster density. I estimated a global detection probability and mean cluster sizes from all the data together, and estimated encounter rates and densities stratified by transect to obtain the densities per transect.

**Table S5.** Models of detection probability for each lemur species from DISTANCE analyses used to calculate densities.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lemur species (# observations) | Model of detection probability | Detection probability within max. width | Effective strip width (in meters) | Truncation distance used (in meters) |
| *Avahi peyrierasi* (121) | Half-normal w/ cosine adjustments (2) | 0.38 | 11.49 | 30 |
| *Propithecus edwardsi* (111) | Uniform w/ cosine adjustment (1) | 0.55 | 24.71 | 45 |
| *Eulemur rufifrons* (79) | Half-normal w/ simple polynomial adjustments (4) | 0.85 | 25.49 | 30 |
| *Varecia variegata* (14) | Uniform  | 1.0 | 35 | 35 |
| *Eulemur rubriventer* (94) | Hazard rate | 0.44 | 17.46 | 40 |
| *Hapalemur griseus* (31) | Uniform w/ cosine adjustment (1) | 0.55 | 16.36 | 30 |
| *Lepilemur microdon* (37) | Uniform | 1.0 | 13 | 13 |
| *Cheirogaleus* sp (196) | Hazard rate | 0.49 | 13.68 | - |
| *Microcebus rufus* (398) | Hazard rate | 0.41 | 8.27 | 20 |

**Phylogeny:** I used the most complete primate phylogeny inferred to date (Springer *et al*. 2012). The tree was based on a total evidence analysis of 69 nuclear and 10 mitochondrial genes for nearly all extant primates. Of the trees made available in the supplement to Springer *et al.* (2012), I selected tree # 4, for which divergence times were inferred using independent rates and soft-bounded constraints (see original reference for details). I rescaled the tree to a total depth of 1.0 using the *rescale* function in the R package *ape* (Paradis *et al.* 2004).

**Figure S1.** Phylogenetic tree from Springer *et al.* (2012) used in the analyses. Note that the total tree depth was scaled to 1.



**Phylogenetic tree in Newick format**

((((Microcebus\_rufus:0.477829953,(Cheirogaleus\_crossleyi:0.3486706697,Cheirogaleus\_sibreei:0.348670589):0.1291592429):0.180240646,Lepilemur\_microdon:0.6580705182):0.06703679958,((Avahi\_peyrierasi:0.4129366558,Propithecus\_edwardsi:0.4129367365):0.2576337139,(((Eulemur\_rufifrons:0.156613371,Eulemur\_rubriventer:0.156613371):0.2701756174,Hapalemur\_griseus:0.4267890288):0.08518450714,Varecia\_variegata:0.5119734351):0.1585968943):0.0545369279):0.2748926015,Daubentonia\_madagascariensis:0.9999998588);

**Phylogenetic structure analyses:**

The following annotated R code gives the analyses of phylogenetic community structure performed in this study.

library(pez); library(cluster); library(picante); library(plotrix); library(dplyr); library(ape)

##############################################################################

##MPD on comm, a communityXspecies matrix, and tree, a phylogenetic tree

##############################################################################

#make phylogenetic distance matrix

phydist=cophenetic(tree)

#conduct MPD analysis

MPD\_fifteen=ses.mpd(comm, phydist, null.model="taxa.labels", abundance.weighted=FALSE, runs=999)

##change abundance.weighted argument to TRUE if data matrix consists of abundances

MPD\_fifteen

#conduct MNTD analysis

MNTD\_fifteen=ses.mntd(comm, phydist, null.model="taxa.labels", abundance.weighted=FALSE, runs=999)

MNTD\_fifteen

#same with functional trait distance matrix; create matrix from a dataframe of traits, #traits\_data.frame

traitdist<-as.matrix(daisy(traits\_data.frame, metric="gower"))

MNTD\_traits=ses.mntd(comm, traitdist, null.model="taxa.labels", abundance.weighted=FALSE, runs=999)

MNTD\_traits

###########################################################################

###############testing effects of functional MNTD / environmental factors on phylo MNTD

###########################################################################

source("r2lmm.R") ##for calculating R^2 with mixed models, see ##http://jonlefcheck.net/2013/03/13/r2-for-linear-mixed-effects-models/

##null model, intercept only with site as a random grouping variable

mntd\_null<-lme(fixed=phymntdz~1, data=dat, random=~1|site, method="ML")

mntd\_null

summary(mntd\_null)

r2<-r.squared.lme(mntd\_null)

r2 #marginal = 0, conditional = 0.646

AIC(mntd\_null2)#86.14

#To test for effects of spatial autocorrelation, I included a distance matrix based on the ratio #geographic correlation matrix calculated using the R package *geoR* (Ribeiro & Diggle 2015). I #ran each linear mixed model (lmm) of predictor variables explaining variation in phylogenetic #MNTD and then tested if including the spatial autocorrelation matrix improved the fit of the #model to the date using the *update* function. For example:

###R code for testing effects of spatial autocorrelation

library(nlme); library(geoR)

null.model<-lme(fixed=mntd~1, data=data, random=~1|site)

ratio.geo<-update(null.model, correlation = corRatio(0.01, form= ~ east + south), method="ML") #east and south are variables in the dataset for latitude and longitude

anova(null.model, ratio.geo)

#Spatial autocorrelation was included in the final models if the model including spatial #autocorrelation had a significantly better fit to the data than the model with no spatial effect.

#alternatively, spatial effects can be included directly in the model specification without using #the function *update*:

mntd\_null\_geo<-lme(fixed=phymntdz~1, data=dat, correlation=corRatio(0.1, form=~east+south|site),random=~1|site, method="ML")

mntd\_null\_geo

summary(mntd\_null\_geo)

r2<-r.squared.lme(mntd\_null\_geo)

r2 #marginal = 0, conditional = 0.625189

AIC(mntd\_null\_geo)#87.94

anova(mntd\_null, mntd\_null\_geo)

#####now to test the effects of independent environmental variables

##proportion of food trees on transects (food), shannon-weinner diversity index (shannon\_all) #and elevation (elev)

mntd\_food\_shannon\_elevation\_mixed<-lme(fixed=phymntdz~scale(food)+scale(shannon\_all)+scale(elev), data=dat, random = ~ 1 | site, method="ML")

summary(mntd\_food\_shannon\_elevation\_mixed)

r2<-r.squared.lme(mntd\_food\_shannon\_elevation\_mixed)

r2 #marginal = 0.60, conditional = 0.62, AIC=85.66

anova(mntd\_null, mntd\_food\_shannon\_elevation\_mixed)

##########################

##assumption checks

############################

#check normality of the residuals:

hist(residuals(mntd\_food\_shannon\_elevation\_mixed))#should look like a normal distribution (or #at least symmetric)

#(particularly worry about outliers, a skewed distribution)

#better option to check for normality

qqnorm(residuals(mntd\_food\_shannon\_elevation\_mixed))

qqline(residuals(mntd\_food\_shannon\_elevation\_mixed))#all points should fall on the straight #line

#and the most important about the residuals, homoscedasticity

plot(fitted(mntd\_food\_shannon\_elevation\_mixed), residuals(mntd\_food\_shannon\_elevation\_mixed))#no pattern should be visible here

# compare sample size and number of parameters:

length(residuals(mntd\_food\_shannon\_elevation\_mixed))

length(coefficients(mntd\_food\_shannon\_elevation\_mixed))#this should be small as compared to #the sample size

library(car)

# Assessing Outliers

mntd\_traits\_lm<-lm(phymntdz~traitmntdz, data=dat)

summary(mntd\_traits\_lm)

outlierTest(mntd\_traits\_lm) # Bonferonni p-value for most extreme obs

qqPlot(mntd\_traits\_lm, main="QQ Plot") #qq plot for studentized resid

leveragePlots(mntd\_traits\_lm) # leverage plots

# Influential Observations

# Cook's D plot

# identify D values > 4/(n-k-1)

cutoff <- 4/((nrow(dat)-length(mntd\_traits\_lm$coefficients)-2))

plot(mntd\_traits\_lm, which=4, cook.levels=cutoff)

# take a look at the rows that were suggested to be influential

dat[27,]

dat[10,]

dat[16,]

# Influence Plot

influencePlot(mntd\_traits\_lm, id.method="identify", main="Influence Plot", sub="Circle size is proportial to Cook's Distance" )

# Global test of model assumptions

library(gvlma)

gvmodel <- gvlma(mntd\_traits\_lm)

summary(gvmodel)

plot(gvmodel)

##########################

###model averaging

##########################

###code adapted from (Garamszegi & Mundry 2014)

library(AICcmodavg)

###create a list of all the model objects

cand.set<-list()

cand.set[[1]]<-mntd\_elev\_mixed

cand.set[[2]]<-mntd\_shannon\_all\_mixed

cand.set[[3]]<-mntd\_elev\_shannon\_all\_mixed

cand.set[[4]]<-mntd\_str\_mixed

cand.set[[5]]<-mntd\_dist\_mixed

cand.set[[6]]<-mntd\_habitat\_mixed

cand.set[[7]]<-mntd\_food\_mixed

cand.set[[8]]<-mntd\_food\_shannon\_elevation\_mixed

cand.set[[9]]<-mntd\_food\_elevation\_mixed

cand.set[[10]]<-mntd\_food\_shannon\_mixed

Modnames = paste(c("elevation", "diversity", "elev+div", "structure", "disturbance",

 "habitat", "food", "food+","food+elev", "food+div"),

 sep = " ")

aic.table<-aictab(cand.set=cand.set, modnames=Modnames, sort=T)

write.table(aic.table, file="aic.table.txt", sep="\t")

modavg(cand.set=cand.set, parm="scale(lemur\_foods)", modnames=Modnames,

 conf.level = 0.95,second.ord = TRUE, nobs = NULL, exclude = NULL,

 warn = TRUE, uncond.se = "revised")

modavg(cand.set=cand.set, parm="scale(shannon\_all)", modnames=Modnames,

 conf.level = 0.95,second.ord = TRUE, nobs = NULL, exclude = NULL,

 warn = TRUE, uncond.se = "revised")

modavg(cand.set=cand.set, parm="elevation.z", modnames=Modnames,

 conf.level = 0.95,second.ord = TRUE, nobs = NULL, exclude = NULL,

 warn = TRUE, uncond.se = "revised")

modavg(cand.set=cand.set, parm="as.factor(dist)1", modnames=Modnames,

 conf.level = 0.95,second.ord = TRUE, nobs = NULL, exclude = NULL,

 warn = TRUE, uncond.se = "revised")

modavg(cand.set=cand.set, parm="as.factor(dist)2", modnames=Modnames,

 conf.level = 0.95,second.ord = TRUE, nobs = NULL, exclude = NULL,

 warn = TRUE, uncond.se = "revised")

modavg(cand.set=cand.set, parm="scale(PC1)", modnames=Modnames,

 conf.level = 0.95,second.ord = TRUE, nobs = NULL, exclude = NULL,

 warn = TRUE, uncond.se = "revised")

##############################################################################

##PGLMM

##############################################################################

nspp<-15

nsite<-121

Vphy <- vcv(tree)

# standardize the phylogenetic covariance matrix to have determinant 1

Vphy <- Vphy/(det(Vphy)^(1/nspp))

##make the phylogenetic dissimilarity matrix as the inverse of the phylo similarity matrix

Vrepul<-solve(Vphy, diag(nspp))

Vrepul<-Vrepul/max(Vrepul)

Vrepul<-1\*Vrepul

#make Y the communityXspecies matrix, the ‘comm’ object

Y<-comm

# name the simulated species 1:nspp and sites 1:nsites

colnames(Y) <- 1:nspp

rownames(Y) <- 1:nsite

#turn the matrix into ‘long’ format

Y<-t(Y)

Y <- matrix(Y, nrow = nspp, ncol = nsite)

YY <- matrix(Y, nrow = nspp \* nsite, ncol = 1)

site <- matrix(kronecker(1:nsite, matrix(1, nrow = nspp, ncol =

 1)), nrow = nspp \* nsite, ncol = 1)

sp <- matrix(kronecker(matrix(1, nrow = nsite, ncol = 1), 1:nspp),

 nrow = nspp \* nsite, ncol = 1)

#make the new PGLMM dataframe, dat

dat <- data.frame(Y = YY, site = as.factor(site), sp = as.factor(sp))

##specify the random effects

# random intercept with species independent; i.e., species-specific effects

re.1 <- list(1, sp = dat$sp, covar = diag(nspp))

# random intercept with overall phylogenetic covariances

re.2 <- list(1, sp = dat$sp, covar = Vphy)

#nested phylogenetic attraction, phylo clustering in each community/transect

re.sp.site <- list(1, sp = dat$sp, covar = Vphy, site = dat$site) # note: nested

# random effect for site

re.site <- list(1, site = dat$site, covar = diag(nsite))

#run PGLMM on presence/absence; if the data are abundances, change family = “Gaussian”

##to test the effect of fixed factors, include them in the model as Y ~ X1\*X2, as in standard #linear model analysis

z.binary\_attraction\_15 <- communityPGLMM(Y ~ 1, data = dat, family = "binomial",

 sp = dat$sp, site = dat$site,

 random.effects = list(re.1, re.2, re.sp.site, re.site), REML = TRUE, verbose = FALSE)

#str(z.binary\_attraction\_15)

z.binary\_attraction\_15$s2r#random effect nonnested variable variance scalars, sp effect, phylo #overdispersion effect, site effect

z.binary\_attraction\_15$s2n#random effect nested scalars

##for binomial regressions, use likelihood ratio tests for the effect of each random effect factor on the presence/absence #matrix; results save the probability of no effect

pr1<-communityPGLMM.binary.LRT(z.binary\_attraction\_15, re.number = 1)$Pr #species effect

pr2<-communityPGLMM.binary.LRT(z.binary\_attraction\_15, re.number = 2)$Pr #phylo effect

pr3<-communityPGLMM.binary.LRT(z.binary\_attraction\_15, re.number = 3)$Pr #nested phylo #effect

pr4<-communityPGLMM.binary.LRT(z.binary\_attraction\_15, re.number = 4)$Pr #site effect

#examine distribution of residuals

hist(as.numeric(z.binary\_attraction\_15$H))

##set up PGLMM to test for phylogenetic overdispersion (or ‘repulsion’)

# random intercept with species independent

re.1 <- list(1, sp = dat$sp, covar = diag(nspp))

# random intercept with species showing phylogenetic dissimilarity

re.2 <- list(1, sp = dat$sp, covar = Vrepul)

#nested phylogenetic attraction

re.sp.site <- list(1, sp = dat$sp, covar = Vrepul, site = dat$site) # note: nested

# random effect for site

re.site <- list(1, site = dat$site, covar = diag(nsite))

z.binary\_fifteen\_repulse <- communityPGLMM(Y ~ 1, data = dat, family = "binomial",

 sp = dat$sp, site = dat$site,

 random.effects = list(re.1, re.2, re.sp.site, re.site), REML = TRUE, verbose = FALSE)

#str(z.binary\_fifteen\_repulse)

z.binary\_fifteen\_repulse$s2r#random effect nonnested variable variance scalars: sp effect, phylo #overdispersion effect, site effect

z.binary\_fifteen\_repulse$s2n#random effect nested phylo repulsion scalars

##likelihood ratio tests of random effects; only works with binomial regressions

communityPGLMM.binary.LRT(z.binary\_fifteen\_repulse, re.number = 1)$Pr #species effect

communityPGLMM.binary.LRT(z.binary\_fifteen\_repulse, re.number = 2)$Pr #phylo effect

communityPGLMM.binary.LRT(z.binary\_fifteen\_repulse, re.number = 3)$Pr #nested phylo #effect

communityPGLMM.binary.LRT(z.binary\_fifteen\_repulse, re.number = 4)$Pr #site effect

##examine residuals

hist(as.numeric(z.binary\_fifteen\_repulse$H))

#################################################################################ecological-evolutionary quantile regressions from Cavender-Bares *et al*. (2004) using pez

##############################################################################

rownames(comm)<-1:121

##these functions in pez use a comparative data object

comp.dat <-comparative.comm(tree, comm)

##conduct the ecological co-existence – phylogenetic distance regression analysis using quantile #regression

eco.phy.reg\_15 <-eco.phy.regression(comp.dat, randomisation="richness", permute=1000,method = "quantile", tau = c(0.1,0.25, 0.5,0.75,0.9))

summary(eco.phy.reg\_15\_attract)

#eco-phy regression with linear model approach

eco.phy.reg\_lm<-eco.phy.regression(comp.dat, randomisation="richness", permute=1000,method = "lm")

summary(eco.phy.reg \_lm)

eco.matrix <- as.dist(1 - as.matrix(comm.dist(comp.dat$comm)))

phy.matrix <- as.dist(cophenetic(comp.dat$phy))

#plot the two matrices with quantile regression slopes and intercepts

plot(eco.matrix~phy.matrix)

abline(0.61,-0.305, lwd=2)##replace 0.61, -0.305 with quantile regression intercept and slope, #respectively

abline(0.79,-0.367, lwd=2, col="blue")

abline(0.93,-0.346, lwd=2, col="purple")

##in these analyses, if phylogenetic clustering is present, the slope of the eco-phy regression #should be negative – the more closely related species should co-occur more frequently than the #distantly related species; if phylogenetic overdispersion is present, the slope should be positive #– higher co-existence among species that are more distantly related to each other

**Appendix S5 Supplementary results**

Lemur species richness and density per transect.

The two bamboo lemur species not observed, *Hapalemur aureus* and *H. simus*, are known to occur at approximately three localities within RNP; their abundance in other areas of the park, however, is unknown. I did not observe the animals during our surveys or their characteristic feeding signs, and thus they were recorded as absent. *Daubentonia* was observed once on a single transect during the surveys, but characteristic feeding signs (bore marks in dead wood) were found on every transect and thus we inferred them to be present but with low abundance. I use a value of 0.001 individuals / km2 for *Daubetonia* at all transects to reflect their presence but low relative abundance. Further, *Cheirogaleus* species are difficult to distinguish from survey data alone. A concomitant study on *Cheirogaleus* using trapping data suggested that at the two localities where *C. sibreei* was present, it was sympatric with *C. crossleyi* and the capture frequency of each species was roughly equal, suggesting the two species were approximately equal in abundance (unpub. personal data). Thus, I inferred the individual density of *Cheirogaleus* from the transect survey data and, at the two localities where the species are sympatric, assigned half the total *Cheirogaleus* density to each species. Finally, on three transects (VOF, VOG, and VOH) nocturnal survey effort was too low (two repetitions) to reliably estimate abundance. I imputed the mean density for the locality at which the transect was located to fill missing values of each nocturnal species observed during the two surveys. Thus, the presence of the species was confirmed during surveys, but the abundance was imputed.

 Neither the Shannon diversity index nor the relative frequency of lemur foods varied significantly in relation to elevation or habitat disturbance (lmm, fixed factors = elevation + disturbance, random factor = locality, likelihood ratio test (LRT) of fixed effects vs. intercept-only model, relative frequency of food trees: χ2= 0.28, p=0.96, Shannon diversity: χ2=5.22, p=0.16).

 Below I present the mean density estimate for each lemur species (Table S6), the calculated MNTD for each community based on phylogenetic and functional distance matrices (Table S7 and S8, respectively), the full results of the linear mixed models testing for effects of environmental variables and functional traits on MNTD (Table S9), the full results of the PGLMMs testing for environmental, phylogenetic, and functional trait effects on species abundances (Table S10), and the results of the linear model predicting ecological co-existence from phylogenetic distance (Table S11).

**Table S6**. Mean estimated individual density (ind/km2) of each lemur species on each transect. *D. mada.ensis* = *Daubentonia madagascariensis*. Each row is a transect, coded such that the first letter(s) indicate the locality, and the subsequent letter indicates the transect. A=Ampatsona, M=Maharira, MI=Miaranony, V=Valohoaka, VO=Vohiparara.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Transect | *D. mada.ensis* | *V. variegata* | *H. griseus* | *E. rufifrons* | *E. rubriventer* | *L. microdon* | *M. rufus* | *C. sibreei* | *C. crossleyi* | *A. peyrierasi* | *P. edwardsi* |
| AA | 0.01 | 0.00 | 1.70 | 0.00 | 8.04 | 6.25 | 42.64 | 2.07 | 2.07 | 34.26 | 11.81 |
| AB | 0.01 | 0.00 | 0.00 | 0.00 | 1.91 | 7.37 | 73.53 | 9.77 | 9.77 | 44.47 | 9.51 |
| AC | 0.01 | 0.00 | 1.12 | 0.00 | 3.99 | 12.53 | 78.91 | 12.45 | 12.45 | 20.61 | 6.01 |
| AD | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 4.17 | 13.12 | 0.00 | 20.71 | 34.26 | 4.72 |
| AE | 0.01 | 0.00 | 2.21 | 0.00 | 0.00 | 10.53 | 56.58 | 0.00 | 16.01 | 35.67 | 7.08 |
| AF | 0.01 | 0.00 | 3.68 | 7.08 | 10.88 | 10.53 | 56.58 | 0.00 | 16.01 | 35.67 | 7.87 |
| AG | 0.01 | 0.00 | 0.00 | 0.00 | 4.48 | 35.71 | 74.97 | 8.87 | 8.87 | 48.95 | 16.19 |
| MA | 0.01 | 0.00 | 1.53 | 0.00 | 8.83 | 3.56 | 84.10 | 7.08 | 7.08 | 23.43 | 0.00 |
| MB | 0.01 | 0.00 | 2.61 | 0.00 | 5.03 | 25.00 | 59.04 | 2.07 | 2.07 | 6.85 | 0.00 |
| MC | 0.01 | 0.00 | 5.35 | 0.00 | 0.00 | 10.42 | 32.80 | 0.00 | 0.00 | 51.40 | 0.00 |
| MIA | 0.01 | 0.00 | 1.95 | 37.76 | 2.51 | 0.00 | 30.07 | 0.00 | 39.69 | 5.71 | 5.86 |
| MIB | 0.01 | 0.00 | 0.00 | 76.10 | 5.30 | 0.00 | 18.22 | 0.00 | 57.52 | 3.81 | 8.84 |
| MIC | 0.01 | 0.00 | 4.84 | 14.18 | 10.37 | 0.00 | 20.87 | 0.00 | 33.88 | 6.23 | 3.46 |
| MID | 0.01 | 0.00 | 0.00 | 9.32 | 6.82 | 4.63 | 10.93 | 0.00 | 25.31 | 3.81 | 1.89 |
| MIE | 0.01 | 0.00 | 2.18 | 15.35 | 25.26 | 0.00 | 49.20 | 0.00 | 46.59 | 14.28 | 10.92 |
| MIF | 0.01 | 0.00 | 5.07 | 0.00 | 16.31 | 0.00 | 13.46 | 0.00 | 14.87 | 21.09 | 14.51 |
| VA | 0.01 | 0.00 | 0.00 | 15.03 | 5.41 | 0.00 | 132.02 | 0.00 | 30.67 | 22.17 | 0.00 |
| VB | 0.01 | 1.60 | 0.00 | 16.97 | 0.00 | 0.00 | 34.99 | 0.00 | 18.41 | 35.53 | 13.64 |
| VC | 0.01 | 3.15 | 1.42 | 19.50 | 4.51 | 1.75 | 98.53 | 0.00 | 13.65 | 49.05 | 10.07 |
| VD | 0.01 | 3.33 | 2.39 | 21.17 | 6.35 | 0.00 | 59.56 | 0.00 | 30.96 | 25.66 | 5.67 |
| VE | 0.01 | 0.81 | 2.89 | 14.24 | 12.31 | 0.00 | 51.70 | 0.00 | 21.76 | 5.90 | 6.86 |
| VF | 0.01 | 1.28 | 0.00 | 9.02 | 9.74 | 1.96 | 72.27 | 0.00 | 12.55 | 16.09 | 10.87 |
| VG | 0.01 | 2.78 | 2.50 | 9.84 | 5.31 | 3.58 | 21.87 | 0.00 | 21.46 | 29.41 | 14.82 |
| VOA | 0.01 | 0.00 | 4.94 | 0.00 | 7.96 | 4.17 | 90.20 | 0.00 | 5.18 | 6.85 | 1.60 |
| VOB | 0.01 | 0.00 | 0.00 | 2.29 | 11.28 | 5.21 | 104.96 | 0.00 | 4.14 | 0.00 | 5.45 |
| VOC | 0.01 | 0.00 | 0.00 | 0.00 | 4.51 | 1.49 | 98.40 | 0.00 | 3.45 | 0.00 | 5.45 |
| VOD | 0.01 | 0.00 | 7.41 | 0.00 | 7.96 | 0.00 | 103.08 | 0.00 | 0.00 | 6.85 | 6.41 |
| VOE | 0.01 | 0.00 | 4.94 | 0.10 | 11.94 | 0.00 | 87.46 | 0.00 | 5.18 | 4.89 | 3.21 |
| VOF | 0.01 | 0.00 | 10.50 | 0.00 | 0.00 | 2.51 | 75.21 | 0.00 | 3.45 | 3.54 | 0.00 |
| VOG | 0.01 | 0.00 | 0.00 | 4.90 | 0.00 | 2.51 | 75.21 | 0.00 | 3.45 | 3.54 | 0.00 |
| VOH | 0.01 | 0.00 | 0.00 | 9.81 | 4.83 | 2.51 | 75.21 | 0.00 | 3.45 | 3.54 | 0.00 |

**Table S7.** Results of phylogenetic community structure analysis comparing the observed mean nearest taxon distance (MNTD) to that expected under a null model that maintains species prevalence and randomizes abundances. P values < 0.05 indicate phylogenetic clustering, highlighted in yellow, and > 0.95 are overdispersed, highlighted in green. Localities and transects labelled as in Table S5.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Transect** | **N taxa** | **Observed MNTD** | **Null MNTD mean** | **Null MNTD SD** | **MNTD effect size** | **p** |
| AA | 9 | 0.907 | 0.922 | 0.108 | -0.137 | 0.430 |
| AB | 8 | 0.907 | 0.917 | 0.114 | -0.087 | 0.440 |
| AC | 9 | 0.919 | 0.923 | 0.117 | -0.033 | 0.488 |
| AD | 6 | 0.909 | 0.914 | 0.115 | -0.043 | 0.461 |
| AE | 6 | 0.943 | 0.920 | 0.114 | 0.199 | 0.632 |
| AF | 9 | 0.860 | 0.924 | 0.112 | -0.573 | 0.239 |
| AG | 8 | 0.960 | 0.916 | 0.119 | 0.373 | 0.721 |
| MA | 8 | 0.993 | 0.912 | 0.113 | 0.718 | 0.813 |
| MB | 8 | 1.041 | 0.920 | 0.112 | 1.087 | 0.892 |
| MC | 5 | 1.330 | 0.912 | 0.105 | 3.969 | 0.997 |
| MIA | 8 | 0.739 | 0.916 | 0.117 | -1.513 | 0.070 |
| MIB | 7 | 0.640 | 0.907 | 0.106 | -2.525 | 0.010 |
| MIC | 8 | 0.740 | 0.910 | 0.107 | -1.589 | 0.063 |
| MID | 8 | 0.817 | 0.925 | 0.107 | -1.010 | 0.132 |
| MIE | 8 | 0.797 | 0.923 | 0.112 | -1.128 | 0.107 |
| MIF | 7 | 0.873 | 0.926 | 0.112 | -0.472 | 0.267 |
| VA | 6 | 0.923 | 0.915 | 0.103 | 0.075 | 0.527 |
| VB | 7 | 0.911 | 0.927 | 0.111 | -0.145 | 0.438 |
| VC | 10 | 0.851 | 0.920 | 0.113 | -0.609 | 0.217 |
| VD | 9 | 0.833 | 0.926 | 0.106 | -0.879 | 0.152 |
| VE | 9 | 0.792 | 0.923 | 0.111 | -1.186 | 0.103 |
| VF | 9 | 0.857 | 0.927 | 0.112 | -0.615 | 0.220 |
| VG | 10 | 0.841 | 0.930 | 0.108 | -0.825 | 0.146 |
| VOA | 8 | 0.947 | 0.929 | 0.114 | 0.157 | 0.631 |
| VOB | 7 | 0.931 | 0.929 | 0.107 | 0.015 | 0.508 |
| VOC | 6 | 0.986 | 0.929 | 0.114 | 0.494 | 0.784 |
| VOD | 6 | 1.336 | 0.926 | 0.110 | 3.722 | 0.997 |
| VOE | 8 | 0.865 | 0.924 | 0.116 | -0.509 | 0.265 |
| VOF | 6 | 0.984 | 0.928 | 0.111 | 0.498 | 0.784 |
| VOG | 6 | 0.986 | 0.919 | 0.108 | 0.621 | 0.821 |
| VOH | 7 | 0.913 | 0.923 | 0.114 | -0.088 | 0.444 |

**Table S8.** Results of functional trait community structure analysis comparing the observed mean nearest taxon distance (MNTD) to that expected under a null model that maintains species prevalence and randomizes abundances. P values < 0.05 indicate phylogenetic clustering, highlighted in yellow, and > 0.95 are overdispersed, highlighted in green. Localities and transects labelled as in Table S5.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Transect** | **N taxa** | **Observed MNTD** | **Null MNTD mean** | **Null MNTD SD** | **MNTD effect size** | **p** |
| AA | 9 | 0.159 | 0.180 | 0.044 | -0.481 | 0.312 |
| AB | 8 | 0.129 | 0.183 | 0.047 | -1.145 | 0.068 |
| AC | 9 | 0.133 | 0.180 | 0.043 | -1.093 | 0.080 |
| AD | 6 | 0.152 | 0.181 | 0.045 | -0.647 | 0.219 |
| AE | 6 | 0.141 | 0.180 | 0.044 | -0.904 | 0.118 |
| AF | 9 | 0.159 | 0.183 | 0.047 | -0.529 | 0.268 |
| AG | 8 | 0.140 | 0.186 | 0.047 | -0.988 | 0.104 |
| MA | 8 | 0.138 | 0.186 | 0.047 | -1.028 | 0.074 |
| MB | 8 | 0.140 | 0.187 | 0.049 | -0.961 | 0.087 |
| MC | 5 | 0.204 | 0.185 | 0.046 | 0.406 | 0.785 |
| MIA | 8 | 0.187 | 0.185 | 0.045 | 0.036 | 0.628 |
| MIB | 7 | 0.190 | 0.183 | 0.043 | 0.174 | 0.677 |
| MIC | 8 | 0.192 | 0.184 | 0.044 | 0.186 | 0.680 |
| MID | 8 | 0.164 | 0.185 | 0.044 | -0.475 | 0.316 |
| MIE | 8 | 0.193 | 0.184 | 0.045 | 0.194 | 0.709 |
| MIF | 7 | 0.283 | 0.184 | 0.044 | 2.251 | 0.959 |
| VA | 6 | 0.180 | 0.183 | 0.045 | -0.074 | 0.593 |
| VB | 7 | 0.247 | 0.185 | 0.042 | 1.474 | 0.934 |
| VC | 10 | 0.155 | 0.186 | 0.047 | -0.667 | 0.204 |
| VD | 9 | 0.198 | 0.186 | 0.048 | 0.258 | 0.744 |
| VE | 9 | 0.183 | 0.184 | 0.045 | -0.015 | 0.634 |
| VF | 9 | 0.171 | 0.184 | 0.047 | -0.269 | 0.441 |
| VG | 10 | 0.173 | 0.185 | 0.044 | -0.280 | 0.458 |
| VOA | 8 | 0.178 | 0.185 | 0.049 | -0.134 | 0.553 |
| VOB | 7 | 0.175 | 0.184 | 0.044 | -0.207 | 0.495 |
| VOC | 6 | 0.178 | 0.187 | 0.049 | -0.178 | 0.542 |
| VOD | 6 | 0.372 | 0.182 | 0.044 | 4.355 | 0.992 |
| VOE | 8 | 0.180 | 0.180 | 0.042 | 0.015 | 0.607 |
| VOF | 6 | 0.162 | 0.185 | 0.047 | -0.505 | 0.306 |
| VOG | 6 | 0.171 | 0.182 | 0.043 | -0.246 | 0.479 |
| VOH | 7 | 0.161 | 0.184 | 0.045 | -0.493 | 0.306 |

**Table S9.** Full results of the linear mixed models testing the effects of environmental variables on phylogenetic community structure (MNTD).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Hypothesis** | **Factors** | **DF** | **Fixed factor coefficient** | **Fixed factor p** | **Model p (against null)** | **Model p (against simpler model)** | **Model R2** | **Model AICc** | **ΔAICc** |
| Relative food tree frequency + elevation | Fixed: food tree | 24 | -0.55 | 0.00 | 0.00 | 0.0025 | 0.58 | 87.62 | 0.00 |
|   | Fixed: elevation | 24 | 0.68 | 0.00 |   |   |   |   |   |
|   | Random: Locality | 4 |   |   |   |   |   |   |   |
| Relative food tree frequency + tree diversity + elevation | Fixed: food tree | 23 | -0.46 | 0.01 | 0.00 | 0.21 | 0.60 | 89.16 | 1.53 |
|   | Fixed: tree diversity | 23 | -0.21 | 0.25 |   |   |   |   |   |
|   | Fixed: elevation | 23 | 0.64 | 0.00 |   |   |   |   |   |
|   | Random: Locality | 4 |   |   |   |   |   |   |   |
| Elevation | Fixed: elevation | 25 | 0.72 | 0.00 | 0.01 | - | 0.35 | 95.46 | 7.84 |
|   | Random: Locality | 4 |   | - |   |   |   |   |   |
| Tree diversity | Fixed: tree diversity | 25 | -0.37 | 0.07 | 0.07 | - | 0.09 | 98.87 | 11.24 |
|   | Random: Locality | 4 |   | - |   |   |   |   |   |
| Tree diversity + elevation | Fixed: tree diversity | 28 | -0.41 | 0.03 | 0.00 | 0.003 | 0.48 | 92.93 | 5.30 |
|   | Fixed: elevation | 28 | 0.43 | 0.02 |   |   |   |   |   |
|   | Random: Locality | 4 |   |   |   |   |   |   |   |
| Forest structure | Fixed: forest structure | 29 | -0.22 | 0.41 | 0.40 | - | 0.03 | 101.39 | 13.77 |
|   | Random: Locality | 4 |   | - |   |   |   |   |   |
| Habitat disturbance | Fixed: disturbance1 | 28 | -0.27 | 0.72 | 0.92 | - | 0.01 | 104.81 | 17.18 |
|   | Fixed: disturbance2 | 28 | -0.20 | 0.81 |   |   |   |   |   |
|   | Random: Locality | 4 |   |   |   |   |   |   |   |
| Relative food tree frequency | Fixed: food tree | 25 | -0.44 | 0.03 | 0.04 | - | 0.14 | 97.66 | 10.03 |
|   | Random: Locality | 4 |   |   |   |   |   |   |   |
| Spatial autocorrelation | Fixed: -  | 30 | - | - |   |   |   | 100.14 | 14.92 |
|   | Random: -  |   |   | - |   |   |   |   |   |
|   | Spatial: Ratio |   |   | - |   |   | <0.001 |   |   |

**Table S10.** Full results of the phylogenetic generalized linear mixed models testing the effects of environmental, phylogenetic, and trait variables on the abundance of species in the communities. See separate Excel file.

**Table S11.** Results of the linear model predicting ecological co-existence from phylogenetic distance (*eco.phy.regression* function in *pez*). The positive slope indicates phylogenetic overdispersion. Degrees of freedom for this regression are 1 and 53.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Parameter estimate** | **SE** | ***t*-value** | **p** |
| Intercept | 0.26 | 0.15 | 1.79 | 0.08 |
| Phylogenetic distance | 0.17 | 0.10 | 1.66 | 0.10 |

**Appendix S6 Supplementary Information on simulation experiments**

I designed a simulation experiment to test the power of commonly-used methods for detecting significant phylogenetic structure in ecological communities. I simulated trees and communities with species-pools of 5, 10, 15, 25, 50, and 100 species to determine if significant phylogenetic clustering or overdispersion could be detected using the three approaches in this study: deviations of the observed mean nearest taxon distance (MNTD) from null expectations, ecological co-existence – phylogenetic distance regressions, and phylogenetic generalized mixed model (PGLMM) regressions.

 I simulated pure-birth trees using the *pbtree* function in the R package *phytools* (Revell 2015)*.* I used the function *scape* in the R package *pez* (Pearse et al. 2015) to simulate communities with phylogenetic clustering and overdispersion. The arguments for this function were set such that the observed communities consisted of 30-40% of the total species pool, which is the size suggested to be the most powerful from previous simulation studies (Kraft *et al.* 2007). The argument ‘scape.size’ was set to 10, which produces 121 communities. To simulate clustered communities, the argument ‘g.center’, which is the strength of clustering, was set to 100, ‘g.repulse’ or the strength of phylogenetic overdispersion was set to 1, ‘wd.all’ or niche width was set to 150 which simulates large range sizes, ‘signal.center’ was set to ‘true’ to simulate phylogenetic signal in the ranges of the species, the ‘sd.center’ was set to 3 for liberal standard deviation around the range center signal, ‘sd.range’ was set to 1 for low standard deviation in range centers, and the ‘th’ or probability of presence threshold argument was set to between 0.25 and 1 to achieve the desired community sizes of 30-40% of total species pool sizes. These simulations also resulted in a median species prevalence of ~40-50% of communities. To simulate overdispersion, ‘g.center’ was set to 1, ‘g.repulse’ was set to 1000, ‘repulse.scale’ was 100, ‘repulse’ was set to ‘true’, and ‘th’ was 2; all other argument settings were the same as for clustering.

 I then performed the three analyses described for the empirical dataset but using the simulated community data. First, I calculated the effect sizes of the mean nearest taxon distance (MNTD) and mean pairwise distance (MPD) using the *ses.mpd* and *ses.mntd* functions in the R package *picante* (Kembel et al. 2010) with 1000 permutations and the ‘taxa.labels’ null model in which the identity of the species present in each community was randomized. Results were comparable using the ‘richness’ null model (results not shown). I performed the ecological co-existence – phylogenetic distance regression analyses as in Cavender-Bares et al. 2004 using the quantile regression (0.25, 0.50, 0.75 quantiles) and linear model approaches and the ‘richness’ null model with 1000 permutations. I performed the PGLMM analyses using the *communityPGLMM* function in *pez*, specifying the following random effects: species identity, non-nested phylogenetic effect, and site. I specified the phylogenetic effect using the variance-covariance distance matrix to test for phylogenetic clustering (using the *vcv* function in the R package *ape* Paradis 2000) and the inverse of the variance-covariance matrix (phylogenetic dissimilarity) to test for phylogenetic overdispersion.

The following code is an example of the *scape* arguments used to simulate communities.

###code to simulate 121 communities with fifteen species

library(pez); library(phytools)

tree\_fifteen<-pbtree(n=15, scale=1)##fifteen species, tree depth = 1

##clustered communities

15\_clust<-scape(tree\_fifteen, scape.size=10, g.center=150, g.range=1, g.repulse=1, wd.all=150,

 signal.center=TRUE, signal.range=FALSE, same.range=FALSE, repulse=FALSE,center.scale = 1,

 range.scale = 1, repulse.scale = 1, site.stoch.scale = 0, sd.center=3, sd.range=1,rho=NULL, th=0.35)

15\_clust$Y

fifteen\_clus\_com\_scape<-15\_clust$Y

##transform to plot community presence/absence data

comm\_15\_clust<-as.data.frame(t(fifteen\_clus\_com\_scape))

comm\_15\_clust<-arrange(comm\_15\_clust, -row\_number())

par(mfrow = c(1, 2), las = 1, mar = c(4, 4, 2, 2) - 0.1)

plot(tree\_fifteen)

color2D.matplot(comm\_15\_clust, extremes=c("white", "black"), ylab = "species", xlab = "sites", main = "presence")

### overdispersed communities

15\_repulse<-scape(tree\_fifteen, scape.size=10, g.center=1, g.range=1, g.repulse=1500, wd.all=150,

 signal.center=TRUE, signal.range=FALSE, same.range=FALSE, repulse=TRUE,center.scale = 1,

 range.scale = 1, repulse.scale = 100, site.stoch.scale = 0, sd.center=3, sd.range=1,rho=NULL, th=0.75)

15\_repulse$Y

fifteen\_com\_scape\_repulse<-15\_repulse$Y

comm\_15\_repulse<-as.data.frame(t(fifteen\_com\_scape\_repulse))

library(dplyr)

comm\_15\_repulse<-arrange(comm\_15\_repulse, -row\_number())

par(mfrow = c(1, 2), las = 1, mar = c(4, 4, 2, 2) - 0.1)

plot(tree\_fifteen)

color2D.matplot(comm\_15\_repulse, extremes=c("white", "black"), ylab = "species", xlab = "sites", main = "presence")

**Simulation results**

The results of the simulation suggest that two commonly-used metrics of phylogenetic community structure, mean nearest taxon distance (MNTD) and mean pairwise distance (MPD) have low power to recover the signal of phylogenetic clustering or overdispersion compared to a null model distribution (Table S12). Communities that were simulated with phylogenetic clustering were detected in ~50 – 60% of communities when the species pool consisted of ≤ 50 species. With a species pool of 100, both metrics recovered ~90% of the communities as significantly clustered. Overdispersion was even more difficult to detect, with < 25% of overdispersed communities having MNTD/MPD scores significantly higher than the 95% null distribution (Table S12).

 In contrast to the MNTD and MPD results, the regression technique of Cavender-Bares *et al*. (2004) detected significantly clustered and overdispersed signals correctly with ≥ 10 species. Clustering was detected with a strong signal (regression slopes > 0.10), while the coefficients for overdispersion were small, suggesting weak effects. In simulations with overdispersed communities and large species pools (50 or 100 species), the regression slope was significantly negative, suggesting that the result is biased in the wrong direction. This may be related to difficulty recovering the signal of overdispersion on large trees.

 Finally, the PGLMM technique of Ives and Helmus (2011) detected significant clustering with ≥ 10 species. The variance scalar associated with the phylogenetic distance matrix was significantly different from zero, suggesting strong effects of phylogenetic similarity on species presence / absence in communities simulated to be clustered (Table S12). For communities simulated to be overdispersed, the PGLMM detected a significant variance scalar for the phylogenetic dissimilarity matrix with n = 10 species, but there was no significant effect of phylogenetic dissimilarity with > 10 species (Table S12). Instead, the species-specific variance scalar became larger (> 1.0, not shown), indicating that the PGLMM ascribes the variance in species presence/absence to species-specific idiosyncrasies, rather than phylogenetic effects. PGLMMs were not run with the communities of 50 and 100 species because of the long run-times required.

**Table S12.** Performance of phylogenetic community ecology approaches in inferring significant structure from simulated communities. The power of the mean nearest taxon distance and mean pairwise distance metrics to detect significant deviations from the null expectation, expressed as the percentage of communities that were inferred to be significantly different from the null expectation; i.e., MNTD or MPD values were outside the 95% distribution of random metrics calculated from the null model. Regression coefficients refer to the 0.25, 0.50, 0.75 quantile regression coefficients of the relationship between ecological co-existence and phylogenetic distance (Cavender-Bares et al. 2004). If there is phylogenetic clustering, these regression coefficients are predicted to be negative, while coefficients for overdispersion should be positive. Phylogenetic effect size refers to the variance scalar inferred for the effect of phylogenetic similarity/dissimilarity on species presence/absence in the phylogenetic generalized linear mixed models (PGLMMs, Ives and Helmus 2011).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **n species** | **Clustered MNTD** | **Clustered MPD** | **Overdispersed MNTD** | **Overdispersed MPD** | **Clustered regression coefficients**  | **Overdispersed regression coefficients** | **Clustered****phylogenetic effect size**  | **Overdispersed phylogenetic effect size** |
| 5 | 5% | 0% | 0% | 0% | 0, **0.04**\*, -0.12\* | 0.06\*, 0.09\*, **-0.14**\* | .007 | 2.04e-5 |
| 10 | 12% | 18% | 5% | 5% | -0.14\*, -0.28\*,-0.33\* | 0.0, 0.03\*, 0.10\* | 0.33\*\* | 3.34\*\* |
| 15 | 68% | 56% | 12% | 16% | -0.28\*, -0.35\*, -0.29\* | 0.02\*, 0.03\*, 0.09\* | 0.36\*\* | 4.86e-4 |
| 25 | 33% | 50% | <1% | 3% | -0.33\*, -0.44\*, -0.6\* | 0.01\*, 0.02\*, 0.01\* | 0.88\*\* | 4.86e-4 |
| 50 | 54% | 54% | 21% | 25% | -0.18\*, -0.22\*, -0.23\* | **-0.07\*, -0.10\*, -0.11\*** | - | - |
| 100 | 93% | 90% | 24% | 5% | -0.18\*, -0.38\*, -0.438 | **-0.005\*, -0.006\*,** 0.003  | - | - |

\*regression coefficients are outside the range of the coefficients in the null model (~0 +/- 0.002 SD).

\*\*p<0.05, assessed using likelihood ratio tests comparing a model in which no phylogenetic effect is included *versus* the model in which phylogenetic effect is included.

- not tested because of computational limitations

**Bold cells** are significant results in the opposite direction of the expected pattern.

1.

Garamszegi, L.Z. & Mundry, R. (2014). Multi-model inference in comparative analyses. In: *Modern phylogenetic comparative methods and their application in evolutionary biology* (ed. Garamszegi, LZ). Springer Berlin, pp. 305-335.

2.

Kraft, N.J.B., Cornwell, W.K., Webb, C.O. & Ackerly, D.D. (2007). Trait evolution, community assembly, and the phylogenetic structure of ecological communities. *The American Naturalist*, 170, 271-283.

3.

Paradis, E., Claude, J. & STrimmer, K. (2004). APE: analsysis of phylogenetics and evolution in R language. . *Bioinformatics*, 20, 289-290.

4.

Pearse, W.D., Cadotte, M., Cavender-Bares, J., Ives, A.R., Tucker, C., Walker, S. *et al.* (2015). pez: Phylogenetics for the Environmental Sciences. *Bioinformatics*, btv277.

5.

Revell, L. (2015). phytools v0.4-45: Phylogenetic tools for comparative biology (and other things). CRAN.

6.

Ribeiro, P. & Diggle, P. (2015). Package 'geoR'. CRAN.