Supporting Information for

Temporal population variability in local forest communities has mixed effects on tree species richness across a latitudinal gradient

Tak Fung, Ryan A. Chisholm*, Kristina Anderson-Teixeira, Norm Bourg, Warren Y. Brockelman, Sarayudh Bunyavejchewin, Chia-Hao Chang-Yang, Rutuja Chitra-Tarak, George Chuyong, Richard Condit, H. S. Dattaraja, Stuart J. Davies, Corneille E. N. Ewango, Gary Fewless, Christine Fletcher, C. V. Savitri Gunatilleke, I. A. U. Nimal Gunatilleke, Zhanqing Hao, J. Aaron Hogan, Robert Howe, Chang-Fu Hsieh, David Kenfack, YiChing Lin, Keping Ma, Jean-Remy Makana, Sean McMahon, William J. McShea, Xiangcheng Mi, Anuttara Nathalang, Perry S. Ong, Geoffrey Parker, E-Ping Rau, Jessica Shue, Sheng-Hsin Su, Raman Sukumar, I-Fang Sun, H. S. Suresh, Sylvester Tan, Duncan Thomas, Jill Thompson, Renato Valencia, Martha I. Vallejo, Xugao Wang, Yunquan Wang, Pushpa Wijekoon, Amy Wolf, Sandra Yap, Jess Zimmerman

* Correspondence: E-mail: ryan.chis@gmail.com

Appendix S1: Summary statistics and plot-specific acknowledgements for forest plots used

Table S1 provides summary statistics for the 21 CTFS–ForestGEO plots considered in our study. Tree census data from these plots were used to quantify tree species richness and temporal population variability of tree species at each plot, which were then used in statistical analyses. For each plot, Table S1 shows the location (latitude and longitude), area, number of tree censuses and total time period of censuses, together with the mean number of species and mean number of individuals for the tree censuses. Here, an individual is defined as one tree, which may have one stem or more than one stem connected either above- or below-ground.

In each plot, all freestanding woody plants with diameter-at-breast-height (DBH; 1.3 m from the ground) ≥ 1 cm were censused (Condit 1998). This threshold of 1 cm was sufficiently small to include many juvenile (non-reproductive) trees – for example, at the Barro Colorado plot, using the reproductive thresholds for tree species determined by Robin Foster (one of the Principal Investigators of the plot; unpublished data), we estimated that 62% of trees were juveniles. Thus, our analyses included both juvenile and adult trees, and hence captured temporal fluctuation-dependent mechanisms acting on both life-history stages. This is important because there is evidence that competition acts most strongly in the juvenile stage (Van Valen 1975; Harcombe 1987; Kitajima & Auspurger 1989), and this might translate into stronger temporal fluctuation-dependent mechanisms.

We included palms in our analyses, but excluded any censused ferns (Smith *et al.* 2006) because their life-histories and growth patterns can be very different to other woody species. Only six plots had fern species, and for these six plots, ferns constituted a small percentage of total abundance in each census (mean of 1.4%). Taxonomic groups representing morphospecies (i.e. not assigned a species name but treated as separate groups at the genus level) were treated as distinct species in our analyses. In addition, unidentified taxonomic groups were treated as distinct species. However, because unidentified groups usually made up only a small percentage of total abundance and total number of taxonomic groups in each census (averages for the 21 plots of 0.08% and 0.9%, respectively), our results are robust to future taxonomic refinements. In addition, for each plot, the species names were checked for synonyms using The Plant List online database (accessed using the R package "Taxonstand" (Cayuela *et al.* 2012)), and abundances were summed for synonymous species identified. This step was necessary because botanists at different plots may have used different rules to determine whether two species names were to be treated as synonymous or not, and these rules needed to be standardized across plots. Synonymous species were identified in 15 of the 21 plots, but usually made up only a small percentage of total abundance and total number of taxonomic groups in each census (means of 1.5% and 2.8% respectively for the 15 plots).

Tree cover at each of the 21 plots was estimated to be 78–106% of the corresponding pristine value (see Table S5 in Anderson-Teixeira *et al.* (2015)), where the pristine value was defined using a global raster map of pre-human modification forest cover produced by UNEP-WCMC (Anderson-Teixeira *et al.* 2015). Thus, there is little evidence of substantial anthropogenic disturbance in the recent past. Also, from Table S1, we see that tree species richness in the 21 plots showed a declining trend away from the tropics. This decreasing trend persisted after rarefaction to account for the different numbers of individuals in the plots (Appendix S4).

Table S2 shows a list of acknowledgements specific to each plot.

Table S1. Summary statistics for the 21 CTFS–ForestGEO plots considered. For each plot, the table shows the location (latitude and longitude), plot area, number of censuses, overall census period, mean number of species in censuses and mean number of individuals in censuses. SCBI and SERC stand for Smithsonian Conservation Biology Institute and Smithsonian Environmental Research Center, respectively.

Forest plot	Latitude (^{o})	Longitude (^{o})	Plot area (ha)	Number of censuses used in analyses	Total time period of censuses used	Mean number of species per census used	Mean number of individuals per census used
Barro Colorado Island	9.15	–79.85	50	7	1981–2010	301	226,000
Changbaishan	42.48	128.08	25	2	2004–2009	52	35,800
Edoro	1.56	28.52	20	3	1994–2007	353	154,000
Fushan	24.76	121.56	25	2	2003–2009	104	112,000
Gutianshan	29.25	118.12	24	2	2005–2011	158	130,000
Huai Kha Khaeng	15.63	99.22	50	4	1992–2010	286	88,500
Khao Chong	7.54	99.80	24	3	2000–2010	554	98,400
Korup	5.07	8.85	50	2	1997–2009	483	321,000
Lambir	4.19	114.02	52	4	1991–2008	1,330	364,000
La Planada	1.16	–77.99	25	2	1997–2003	221	102,000
Lenda	1.31	28.65	20	3	1994-2007	340	131,000
Luquillo	18.33	–65.82	16	5	1990–2012	138	57,600
Mo Singto	14.43	101.35	30.5	2	2003–2011	263	133,000
Mudumalai	11.60	76.53	50	6	1988–2008	71	21,500
Palanan	17.04	122.39	16	3	1998–2010	310	70,000
Pasoh	2.98	102.31	50	5	1986–2006	794	312,000
SCBI	38.894	–78.15	25.6	2	2008–2013	65	32,400

SERC	38.889	–76.56	16	2	2008-2014	71	23,600
Sinharaja	6.40	80.40	25	3	1993–2008	215	193,000
Wabikon	45.55	–88.79	25.2	2	2008–2013	36	47,500
Yasuni	-0.69	–76.40	25	2	1995–2003	1,070	144,000

Table S2. **Plot-specific acknowledgements.** Acknowledgements specific to the different CTFS–ForestGEO plots considered in our study.

Forest plot	Acknowledgement
Barro Colorado Island (BCI)	The BCI forest dynamics research project was founded by S.P. Hubbell and R.B. Foster and is now managed by R. Condit, S. Lao, and R. Perez under the Center for Tropical Forest Science and the Smithsonian Tropical Research in Panama. Numerous organizations have provided funding, principally the U.S. National Science Foundation, and hundreds of field workers have contributed.
Fushan	Fushan FDP is supported by the Taiwan Forestry Bureau, the Taiwan Forestry Research Institute and the Ministry of Science and Technology of Taiwan. We thank the staff at Fushan Research Center for providing logistic support.
Huai Kha Khaeng	Direct financial support for the Huai Kha Khaeng plot has been provided by the Royal Thai Forest Department and the National Parks Wildlife and Plant Conservation Department, the Arnold Arboretum of Harvard University (under NSF award #DEB-0075334, and grants from USAID and the Rockefeller Foundation), the Smithsonian Tropical Research Institute, and the National Institute for Environmental Studies, Japan. The Huai Kha Khaeng Forest Dynamics Plot is part the Center for Tropical Forest Science, a global network of large-scale demographic tree plots. We acknowledge the Royal Thai Forest Department for supporting and maintaining the project in Huai Kha Khaeng Wildlife Sanctuary, Thailand.
Lambir	The 52-ha Long-Term Ecological Research Project is a collaborative project of the Forest Department of Sarawak, Malaysia, the Center for Tropical Forest Science of the Smithsonian Tropical Research Institute, the Arnold Arboretum of Harvard University, USA (under NSF awards DEB-9107247 and DEB-9629601), and Osaka City, Ehime & Kyoto Universities, Japan (under Monbusho grants 06041094, 08NP0901 and 09NP0901). The Lambir Forest Dynamics Plot is part of the Center for Tropical Forest Science, a global network of large-scale demographic tree plots. We acknowledge the Sarawak Forest Department for supporting and maintaining the project in Lambir Hills National Park.
Luquillo	The 16-ha Luquillo Forest Dynamics Plot was supported by grants BSR-8811902, DEB 9411973, DEB 0080538, DEB 0218039, DEB 0620910, DEB 0963447 AND DEB-129764 from NSF to the Department of Environmental Science, University of Puerto Rico, and to the International Institute of Tropical Forestry USDA Forest Service, as part of the Luquillo Long-Term Ecological Research Program. The Andrew Mellon Foundation and the University of Puerto Rico gave additional support. We especially thank the hundreds of field workers who contributed to the tree censuses.

Mo Singto	The 30-ha plot is supported by National Science and Technology Development Agency (Thailand); The Department of National Park, Wildlife and Plant Conservation; and Thai Ministry of Natural Resources and Environment.
Mudumalai	The 50 hectare Mudumalai Forest Dynamics plot was set up by the Centre for Ecological Sciences, Indian Institute of Science, Bangalore. Most of the long-term funding for running the plot has come from the Ministry of Environment, Forest and Climate Change (Government of India). In recent years this has been supplemented with funding from the Department of Biotechnology (Government of India), the JC Bose National Fellowship (Department of Science and Technology), and the Divecha Centre for Climate Change, Indian Institute of Science. We acknowledge the support of Tamil Nadu Forest Department for this long-term monitoring.
Palanan	Research in the 16-ha Palanan Forest Dynamics Plot is collaboratively managed by the Institute of Biology, University of the Philippines Diliman and the Smithsonian Tropical Research Centre/Centre for Tropical Forest Science while additional grants were provided by the University of the Philippines Office of the Vice President For Academic Affairs, Commission on Higher Education, the Department of Science and Technology and the Energy Development Corporation. The assistance and continuing support of the Protected Areas Management Board of the Northern Sierra Madre Natural, the Department of Environment and Natural Resources and the local government unit of Palanan, Isabela as well as the local community are gratefully acknowledged.
Pasoh	Data from the Pasoh Forest Reserve was provided by the Forest Research Institute Malaysia – Smithsonian Tropical Research Centre/Centre for Tropical Forest Science collaborative research and support from the Negeri Sembilan State Forestry Department.
Smithsonian Conservation Biology Institute (SCBI)	Funding for the establishment of the SCBI ForestGEO Large Forest Dynamics Plot was provided by the Smithsonian Global Earth Observatory initiative, the Smithsonian Institution, National Zoological Park and the HSBC Climate Partnership. We especially thank the numerous technicians, interns and volunteers of the Conservation Ecology Center at the SCBI who were essential in assisting with plot establishment and data collection. Support for the original exclosure fence installation was provided by the Friends of the National Zoo and Earthwatch Foundation.
Sinharaja	The 25-ha Long-Term Ecological Research Project at Sinharaja World Heritage Site is a collaborative project of the University of Peradeniya, the Center for Tropical Forest Science of the Smithsonian Tropical Research Institute and the Arnold Arboretum of Harvard University, USA, with supplementary funding received from the John D. and Catherine T. Macarthur Foundation, the National Institute for Environmental Science, Japan, and the Helmholtz Centre for Environmental Research-UFZ, Germany, for past censuses. The PIs gratefully acknowledge the

	Forest Department and the Post-Graduate Institute of Science at the University of Peradeniya, Sri Lanka for supporting this project, and the local field and lab staff who tirelessly contributed in the repeated censuses of this plot.
Wabikon	Major support for research at the Wabikon Forest Dynamics Plot has been provided by The 1923 Fund, the Smithsonian Tropical Research Institute, and the U.S. Forest Service. Gary Fewless, Kathryn Corio, and Juniper Sundance have been key contributors to research at the site, which has engaged more than 50 students from UW-Green Bay and other institutions.
Yasuni	The Yasuni plot establishment and censuses were supported by Pontifical Catholic University of Ecuador (Donaciones del Impuesto a la Renta from the government of Ecuador and PUCE grants L13251, M13373 in recent years) and the Center for Tropical Forest Science of the Smithsonian Tropical Research Institute. We are grateful to researchers, field-workers, students and volunteers who have helped with collecting and processing census data. The Yasuni Scientific Station offered logistic support. The continuous research in the Yasuni plot is endorsed by the Ministerio de Ambiente del Ecuador through several research permits (in particular permits: No. 002-015-IC-FLO-PNY-DPAO; No. 025-2016-IC-FAU-FLO-DPAO-PNY and No. 007-2018-IC-PNY-DPAO/AVS to Renato Valencia).

Appendix S2: Assessing bias in metrics of temporal population variability

Assessing bias with respect to changing species richness

In our study, the aim was to assess the degree of temporal population variability in different tree communities and how this variability affects coexistence of tree species in the communities. To avoid confounding effects of species richness on temporal population variability, we needed to use a metric of temporal population variability that was largely invariant to changes in species richness, given a particular environmental regime. In other words, we needed a metric with a low bias with respect to species richness, where the bias is defined as the change in the metric's expected value with species richness. To assess the bias of candidate metrics, we applied these metrics to simulated data from model communities with varying species richness but in a fixed environment regime. Each model community was constructed using the dynamic, mechanistic model developed by Danino *et al*. (2016).

Specifically, the mechanistic model (Danino *et al*., 2016) represented the abundance dynamics of a community of species competing for a fixed amount of limiting resources, given by a fixed carrying capacity of J individuals. At the beginning of a model simulation, the model community was initialized with J individuals distributed among S_0 species. The fitness of each species, which determined its recruitment rate, was chosen randomly from a lognormal distribution with mean 1 and variance A. The different fitness values among species represented different responses of the species to the prevailing environmental conditions. At the beginning of each subsequent time-step of the model, the environmental conditions changed with probability $1/\tau$, and this change was represented indirectly by the fitness values of all species being redrawn randomly and independently from the lognormal distribution (Fig. S1). Thus, τ measured the temporal correlation in environmental conditions. Together, A and τ indirectly determined the temporal sequence of sets of environmental conditions that species in a model community experienced. Thus, a particular combination of A and τ defined an environmental regime. After determining whether the environmental conditions had changed, an individual was chosen at random to die, and then another individual was chosen at random as a replacement. With probability ν the replacement individual was of a new species. The introduction of new species prevented the model community from collapsing to a monoculture after a long period of time, such that the model was like Hubbell's neutral model (Hubbell 1997, 2001) except with species that have different fitness values, which fluctuate in a temporally changing environment. With probability $1 - \nu$ the replacement individual was chosen randomly from the pool of potential recruits produced by existing species (Fig. S1). The contribution of each existing species to the pool of potential recruits was proportional to the product of the abundance of the species and its recruitment rate, which was assumed to be proportional to its fitness. Thus, existing species compete for the resources vacated by the dead individual according to lottery dynamics (as in the model of Chesson & Warner (1981), except that in the model we use, each time-step involved the death and replacement of only one individual).

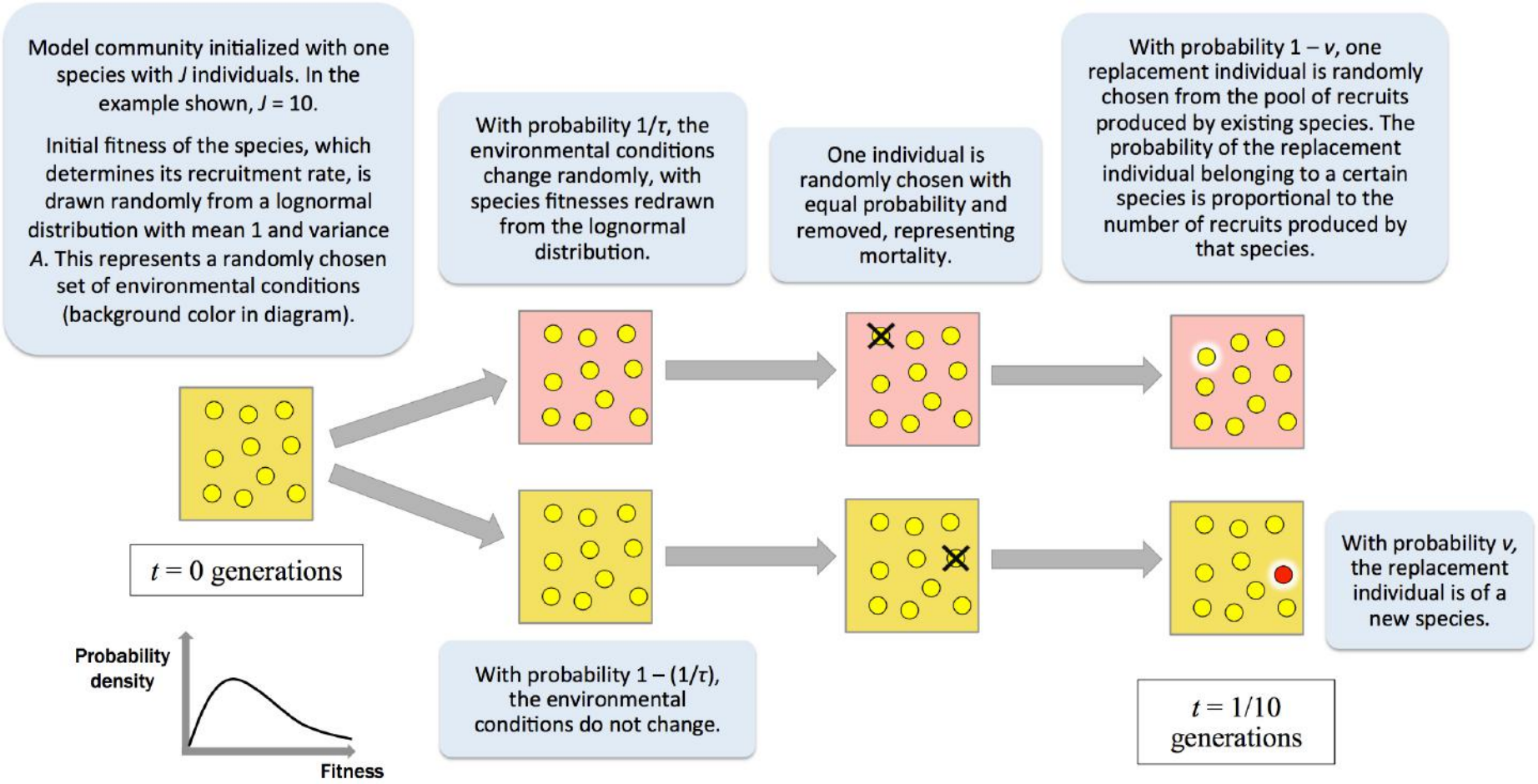


Fig S1. Schematic diagram showing the processes operating in the dynamic, mechanistic model used to assess bias of metrics of temporal population variability. The diagram shows the processes operating over one model time-step.

For a given environmental regime (A and τ, where τ is expressed in generations), we constructed a set of 20 model communities with different species richness by using 20 values of ν from 10^{-4} to 0.1 (equally spaced on a natural log-scale). For each value of ν, we initialized a model community with $S_0 = 1$ species with $J = 15{,}299$ individuals, corresponding to the number of individuals in the 21 tree communities that we examined, after rarefaction to standardize the number of individuals. Model dynamics were then simulated for 1,000 generations (gen; each generation consists of J individuals dying and being replaced), which was sufficient time for the emergence of model communities spanning three orders of species richness S (Fig. S2), covering the empirical values for the 21 tree communities (Fig. S8). After 1,000 gen, model dynamics were simulated for a further 500 gen and the changes in abundance used to calculate metrics of temporal population variability. We examined two candidate metrics: the first measures the mean absolute change in abundance in a species in a year ($\overline{\Delta N}$; calculated using eq. (1) in main text), whereas the second is the same as the first but with correction of the simulated data for different sets of initial abundances among communities ($\overline{\Delta N_c}$; calculated as described in Appendix S3 below – here, we drop the subscript "r" that refers to rarefaction). Time in the model was measured as the number of generations, so to calculate the metrics using data from the model, we specified an inter-census interval length of $T = 2{,}500/J$ gen $= 0.163$ gen, based on the average for the 21 forest tree communities (an empirical value of T for a pair of consecutive censuses at a plot was calculated as the number of deaths over the inter-census interval divided by community size). To convert the units of the metrics from gen to yr^{-1}, we took the inter-census interval length to be equal to 5.14 yr, based on the average for the 21 forest tree communities. The simulation of 500 gen provided many pairs of consecutive intervals, which allowed the expected changes in absolute species abundances in a year to be estimated with high accuracy. We performed these simulations for 18 environmental regimes represented by all 18 combinations of $A \in \{10^{-4}, 0.01, 1, 100, 10^4, 10^6\}$ and $\tau \in \{0.01\text{ gen}, 0.1\text{ gen}, 1\text{ gen}\}$, encompassing nearly all of the values of A and τ that we examined when fitting the model (without introduction of new species) to the 21 tree communities (see main text).

We found that for each of the 18 simulated environmental regimes, the first metric of population variability ($\overline{\Delta N}$) showed a large bias with respect to species richness (S) (Fig. S2). The value of $\overline{\Delta N}$ almost always decreased with S by substantial amounts, sometimes by several orders of magnitude (Fig. S2). In contrast, the second metric of population variability ($\overline{\Delta N_c}$) showed little bias with respect to S (Fig. S2). For each of the 18 environmental regimes, the coefficient of variation (CV) of $\overline{\Delta N_c}$ was small, ranging from 0.00574–0.0321 (Fig. S2). This meant that for each of the 18 environmental regimes, the ratio of $\overline{\Delta N_c}$ at the smallest value of S to $\overline{\Delta N_c}$ at the largest value of S was reasonably close to 1, ranging from 0.932–1.07 (Fig. S2).

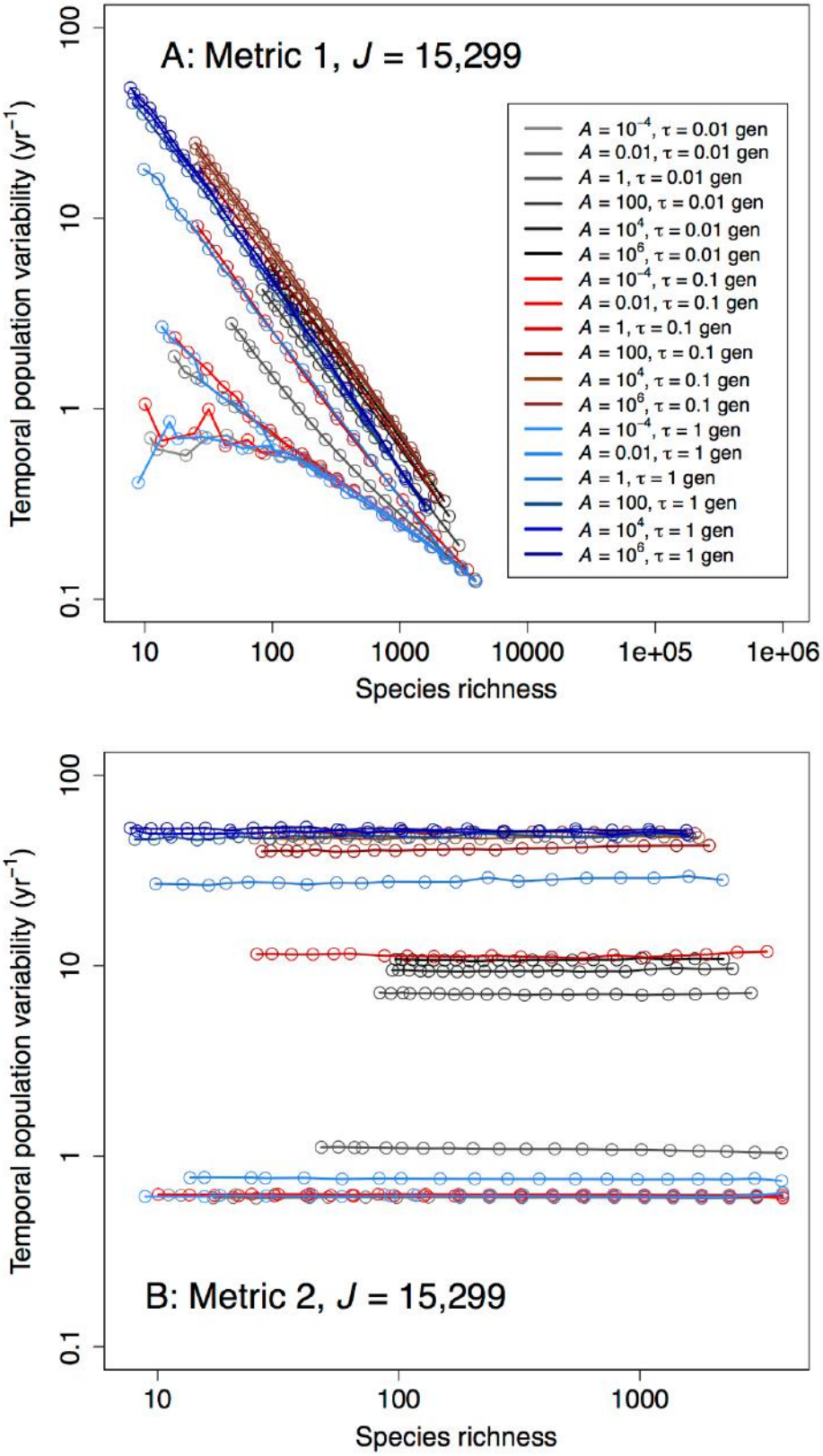


Fig. S2. Values of two metrics of temporal population variability ($\overline{\Delta N}$ and $\overline{\Delta N_c}$) across model communities with the same community size ($J = 15{,}299$) but different species richness (S), for 18 environmental regimes. Panels **(A)** and **(B)** show results for $\overline{\Delta N}$ (metric 1) and $\overline{\Delta N_c}$ (metric 2) respectively. Each environmental regime was defined by two parameters: the variance in fitness of species (A) and the temporal correlation of environmental conditions

(τ). For an environmental regime, $\overline{\Delta N}$ and $\overline{\Delta N_c}$ were calculated using data from the last 500 gen of model simulations lasting 1,500 gen. In addition, *S* was calculated as the mean number of species in the last 500 gen.

To examine whether our results held for different spatial scales, we repeated the analyses above for model communities that were about half and twice the size, i.e. with $J = 7{,}500$ and $J = 30{,}000$ respectively. We found that with either $J = 7{,}500$ and $J = 30{,}000$, the trends of $\overline{\Delta N}$ and $\overline{\Delta N_c}$ with *S* were essentially the same as with $J = 15{,}299$ (Figs. S3 and S4). From these results, we also found that the values of $\overline{\Delta N_c}$ among the 18 different environmental regimes retained largely the same ranking regardless of *J* (Figs. S2B, S3B and S4B). The rank of an environmental regime changed by an average of only 0.667 when *J* decreased from 15,299 to 7,500 and an average of only 1.44 when *J* increased from 15,299 to 30,000. This result indicated a similar ranking of mean extinction times among the different environmental regimes. Thus, we expect results from the main text to hold qualitatively at larger spatial scales.

We conclude that the first metric of temporal population variability, $\overline{\Delta N}$, was inappropriate due to its large bias with respect to species richness *S*. This large bias in $\overline{\Delta N}$ was not surprising because, as explained in the main text, a community with higher *S* would necessarily have species with smaller population sizes on average, and species with smaller population sizes tend to have lower temporal population variability (Chisholm *et al*. 2014). This bias can be corrected for by considering only those species with initial population sizes common to all plots when calculating temporal population variability, as for the second metric of temporal population variability, $\overline{\Delta N_c}$. Our simulation results showed that the second metric had little bias, at least for the range of *S* and inter-census intervals pertaining to the 21 forest tree communities. Thus, we conclude that $\overline{\Delta N_c}$ was an appropriate metric of temporal population variability for the empirical analyses in our study. We also note that $\overline{\Delta N_c}$ differed substantially among the 18 environmental regimes (Fig. S2B), which showed that the metric was able to differentiate among environmental regimes with respect to how they affect temporal population variability.

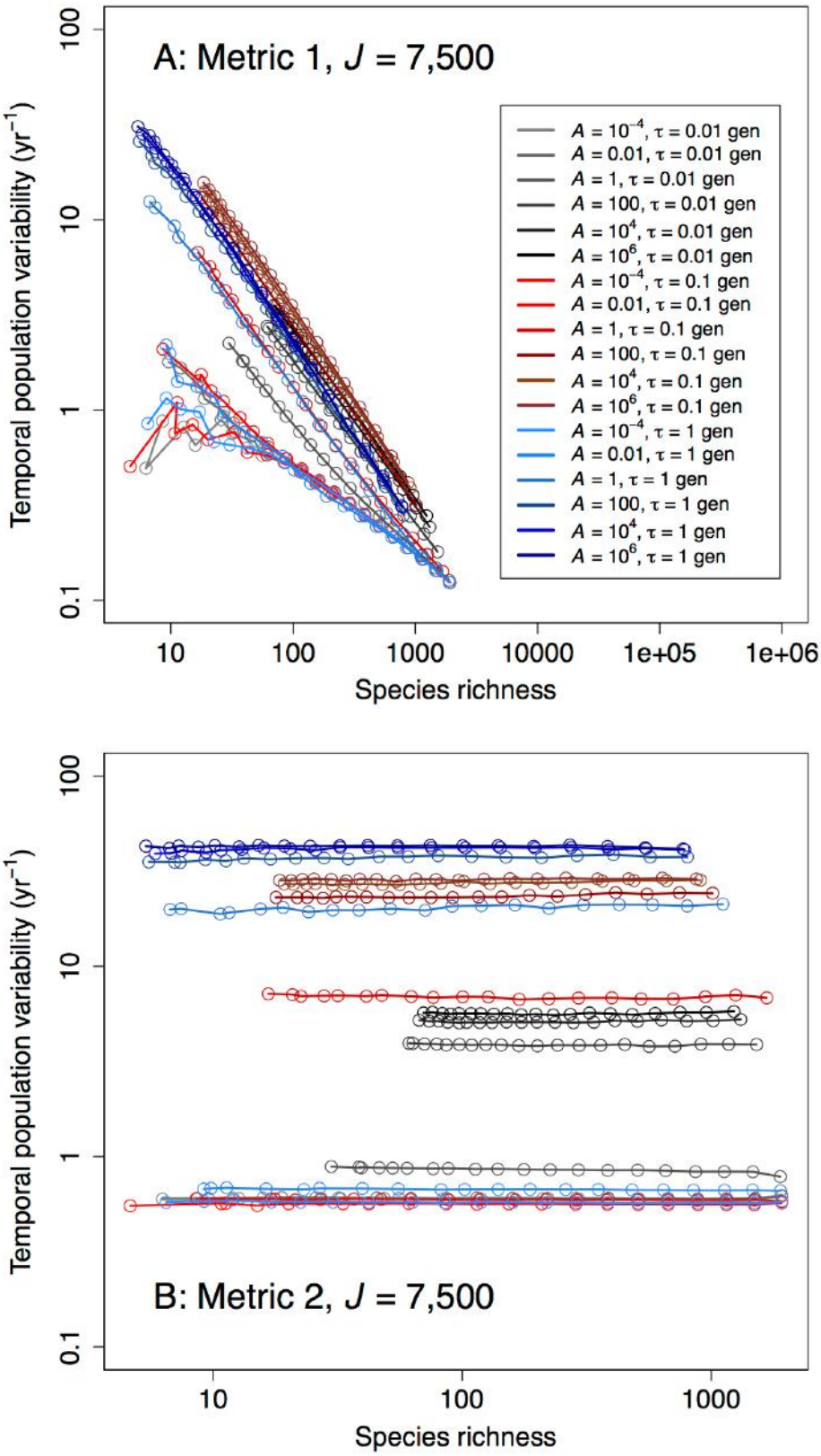


Fig. S3. Same as Fig. S2, except with $J = 7{,}500$ instead of $J = 15{,}299$.

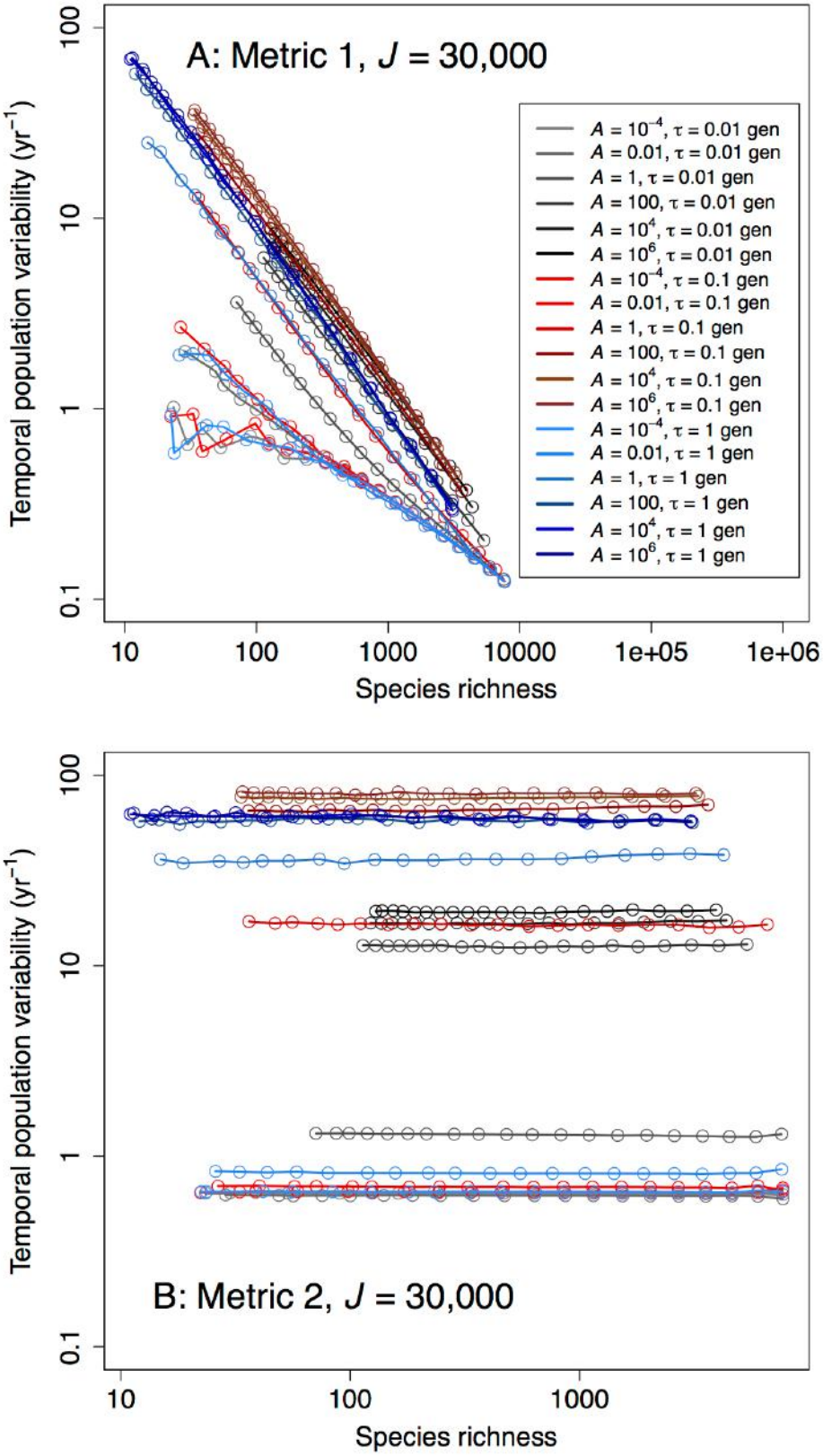


Fig. S4. Same as Fig. S2, except with $J = 30{,}000$ instead of $J = 15{,}299$.

Assessing bias with respect to sample size

In our study, we calculated temporal population variability using $\overline{\Delta N_c}$, for 21 forest tree communities. Before we calculated $\overline{\Delta N_c}$, we performed rarefaction to standardize the number of individuals for each community. This rarefaction consisted of random sampling down to a common number of individuals. Ideally, for a fixed environmental regime, the sampling procedure would give the same value of $\overline{\Delta N_c}$ when sampling from different community sizes. We refer to the change in the expected value in $\overline{\Delta N_c}$ with sampling from different community sizes as the metric's bias with respect to sample size. To assess this bias, we considered three environmental regimes and for each one, we ran simulations of our model to construct 20 model communities with the same community size *J* but a range of species richness *S*. We did this for $J = 15{,}299$, 30,000 and 60,000. For each *J*, we calculated $\overline{\Delta N_c}$ for the 20 communities after sampling to a sample size of 15,299. In calculating $\overline{\Delta N_c}$, we used the inter-census interval of $T = 0.163$ gen, as in our tests of bias in $\overline{\Delta N_c}$ with respect to *S* (described in the previous subsection). The three environmental regimes we examined were $A = 10^{-4}$ and $\tau = 0.01$ gen; $A = 1$ and $\tau = 0.1$ gen; and $A = 10^{6}$ and $\tau = 1$ gen, representing weak, moderate and strong (temporal) environmental variance.

We found that for the moderate and strong environmental variance regimes, $\overline{\Delta N_c}$ for the 20 communities remained largely invariant to sampling from different community sizes, with the mean $\overline{\Delta N_c}$ across the 20 communities changing by less than 5.73% when J increased four-fold from 15,299 to 60,000 (Fig. S5). For the weak environmental regime, mean $\overline{\Delta N_c}$ showed a large percentage increase of 47.0% when J increased four-fold, but a small absolute increase of 0.269 yr^{-1} (Fig. S5). The small absolute change reflects $\overline{\Delta N_c}$ always being small, ranging from 0.565–1.12 yr^{-1} (Fig. S5; for comparison, the empirical values of $\overline{\Delta N_c}$ varied from 1.40 – 9.95 yr^{-1}). We note that for the weak environmental regime, $\overline{\Delta N_c}$ showed an upward trend for the 20 communities at very high values of $S > 3{,}500$, but this is far beyond the empirical upper limit of 881 (Fig. S5).

From our simulation results, we conclude that the relative bias with respect to sample size was small when the environmental regime was moderate or strong. We also conclude that although the relative bias was large when the environmental regime was weak, the absolute bias was small. In addition, we note that the weak environmental regime was unlikely to be relevant to the forest tree communities that we examined, because it produced values of $\overline{\Delta N_c}$ below our empirical values.

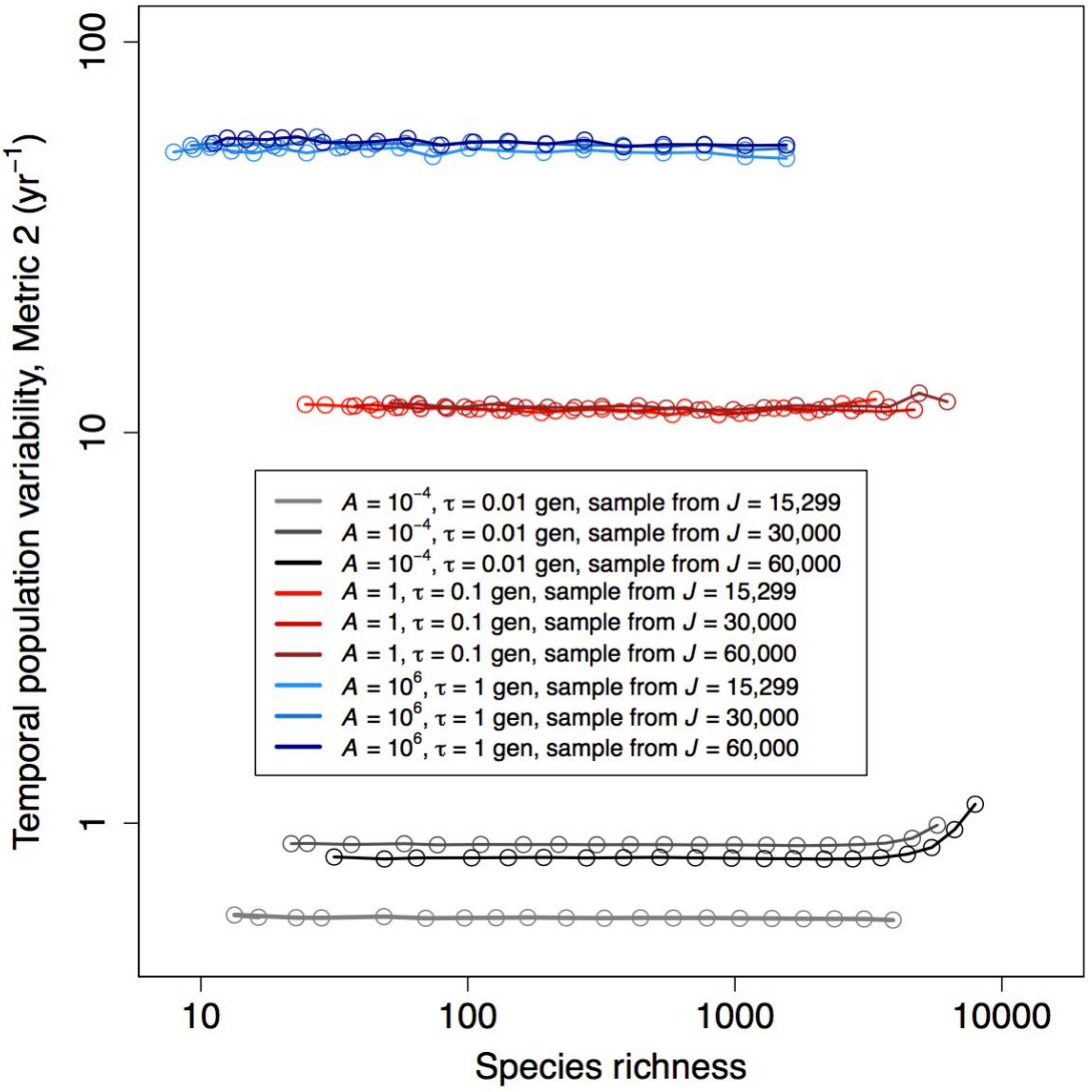


Fig. S5. Values of temporal population variability ($\overline{\Delta N_c}$) in samples of model communities with the same community size (J) but different species richness (S), for three environmental regimes. Results are shown for $J = 15{,}299$, $J = 30{,}000$ and $J = 60{,}000$. Other details are as in caption for Fig. S2.

Appendix S3: Further details on calculation of metrics of temporal population variability

Correcting for different sets of species abundances among plots

In our study, as a metric of temporal population variability for the tree community in each of the 21 forest plots examined, we first considered the mean absolute change in abundance of a species in a year (eq. (1) in main text), calculated using tree census data rarefied to a chosen number of individuals ($\overline{\Delta N_r}$). However, as discussed in the main text, there remained an important potential source of bias: given a fixed total tree abundance, abundant species were over-represented in species-poor plots, while rare species were over-represented in species-rich plots. Thus, for the different plots, there was a need to correct for the different sets of species abundances.

To do this correction, for each plot, we considered all pairs of consecutive censuses in all of the sampled datasets (resulting from rarefaction), and extracted the set of all species abundances in the first censuses of these pairs (set of all "initial abundances"). Considering the sets of initial abundances for the 21 plots, we determined those initial abundances that were shared among all 21 plots. There were 222 of these initial abundances, ranging from 1 to 394. The number of initial abundances here was greater than the number of species in some of the plots, which was possible because a species at a plot sometimes had more than one initial abundance across censuses and sampled datasets. For each plot and each of the 222 initial abundances, we filtered the sampled datasets such that for each pair of consecutive censuses, only species with the initial abundance considered were retained, and then used the filtered datasets with eq. (1) to calculate the mean absolute change in abundance in a year. Then for each plot *i*, we calculated the average of the mean absolute change in abundance across the 222 initial abundances, which we then used as the metric of temporal population variability corrected for initial abundances, denoted by $\overline{\Delta N_{r,c,i}}$. This method corrected for different sets of initial abundances across plots because it ensured that: (i) only the same initial abundances were considered across plots and (ii) for each of these initial abundances, only one (mean) value of change in abundance was used for each plot, which avoided confounding effects arising from different plots having differing numbers of data points for each initial abundance.

Rarefaction procedure that standardized area and the number of individuals, in a way that conserved the temporal correlations of species abundances

In the main text, we used a metric of temporal population variability ($\overline{\Delta N_{ra,c}}$) that was calculated using data rarefied according to sample area and the number of individuals, in a way that conserved (pairwise) temporal correlations of species abundances. This rarefaction procedure consisted of two steps. In the first step, we standardized the area of each of the 21 forest plots that we examined to 16 ha, which was the area of the three smallest plots (Luquillo, Palanan and SERC). To perform this standardization, for each of the 18 plots with area greater than 16 ha, we randomly sampled 1,000 16 ha areas. Specifically, 15 of these 18 plots were square or rectangular, and for these 15 plots, we randomly sampled 1,000 16 ha rectangles. Each random sample was achieved by randomly sampling the *x*- and *y*-coordinates of the bottom-left and bottom-right corners of the 16 ha rectangle (Fig. S6). These coordinates fixed the coordinates of the top-left and top-right corners of the rectangle (due to the constraint that the area was 16 ha). Two of the remaining three plots (Lenda and

Edoro) each consisted of two rectangular sub-plots of area 10 ha, separated in space. Therefore, for each of these two plots, each random sample of 16 ha was achieved by randomly sampling a 8 ha rectangle from each sub-plot. The remaining plot (Mo Singto) consisted of two rectangles joined together. For this plot, the randomly sampled 16 ha area was either a rectangle or two rectangles joined together. However, the sampling procedure was the same as when the plot was rectangular – each random sample was achieved by randomly sampling the x- and y-coordinates of the bottom-left and bottom-right corners of the 16 ha area, which fixed the other coordinates of the area.

After standardizing according to area, the second step involved standardizing according to the number of individuals, in a way that conserved temporal correlations of species abundances in the 16 ha areas. This second step was based on the idea that if the abundances of species populations in a 16 ha area were all multiplied by the same scaling factor in all censuses, then the temporal correlations of the species abundances would remain the same. The scaling factor for each 16 ha area was chosen such that after scaling, the mean number of individuals across censuses was the same for all 16 ha areas. Specifically, we first calculated the minimum mean number of individuals across censuses for the 16 ha areas for all plots (i.e., the 1,000 randomly sampled 16 ha areas for each of the 18 plots with area greater than 16 ha, together with the three 16 ha plots). We found that this minimum mean number of individuals was $\overline{N}_{\min} = 4{,}713$. Afterwards, for each 16 ha area, we scaled the species abundances in each census by the factor $\overline{N}_{\min}/\overline{N}$, where $\overline{N}$ was the mean number of individuals in the 16 ha area across all censuses, and rounded fractional abundances to the nearest integer (Fig. S6). This rarefaction step ensured that in each 16 ha area, the mean number of individuals was approximately $\overline{N}_{\min}$ and the temporal correlations of species abundances were approximately the same as that before the rarefaction step. The temporal correlations were not exactly the same because of rounding the fractional abundances to integers.

For the 50 ha Barro Colorado Island plot, rarefaction according to area in the first step produced 1,000 randomly sampled 16 ha areas. Within each sampled 16 ha area, the mean absolute difference in temporal correlation of species abundances before and after rarefaction in the second step was calculated. The average value of this mean absolute difference across all 1,000 sampled 16 ha areas was 0.151, which was quite small (Fig. S7). Similarly, for the other 20 plots, the average value of the mean absolute difference in temporal correlation was 2.38×10^{-18}–0.228 (median for all 21 plots of 6.66×10^{-13}). In contrast, for our method of rarefaction that did not conserve temporal correlations (simple random sampling down to the smallest number of individuals found in a census in a plot), the average value of the mean absolute difference before and after rarefaction was 0.337–0.960 for the 21 plots (median of 0.796), which was much larger (Fig. S7).

After carrying out the two steps, we ended up with 1,000 random samples for the 18 forest plots with area greater than 16 ha, and one random sample for the remaining three 16 ha forest plots. These samples were used to calculate our metric of temporal population variability $\overline{\Delta N_{ra,c}}$, which involved correcting for the different species abundances among the samples for the different plots (as described in the previous subsection).

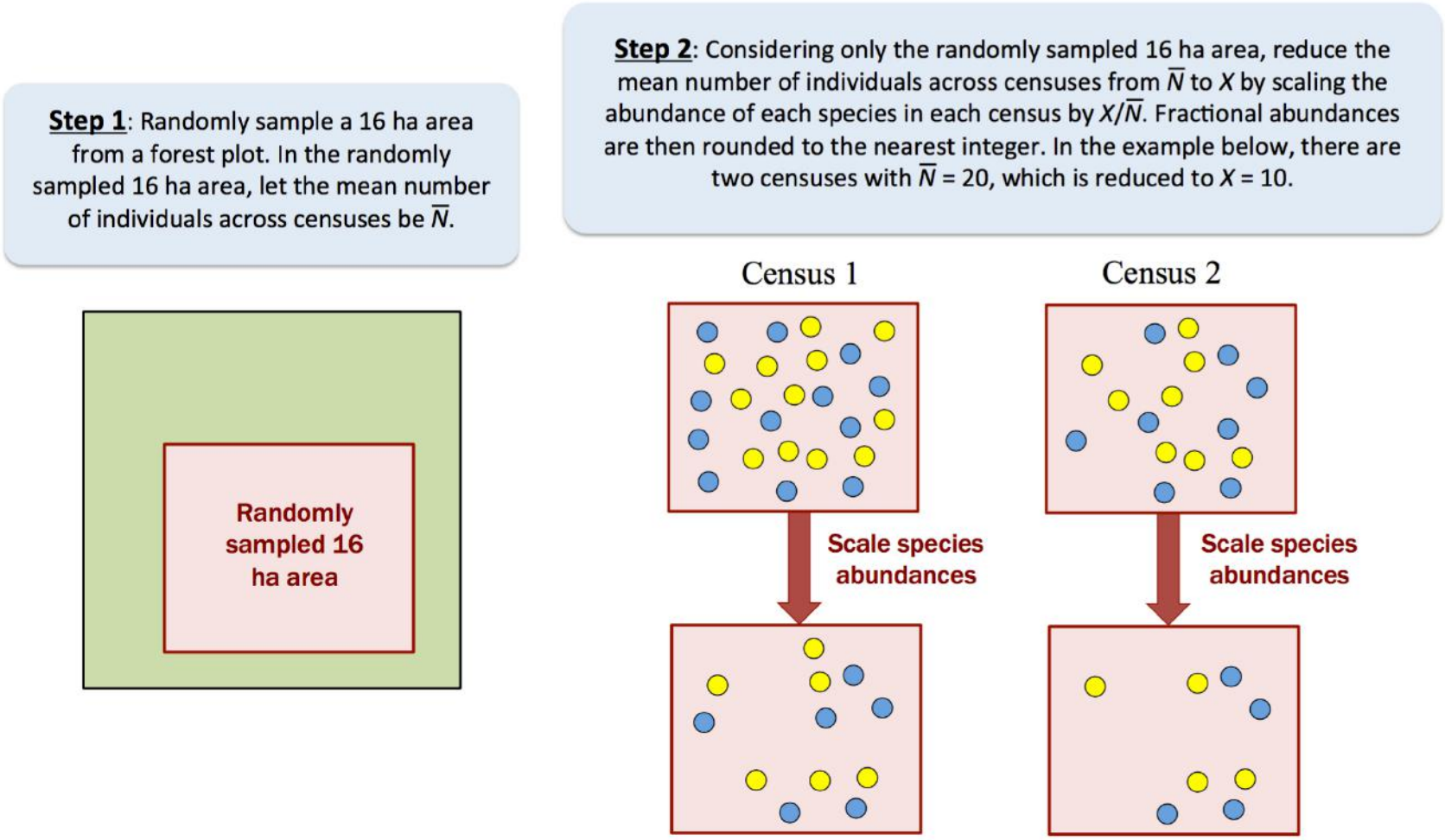


Fig. S6. Schematic diagram showing the two-step rarefaction procedure used for standardizing area and the number of individuals, in a way that conserved temporal correlations of species abundances.

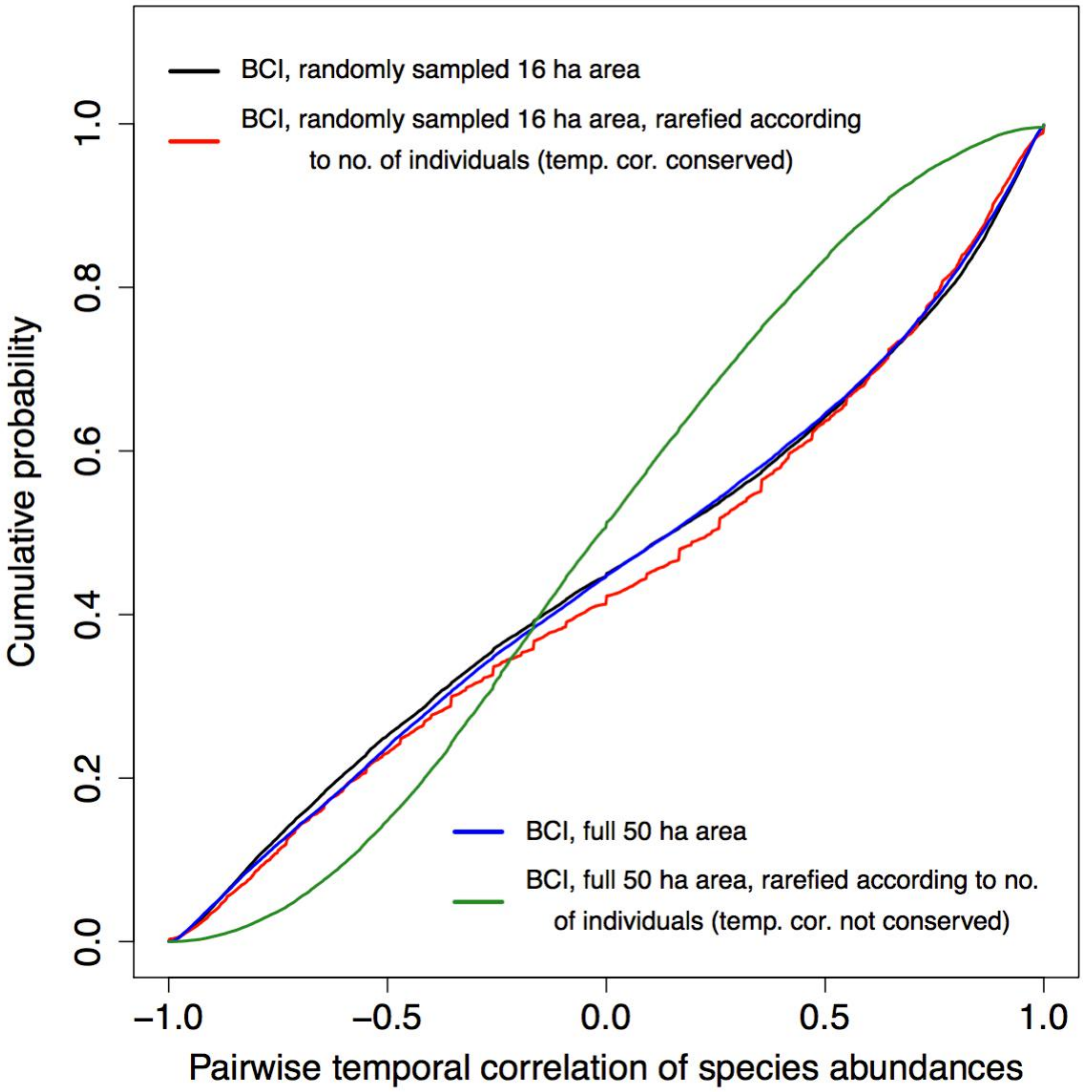


Fig. S7. Comparison of cumulative distribution functions (cdfs) of temporal correlations of species abundances at Barro Colorado Island (BCI) plot, before and after rarefaction according to the number of individuals. Rarefaction was performed with and without conserving temporal correlations of species abundances.

When calculating $\overline{\Delta N_{ra,c}}$, there were 16 initial abundances common to all 21 plots, ranging from 1 to 52. The upper limit of 52 was small, with the main reasons being the low number of individuals at Mudumalai plot, which reduced the common sample size for all plots, and the lack of random sampling over space for the three 16 ha plots (Luquillo, Palanan and SERC). Thus, we also calculated $\overline{\Delta N_{ra,c}}$ using the 20 plots that excluded Mudumalai and using the 17 plots that excluded Mudumalai, Luquillo, Palanan and SERC. Excluding Mudumalai resulted in 20 initial abundances common to the remaining 20 plots, ranging from 1 to 75. Excluding Mudumalai, Luquillo, Palanan and SERC resulted in 194 initial abundances common to the remaining 17 plots, ranging from 1 to 355.

Appendix S4: Further details on regression analyses

Regression of tree species richness and absolute latitude for forest plots used

A recent study by Ricklefs & He (2016) used tree census data from 47 CTFS–ForestGEO plots and found that tree species richness declined with absolute latitude, from the tropics to the poles (their Fig. S3). However, the study did not account for the different number of individuals in each plot. Therefore, in our study, we verified whether tree species richness declined with absolute latitude for the 21 CTFS–ForestGEO plots that we used. Specifically, the rarefied species richness for plot i, $S_{r,i}$, was calculated as the average rarefied number of species over all censuses apart from the last. The last census was not considered because species that appeared for the first time in the last census were not used to calculate temporal population variability, as this calculation required the abundances of a species over two consecutive censuses. The rarefied species richness for census j at plot i, $S_{r,i,j}$, was estimated as the average number of species for census j over the 1,000 resampled datasets for plot i (using the rarefaction method that does not conserve temporal correlations of species abundances; Hurlbert 1971). Rarefying by sample coverage (Chao & Jost 2012) instead of sample size produced species richness values for each plot that were highly correlated with $S_{r,i}$ ($R^2 = 0.998$, $n = 21$), and hence we only used $S_{r,i}$. We subsequently regressed $S_{r,i}$ against the absolute latitude for each plot. In this regression, a log-transformation was applied to both variables to reduce their skewness and help to meet the assumption of normality and homoscedasticity (see subsection *Other details on regression analyses* below). The subsequent regression showed a strong negative correlation between $S_{r,i}$ and absolute latitude ($R^2 = 0.533$, slope = –0.537, $P = 1.73\times10^{-4}$, $n = 21$; Figs. S8, S11, S12).

Regressions using climate data

We performed multiple regressions to determine how much variation in temporal population variability in the 21 forest plots could be explained by temporal variability in two simple climate variables: mean monthly temperature, *MT*, and total monthly precipitation, *MP*. We examined variation in monthly instead of annual climate variables because the former captures climatic variation over both monthly and annual timescales. Specifically we estimated the coefficient of variation (CV) of *MT* (in units of K) and *MP* (in units of mm) for the census periods of each plot, using local measurements of *MT* and *MP* where possible and using measurements of *MT* and *MP* from the CRU dataset (CRU RS v.3.23, https://crudata.uea.ac.uk/cru/data/hrg/) otherwise. The CRU dataset is at a coarse spatial resolution of 0.5^{o} by 0.5^{o} (approximately 300,000 ha). Below we detail which plots had local data available for calculation of the CV of *MT* and *MP* during the census periods, and our rationale for using data from the CRU dataset for calculation of the CV of *MT* and *MP* for the remaining plots.

Local measurements of *MT* were available for six of the 21 plots (Table S3) and local measurements of *MP* were available for eight plots (Table S3). To assess how well *MT* and *MP* values from the CRU dataset reflected local values, we compared values of *MT* and *MP* from the local data with those from the CRU dataset. The mean *MT* values calculated for the six plots using the local data exhibited a very high correlation ($R^2 = 0.94$) with the corresponding mean *MT* values calculated using *MT* time-series extracted from the CRU dataset, covering the same time period. Similarly, the coefficient of variation (CV) of *MT* values calculated for the six plots using local data exhibited a very high correlation ($R^2 =$

1.00) with the corresponding CV of *MT* values calculated using the CRU dataset. We also found that mean *MP* values calculated for the eight plots using the local data exhibited a high correlation ($R^2 = 0.56$) with the corresponding mean *MP* values calculated using the CRU dataset. However, the CV of *MP* values calculated using the local data exhibited a low correlation ($R^2 = 0.27$) with the corresponding CV of *MP* values calculated using the CRU dataset. The main reason was that the local variability in precipitation at the Fushan plot was much higher than the variability in precipitation in the 0.5° by 0.5° grid cell containing the Fushan plot, because of high local topographic heterogeneity combined with the presence of typhoons (standard deviation in *MP* was about three times greater, with the CV of *MP* being 35% greater). Excluding the Fushan plot, the CV of *MP* values calculated using local data for the remaining seven plots exhibited a high correlation ($R^2 = 0.66$) with the corresponding CV of *MP* values for the CRU dataset.

We also assessed the correspondence between values of *MT* and *MP* from the WorldClim database (Hijmans *et al.* 2005) and those from the CRU dataset. Data from the WorldClim database (Hijmans *et al.* 2005) is at a resolution of 0.00833° by 0.00833° (approximately 100 ha), which is smaller than the resolution of data from the CRU database. Specifically, we calculated the mean *MT* values for all 21 plots in our study using 1951–2000 time-series of *MT* values from the WorldClim database and the CRU dataset. The period of 1951–2000 was chosen because the WorldClim database only had data for this period. We found that the two sets of mean *MT* values were highly correlated ($R^2 = 0.99$). Similarly, mean *MP* values for all 21 plots calculated using the CRU dataset and the WorldClim database were highly correlated ($R^2 = 0.96$).

Thus, overall, there is evidence that *MT* and *MP* values from the CRU dataset are generally representative of those measured at smaller scales, including those obtained locally. Therefore, in our multiple regression analysis, we used *MT* and *MP* values from the CRU dataset when these were not available from local sources. The SERC plot was the exception: we used values from the CRU dataset even though local data was available. This was because the local data only covered 15.1% of the total tree census period (Table S3).

The plot at Lenda is approximately 20 km away from the plot at Edoro in the same country (Democratic Republic of the Congo), and this distance is smaller than the spatial scale of the CRU climate data for these two plots. Thus, the Lenda plot was removed from the multiple regression analyses to ensure statistical independence of data points. Removal of the Edoro plot instead of the Lenda plot produced similar results. The time-series of climate for each of the 20 remaining plots were used to calculate the CV of *MT* ($\mathrm{CV}(MT)$) and of *MP* ($\mathrm{CV}(MP)$) for each of the 20 plots. These were used as explanatory variables in a multiple regression, with the response variable being our measure of temporal population variability that involved rarefaction to standardize the number of individuals ($\overline{\Delta N_{r,c}}$). All three variables (the two explanatory variables and the response variable) were log-transformed because this reduced the skewness of each variable, which helped to meet the assumptions of normality and homoscedasticity of residuals for a multiple regression. The two explanatory variables $\ln(\mathrm{CV}(MT))$ and $\ln(\mathrm{CV}(MP))$ were only weakly correlated ($R^2 = 0.153$) and therefore met the assumption of independence of explanatory variables. Results from this multiple regression are presented in Figs. S9, S11 and S12, and Table S4. We also did a multiple regression using our measure of temporal population variability that involved rarefaction to standardize area and the number of individuals, in a way that conserved temporal correlations of species abundances ($\overline{\Delta N_{ra,c}}$; Figs. S10–S12, Table S5). In this regression, we excluded

Mudumalai, Luquillo, Palanan and SERC in order to increase the number of initial abundances used to calculate $\overline{\Delta N_{ra,c}}$ (see Appendix S3 for details).

Other details on regression analyses

In our study, we performed a total of five bivariate linear regressions: (i) temporal population variability ($\overline{\Delta N_{r,c}}$) regressed against absolute latitude (Fig. 4A); (ii) temporal population variability ($\overline{\Delta N_{ra,c}}$) regressed against absolute latitude, using all 21 forest plots (Fig. 4B); (iii) $\overline{\Delta N_{ra,c}}$ regressed against absolute latitude, using the 20 forest plots that excluded Mudumalai (Fig. 4C); (iv) $\overline{\Delta N_{ra,c}}$ regressed against absolute latitude, using the 17 forest plots that excluded Mudumalai, Luquillo, Palanan and SERC (Fig. 4D); and (v) rarefied tree species richness (S_r) regressed against absolute latitude (Fig. S8). In each of these five regressions, a log-transformation was applied to all variables, to reduce their skewness and help to meet assumptions of normality and homoscedasticity. Normality of residuals was assessed graphically using quantile plots (Fig. S11) and using a Shapiro–Wilk test (P = 0.756, 6.87×10^{-4}, 0.428, 0.715 and 0.759 for the five regressions, respectively). The quantile plots and Shapiro–Wilk tests indicated normality of residuals except for the second regression. Homoscedasticity of residuals was confirmed graphically using residual plots for the five regressions (Fig. S12). In addition, independence of residuals was assessed for all regressions using a 2-sided Durbin–Watson test (P = 0.153, 0.718, 0.124, 0.111 and 0.0450, respectively). This test indicated independence of residuals for all regressions apart from the fifth regression, for which there was a marginal indication of non-independent residuals.

We also performed two multiple linear regressions, one that used $\overline{\Delta N_{r,c}}$ as the response variable and one that used $\overline{\Delta N_{ra,c}}$ as the response variable (see subsection *Regression using climate data* above). For these multiple regressions, normality of residuals was confirmed graphically (Fig. S11) and using a Shapiro–Wilk test (P = 0.452 and 0.327, respectively), whereas homoscedasticity of residuals was confirmed graphically (Fig. S12). Independence of residuals was confirmed using a 2-sided Durbin–Watson test (P = 0.971 and 0.710, respectively).

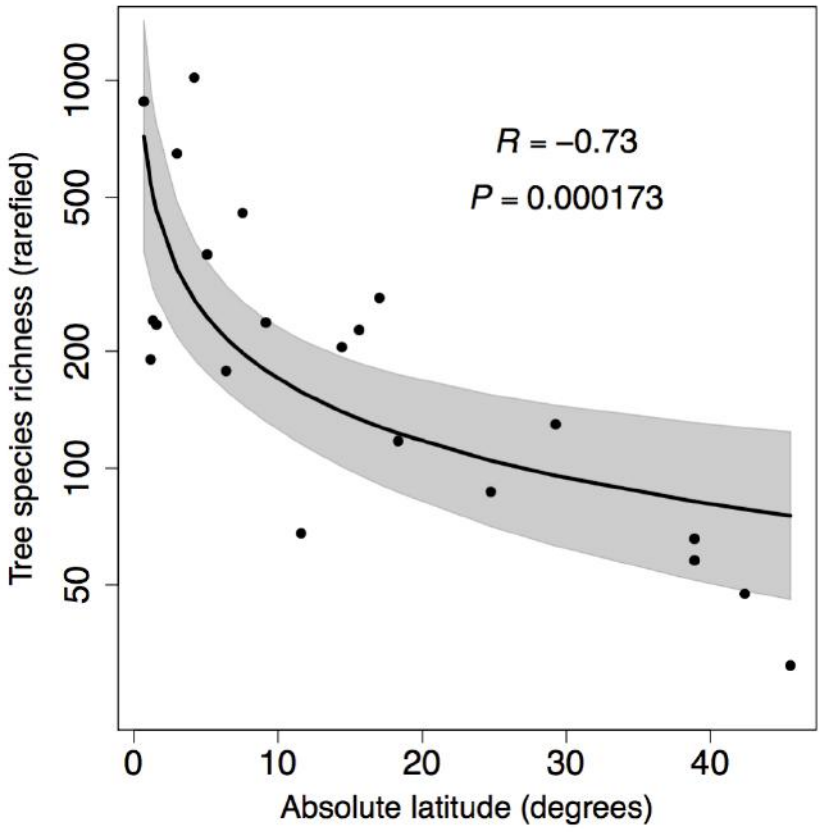


Figure S8. Linear regression between rarefied tree species richness and absolute latitude for the 21 CTFS–ForestGEO forest plots considered in this study. The shaded region represents the 95% confidence interval.

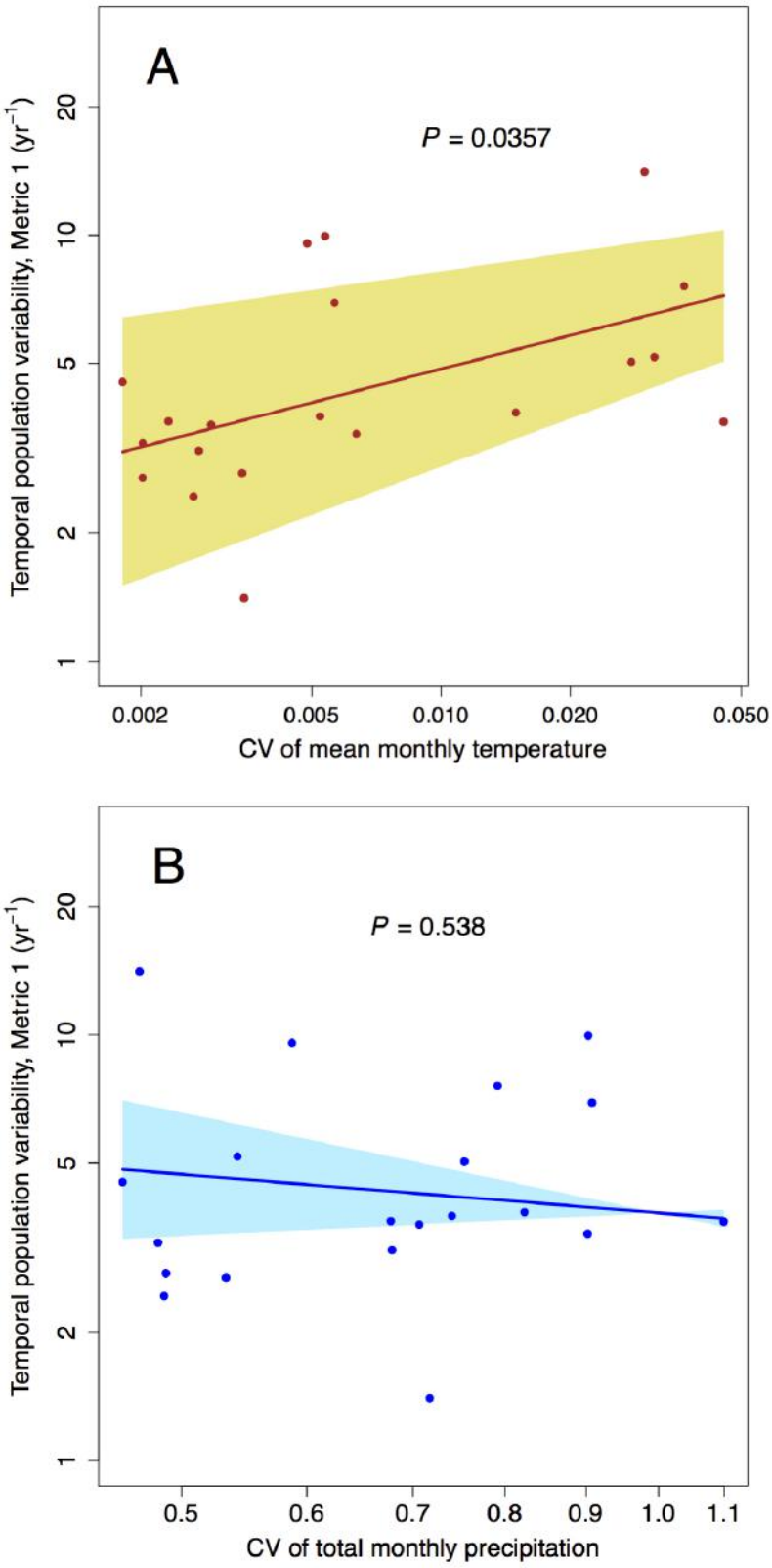


Figure S9. Relationships between temporal population variability and **(A)** the coefficients of variation of mean monthly temperature (CV(MT)) and **(B)** of total monthly precipitation (CV(MP)) for 20 CTFS–ForestGEO forest plots. Temporal population variability was measured as the mean absolute change in species abundance per year, with rarefaction to standardize the number of individuals ($\overline{\Delta N_{r,c}}$). The solid lines represent values from a

multiple regression model of $\ln\left(\overline{\Delta N_{r,c}}\right)$ against $\ln(\mathrm{CV}(MT))$ and $\ln(\mathrm{CV}(MP))$, fitted to the empirical values represented by the dots. In drawing the relationship between $\ln\left(\overline{\Delta N_{r,c}}\right)$ and one climate variable, the value of the other climate variable was fixed at the mean observed value. The shaded regions represent standard errors in the coefficient of the changing climate variable.

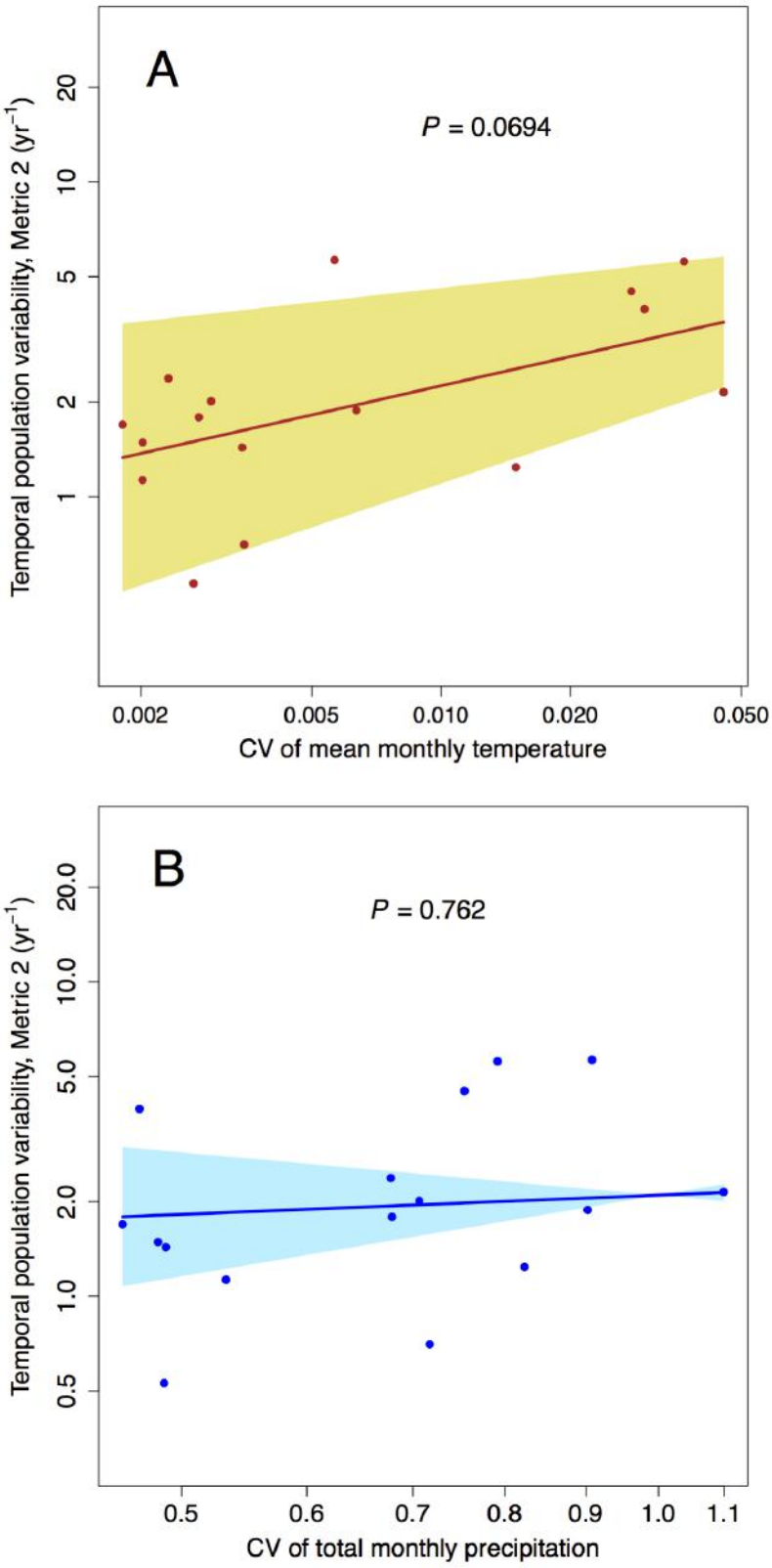
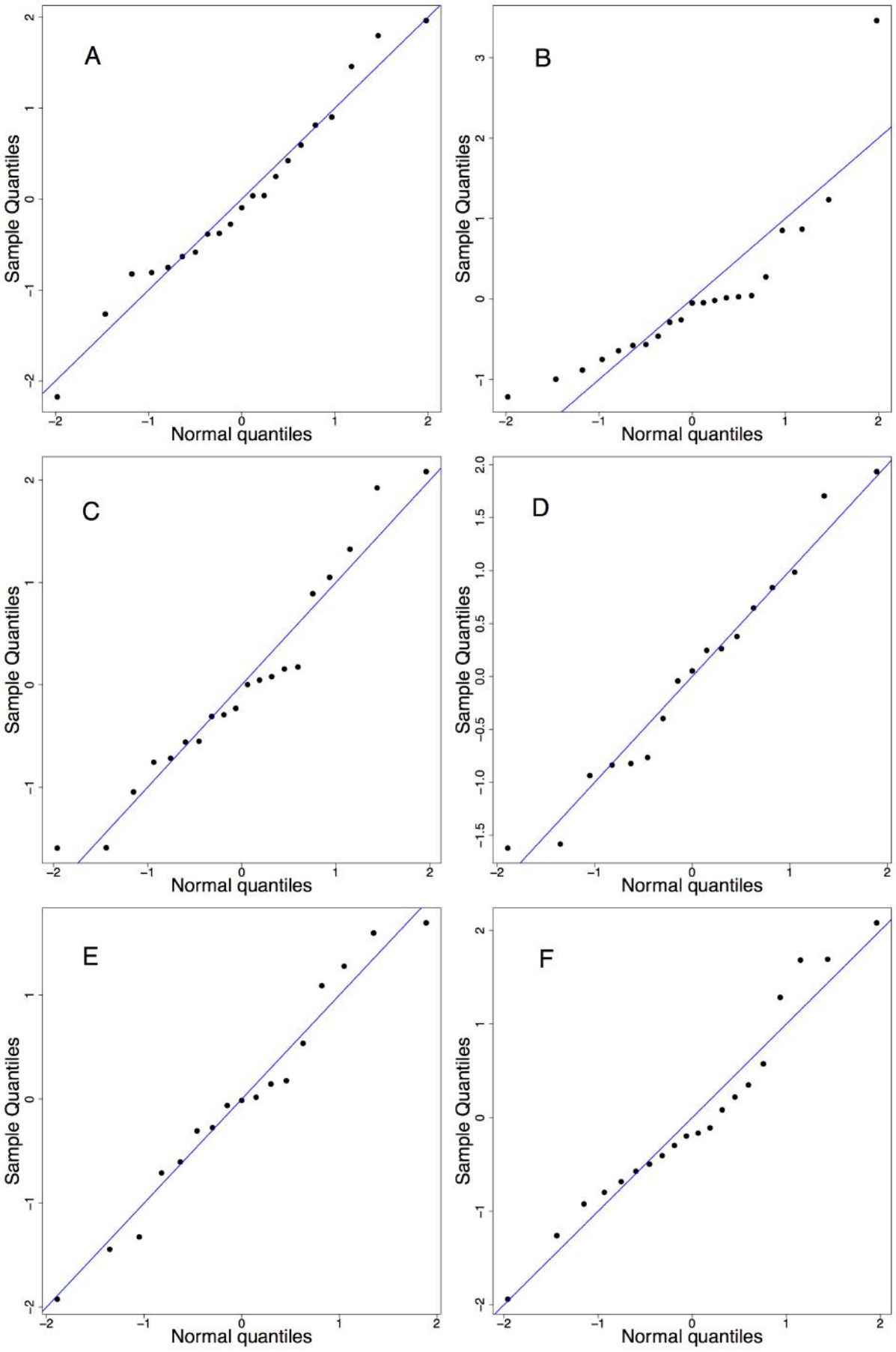


Figure S10. Relationships between temporal population variability and **(A)** the coefficients of variation of mean monthly temperature (CV(MT)) and **(B)** of total monthly precipitation (CV(MP)) for 16 CTFS–ForestGEO forest plots. Temporal population variability was measured as the mean absolute change in species abundance per year, with rarefaction to standardize area and the number of individuals, in a way that conserved temporal correlations of species abundances ($\overline{\Delta N_{ra,c}}$). Other details as in Fig. S9.


A
B
C
D
E
F
Sample Quantiles
Normal quantiles

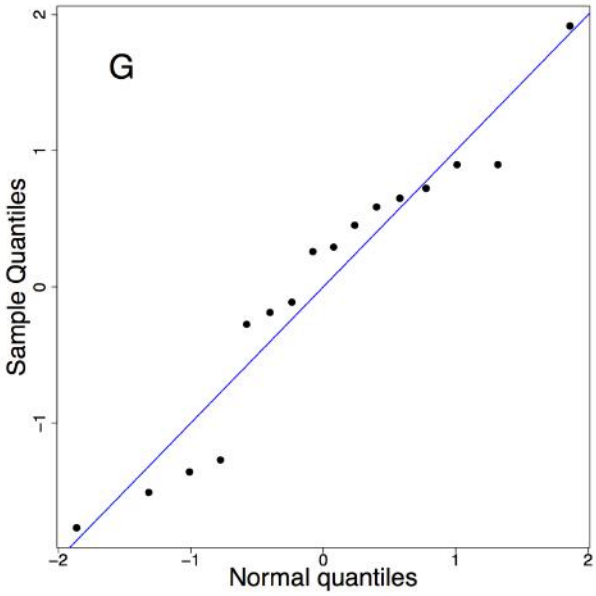
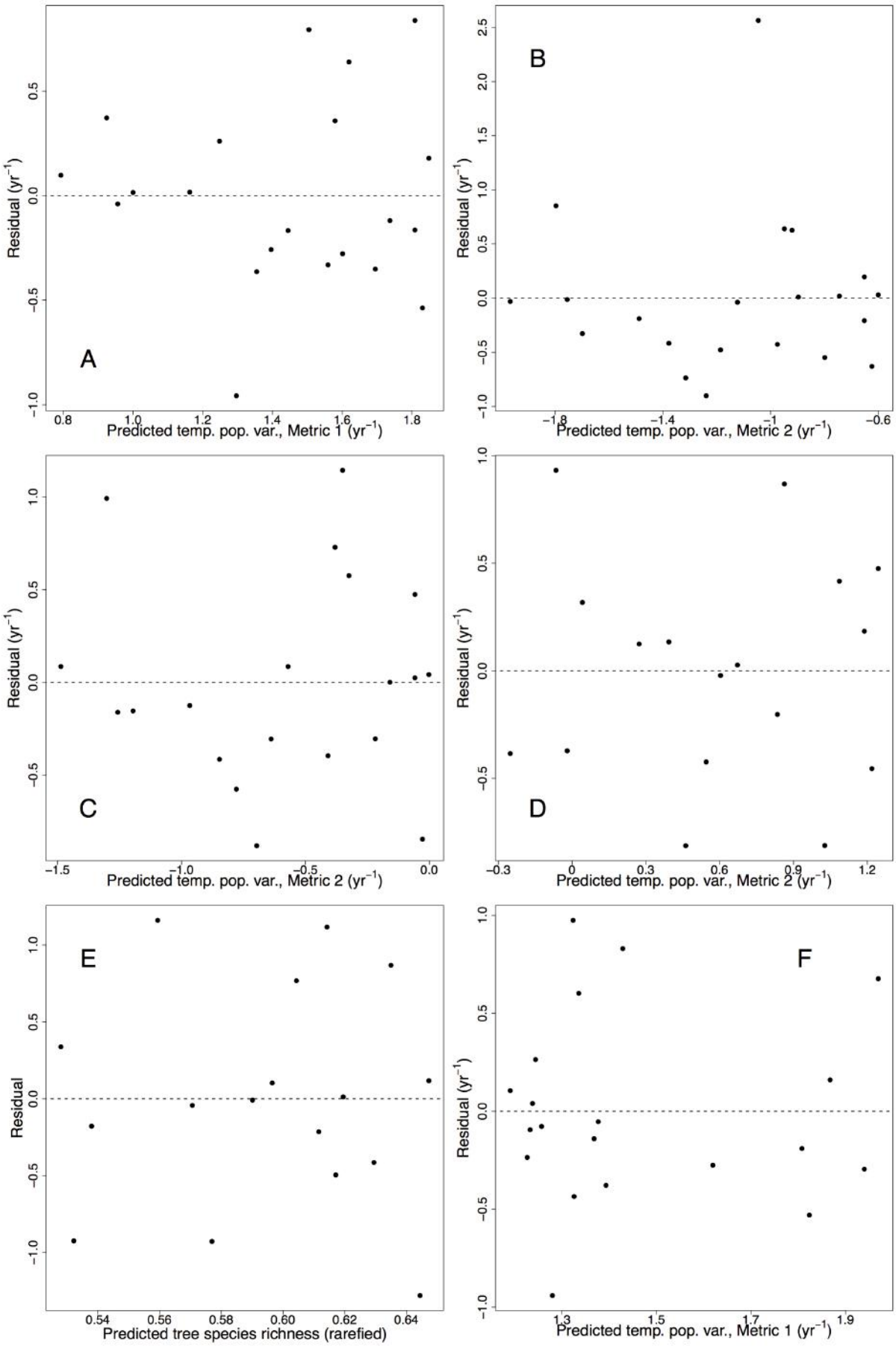


Figure S11. Quantile plots for the standardized residuals of the regressions in **(A)** Fig. 4A, **(B**) Fig. 4B, **(C)** Fig. 4C, **(D)** Fig. 4D, **(E)** Fig. S8, **(F)** Fig. S9 and **(G)** Fig. S10.


A
B
C
D
E
F
Residual (yr^{-1})
Residual
Predicted temp. pop. var., Metric 1 (yr^{-1})
Predicted temp. pop. var., Metric 2 (yr^{-1})
Predicted tree species richness (rarefied)

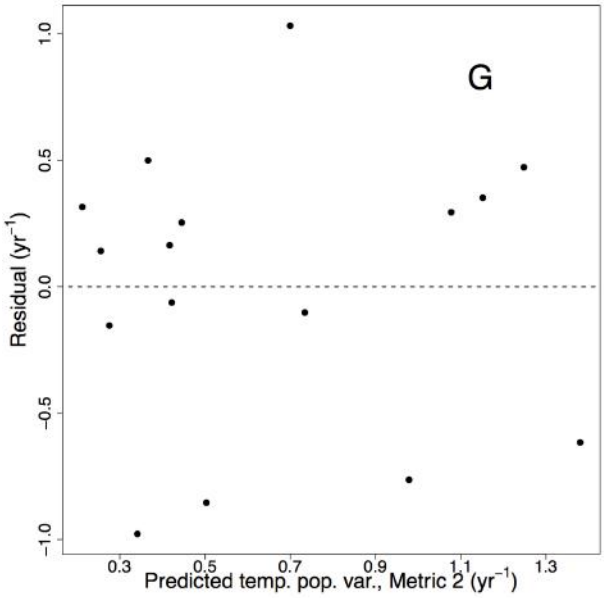


Figure S12. Residual plots for the regressions in **(A)** Fig. 4A, **(B**) Fig. 4B, **(C)** Fig. 4C, **(D)** Fig. 4D, **(E)** Fig. S8, **(F)** Fig. S9 and **(G)** Fig. S10.

Table S3. Details of local climate data for eight of the 21 CTFS–ForestGEO plots used in the study. *MT* and *MP* stand for "Mean monthly temperature" and "Total monthly precipitation", respectively. SERC stands for Smithsonian Environmental Research Center. The coverage of the total census period was calculated as the percentage of census months for which local *MT* or *MP* measurements were available. Climate data for Barro Colorado Island was collected as part of the STRI Physical Monitoring Program and can be downloaded at http://biogeodb.stri.si.edu/physical_monitoring/research/barrocolorado; climate data for Luquillo can be downloaded at http://luq.lternet.edu/data/luqmetadata14; and climate data for Wabikon can be downloaded at https://www.ncei.noaa.gov/.

Forest plot	Type of climate data	Location where climate data was collected	Coverage of total tree census period
Barro Colorado Island	*MT*, *MP*	Clearing adjacent to plot	92.7% for *MT*, 98.2% for *MP*
Fushan	*MT*, *MP*	Fushan Botanical Garden, approx. 3 km east of plot	100%
Gutianshan	*MT*, *MP*	Kaihua Meteorological Station, approx. 50 km from plot	100%
Luquillo	*MP*	El Verde field station, a few hundred m from plot	100%
Mo Singto	*MT*, *MP*	Khao Yai National Park, approx. 0.8 km east of plot	57.1% for *MT*, 100% for *MP*
SERC	*MT*, *MP*	Clearing adjacent to plot	15.1%
Sinharaja	*MP*	Sinharaja field research station, approx. 1 km northeast of plot	98%
Wabikon	*MT*, *MP*	Laona WI weather station, approx. 6 km from plot	78.2%

Table S4. Multiple regression of temporal population variability against two measures of climatic variability, for 20 CTFS–ForestGEO forest plots ($F_{2,17} = 2.64$, $P = 0.101$, $R^2 = 0.237$). Temporal population variability was measured as the mean absolute change in species abundance per year, with rarefaction to standardize the number of individuals ($\overline{\Delta N_{r,c}}$). Log-transformations were applied to all three variables to meet normality and homoscedasticity assumptions.

Response variable: Log of mean absolute change in species abundance per year ($\overline{\Delta N_{r,c}}$)			
	Coefficient	**s.e.**	***P***
Intercept	2.67	0.546	1.43×10^{-4}
Log of CV of mean monthly temperature	0.262	0.115	0.0357
Log of CV of total monthly precipitation	–0.303	0.481	0.538

Table S5. Multiple regression of temporal population variability against two measures of climatic variability, for 16 CTFS–ForestGEO forest plots ($F_{2,13} = 3.19$, $P = 0.0748$, $R^2 = 0.329$). Temporal population variability was measured as the mean absolute change in species abundance per year, with rarefaction to standardize area and the number of individuals, in a way that conserved temporal correlations of species abundances ($\overline{\Delta N_{ra,c}}$). Log-transformations were applied to all three variables to meet normality and homoscedasticity assumptions.

Response variable: Log of mean absolute change in species abundance per year ($\overline{\Delta N_{ra,c}}$)			
	Coefficient	**s.e.**	***P***
Intercept	2.31	0.709	0.00621
Log of CV of mean monthly temperature	0.308	0.156	0.0694
Log of CV of total monthly precipitation	0.202	0.654	0.762

Appendix S5: Further details and results pertaining to simulations of our mechanistic model

For each of the 21 CTFS–ForestGEO forest plots except Mudumalai, we fitted a dynamic mechanistic model (Danino *et al*. 2016) to the observed temporal variability in abundances of tree species populations (temporal population variability) and the observed cumulative distribution function (cdf) of pairwise temporal correlations of abundances of tree species populations. The model was fitted to the value of temporal population variability given by eq. (1), as applied to the data rarefied according to area and the number of individuals, in a way that conserved temporal correlations of species abundances (Appendix S3). We denote this metric by $\overline{\Delta N_{ra}}$ (as opposed to $\overline{\Delta N_{r}}$, which is for the data rarefied according to number of individuals only). We did not fit the model to values of temporal population variability given by the metric that corrected for different sets of (initial) species abundances among plots ($\overline{\Delta N_{ra,c}}$). This was because $\overline{\Delta N_{ra,c}}$ could only be calculated using simulated data from all 20 plots simultaneously, and hence would necessitate searching $448^{20} = 1.06 \times 10^{53}$ parameter combinations to find the best-fit models, which would be computationally infeasible. The observed and model cdfs of temporal correlations were evaluated at 1,000 equally spaced values of temporal correlation ranging from –1 to 1. Time in the model was measured as the number of deaths, so when fitting the model to a rarefied dataset for a plot, the length of time in between a pair of consecutive censuses in the model was defined as the empirically observed number of deaths in between the pair of censuses in the original dataset for the plot (before rarefaction), scaled to the number of individuals in the first census of the pair in the rarefied dataset.

In a simulation of the model for a tree community in a forest plot, the initial fitness values of model species were drawn randomly and independently from a lognormal distribution with mean 1 and variance A (or equivalently, the logarithm of the initial value was drawn randomly from a normal distribution with mean $-0.5\ln(1+A)$ and variance $\ln(1+A)$). More technically, the "fitness value" of a species was proportional to the relative fecundity of the species and hence determined its recruitment rate (see below for further details). We note that as A increased, the quantile of the lognormal distribution at the mean increased from 0.5 and approached 1, such that the expected proportion of species with fitness values above the mean decreased from 0.5 and approached 0, representing a smaller expected proportion of species benefitting from the prevailing environmental conditions at the expense of the remaining species. The fitness values of all species were assumed to vary over time, representing changes in environmental conditions. Specifically, at the beginning of each discrete model time-step, there was a probability $1/\tau$ that the environmental conditions changed, resulting in the fitness values of all species being randomly and independently redrawn from the lognormal distribution. Thus, τ measured the temporal correlation in environmental conditions and we assumed that each species had the same distribution of possible fitness values – a parsimonious assumption that has often been used in community models with changing environmental conditions (e.g., Kalyuzhny *et al*. 2015; Danino *et al.* 2016; Fung *et al*. 2016; Danino & Shnerb 2018).

After determining whether the environment changed or not in a time-step, an individual was randomly chosen to die, with each individual having an equal probability of being chosen. This represented species having an equal mortality rate of d. Therefore, the expected length of each time-step was $1/Jd$, or equivalently $1/J$ generations, where J is the total number of individuals at the start of the time-step. The value of J was assumed to change according to

the rarefied dataset being fitted to – at the census times, J was set equal to the total number of individuals in the rarefied dataset at those times, with linear changes in J in between the censuses. After an individual was chosen to die, X replacement individuals were chosen, with X chosen to satisfy the changes in J in between censuses (X can be zero). A replacement individual was the offspring of an existing individual in the modeled community, representing local recruitment. The local recruit was chosen randomly from the pool of local recruits produced by all species in the modeled community. Specifically, if there were S species and species i had fitness f_i and n_i individuals, then the probability of the recruit belonging to species i was $f_i n_i / \sum_{j=1}^{S} f_j n_j$. If the fitness values of all species are fixed to the same value, then the model becomes a neutral model, a zero-sum version of which has often been used to model forest tree dynamics (Hubbell 2001; Volkov *et al*. 2003, 2007). Therefore, the model we use can be conceptualized as a neutral model parsimoniously extended to include temporal fluctuations in environmental conditions (with the resulting model being non-neutral).

It can be seen that in the model, changes in environmental conditions acted on the recruitment rates of species rather than their mortality rates. This approximation is consistent with empirical evidence from two tropical tree communities at the Barro Colorado Island (BCI) plot in Panama and the Pasoh plot in peninsular Malaysia. For these communities, counting all trees with DBH ≥ 1 cm, the effect of changes in environmental conditions on recruitment was found to be about two to three times greater than on mortality, respectively (Appendix S1 of Fung *et al*. (2016)).

For the tree community in each of the 20 forest plots, we performed simulations for all 448 combinations of 32 values of A and 14 values of τ. The values of A were 0, 0.001, 0.002, 0.004, 0.006, 0.008, 0.01, 0.02, 0.04, 0.06, 0.08, 0.1, 0.2, 0.4, 0.6, 0.8, 1, 2, 4, 6, 8, 10, 20, 40, 60, 80, 100, 200, 400, 600, 800 and 1000. The values of τ were 1, 10, 25, 50, 75, 100, 250, 500, 750, 1000, 2500, 5000, 7500, 10000. We note that although there were two fitting parameters (A and τ), a priori the model was not guaranteed to give good fits. This was because the model was mechanistic and the dynamics that it could have produced were constrained by the mechanisms built into the model together with the parameter space explored.

There were 17 plots with an area greater than 16 ha, and for each of these plots there were 1,000 rarefied datasets resulting from 1,000 randomly sampled 16 ha sub-areas (Appendix S3). For each of the 17 plots, we performed 1,000 simulations for each combination of A and τ, corresponding to the 1,000 rarefied datasets. From these simulations, we calculated the mean temporal population variability and the mean cdf of pairwise temporal correlations of species abundances, and calculated the error as the average of (i) the percentage absolute difference between the model and observed temporal population variability and (ii) the percentage absolute difference between the model and observed cdf of temporal correlations (as described in subsection "Relating temporal population variability to mechanisms maintaining species richness" in the *Materials and methods* section of the main text). Let this error be denoted by $\varepsilon_{1,A,\tau}$, where the subscript "1" refers to the one set of 1,000 simulations used. In general, the error $\varepsilon_{A,\tau}$ from a set of 1,000 simulations is a random variable because the simulations are stochastic. However, the variation in this error is small because the error is calculated using mean quantities over 1,000 simulations. As described below, we verified this small variation in error for 10 of the 17 plots with a good model fit, i.e. the 10 plots with at least one combination of A and τ giving $\varepsilon_{1,A,\tau} < 10\%$ (Table S6). Thus, $\varepsilon_{1,A,\tau}$ can be used as an approximate measure of the typical error. We also found that one plot (Lenda) had a

combination of A and τ giving $\varepsilon_{1,A,\tau}$ very close to 10% (10.1%; Table S6), which we refer to as a "marginally good fit".

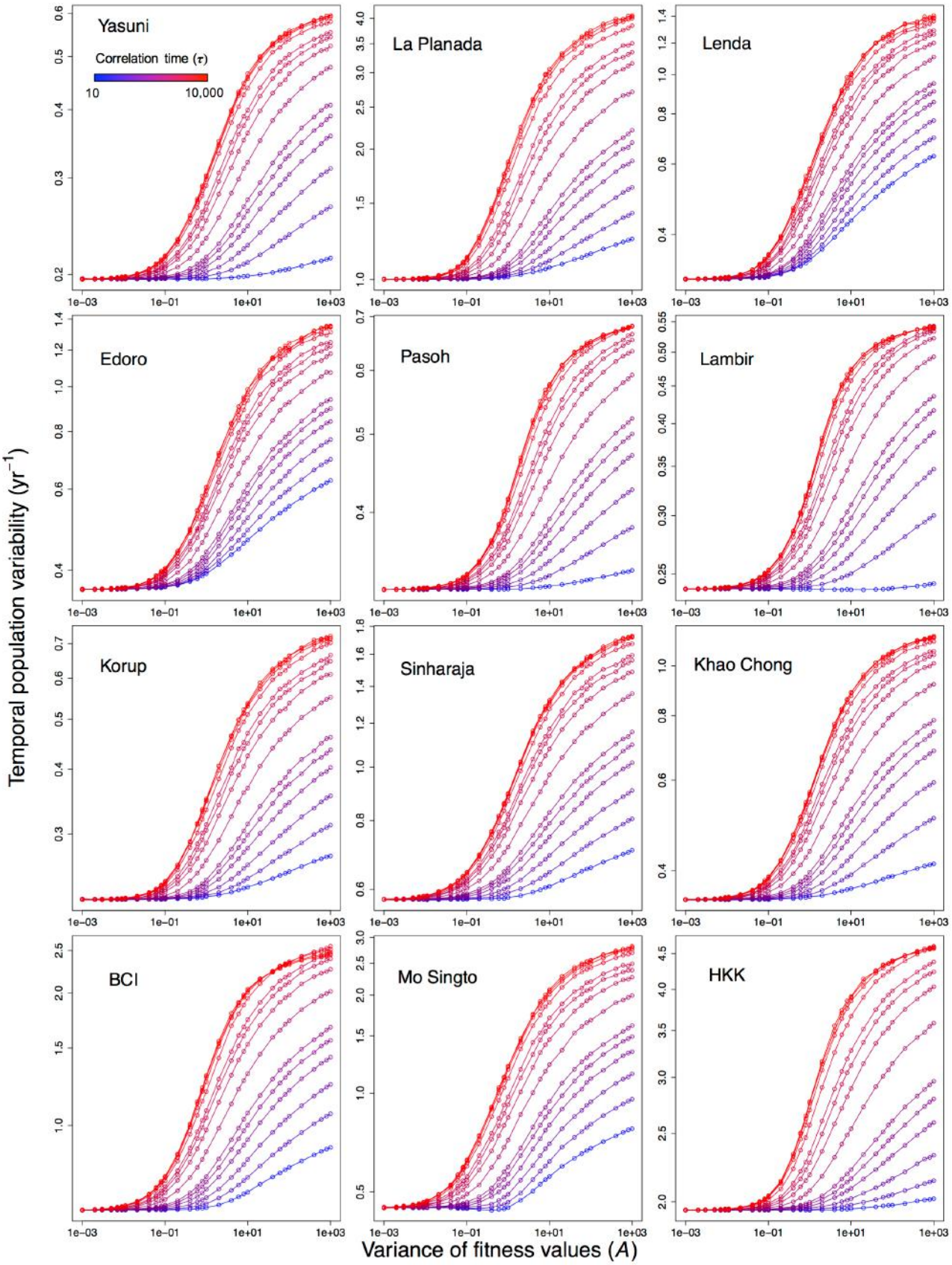
There were three plots with area 16 ha, and for each of these plots there was only one rarefied dataset, because there was no need to randomly sample 16 ha sub-areas (Appendix S3). For each of these three plots, it is possible to perform one simulation for each combination of A and τ, corresponding to the single rarefied dataset, and then calculate the error $\varepsilon_{1,A,\tau}$. However, the variation in $\varepsilon_{A,\tau}$ is no longer small because it is calculated using only one simulation. Thus, $\varepsilon_{1,A,\tau}$ cannot be used as an approximate measure of the typical error. So instead, for each of the three plots and each combination of A and τ, we ran 1,000 simulations for the rarefied dataset, to produce a distribution of realized errors (i.e., realized values of $\varepsilon_{A,\tau}$). We then used the median of these realized errors, $\varepsilon_{j_{50},A,\tau}$, as a measure of the typical error. Thus, for a particular combination of A and τ, if simulations for a plot have $\varepsilon_{j_{50},A,\tau} < 10\%$, then this was considered to represent a good model fit. We found that one of the three plots (Palanan) had at least one combination of A and τ giving $\varepsilon_{j_{50},A,\tau} < 10\%$ (Table S6), and one plot (Luquillo) had a combination giving $\varepsilon_{j_{50},A,\tau}$ very close to 10% (10.5%; Table S6), i.e. a marginally good fit.

Fig. S13 shows the temporal population variability from the model for all 20 plots, across all combinations of A and τ. For the 17 plots with area greater than 16 ha, the temporal population variability plotted is the mean value of $\overline{\Delta N_{ra,c}}$ across the 1,000 simulations using 1,000 rarefied datasets. For the three plots with area 16 ha, the variability plotted is the median value of $\overline{\Delta N_{ra,c}}$ across the 1,000 simulations for one rarefied dataset. Fig. S14 shows the cdfs of temporal correlations from the model for two plots, for combinations of A and τ minimizing the typical error.

Table S6. For the dynamic mechanistic model that we fitted to tree census data from each of 20 CTFS–ForestGEO plots, this table shows the minimum value of the error metric across all combinations of model parameter values tested. This error metric is the average of the percentage absolute error in temporal population variability and the percentage absolute error in the cumulative distribution function (cdf) of temporal correlations of species abundances from the model (corresponding to $\varepsilon_{1,A,\tau}$ for the 17 plots with area greater than 16 ha and $\varepsilon_{j_{50},A,\tau}$ for the three plots with area 16 ha; see text in Appendix S5 for details). The modeled and observed temporal population variability corresponding to the minimum value of the error metric are also shown, together with the percentage error in the modeled temporal population variability and the percentage error in the cdf of temporal correlations of species abundances from the model. SCBI and SERC stand for Smithsonian Conservation Biology Institute and Smithsonian Environmental Research Center, respectively.

Forest plot	**Average of % absolute errors in modeled temporal population variability and cdf**	**Modeled temporal population variability (yr^{-1})**	**Observed temporal population variability (yr^{-1})**	**% error in modeled temporal population variability**	**% error in modeled cdf of temporal correlations of species abundances**
Barro Colorado Island	3.125 %	0.891	0.898	–0.827 %	5.424 %
Changbaishan	2.217 %	3.139	3.124	0.472 %	3.962 %
Edoro	2.098 %	0.497	0.497	–0.079 %	4.117 %
Fushan	1.798 %	2.422	2.429	–0.297 %	3.298 %
Gutianshan	17.488 %	3.077	3.244	–5.146 %	29.830 %
Huai Kha Khaeng	2.263 %	2.604	2.610	–0.235 %	4.291 %
Khao Chong	2.654 %	0.376	0.376	–0.076 %	5.232 %
Korup	0.824 %	0.291	0.286	1.640 %	0.009 %
La Planada	0.611 %	1.404	1.407	–0.212 %	1.010 %
Lambir	12.612 %	0.237	0.193	23.011 %	2.213 %
Lenda	10.071 %	0.358	0.359	–0.394 %	19.748 %
Luquillo	10.542 %	8.444	8.761	–3.616 %	17.469 %
Mo Singto	1.561 %	0.976	0.969	0.753 %	2.369 %

Palanan	5.051 %	1.193	1.219	–2.160 %	7.942 %
Pasoh	14.203 %	0.321	0.281	14.064 %	14.342 %
SCBI	6.661 %	10.348	10.396	–0.465 %	12.858 %
SERC	11.277 %	4.000	4.441	–9.932 %	12.622 %
Sinharaja	121.098 %	0.574	0.419	37.075 %	205.121 %
Wabikon	14.249%	12.445	12.454	–0.069 %	28.428 %
Yasuni	17.667 %	0.196	0.148	32.389 %	2.943 %


Yasuni
Correlation time (τ)
10
10,000
La Planada
Lenda
Edoro
Pasoh
Lambir
Korup
Sinharaja
Khao Chong
BCI
Mo Singto
HKK
Temporal population variability (yr⁻¹)
Variance of fitness values (A)
1e−03
1e−01
1e+01
1e+03

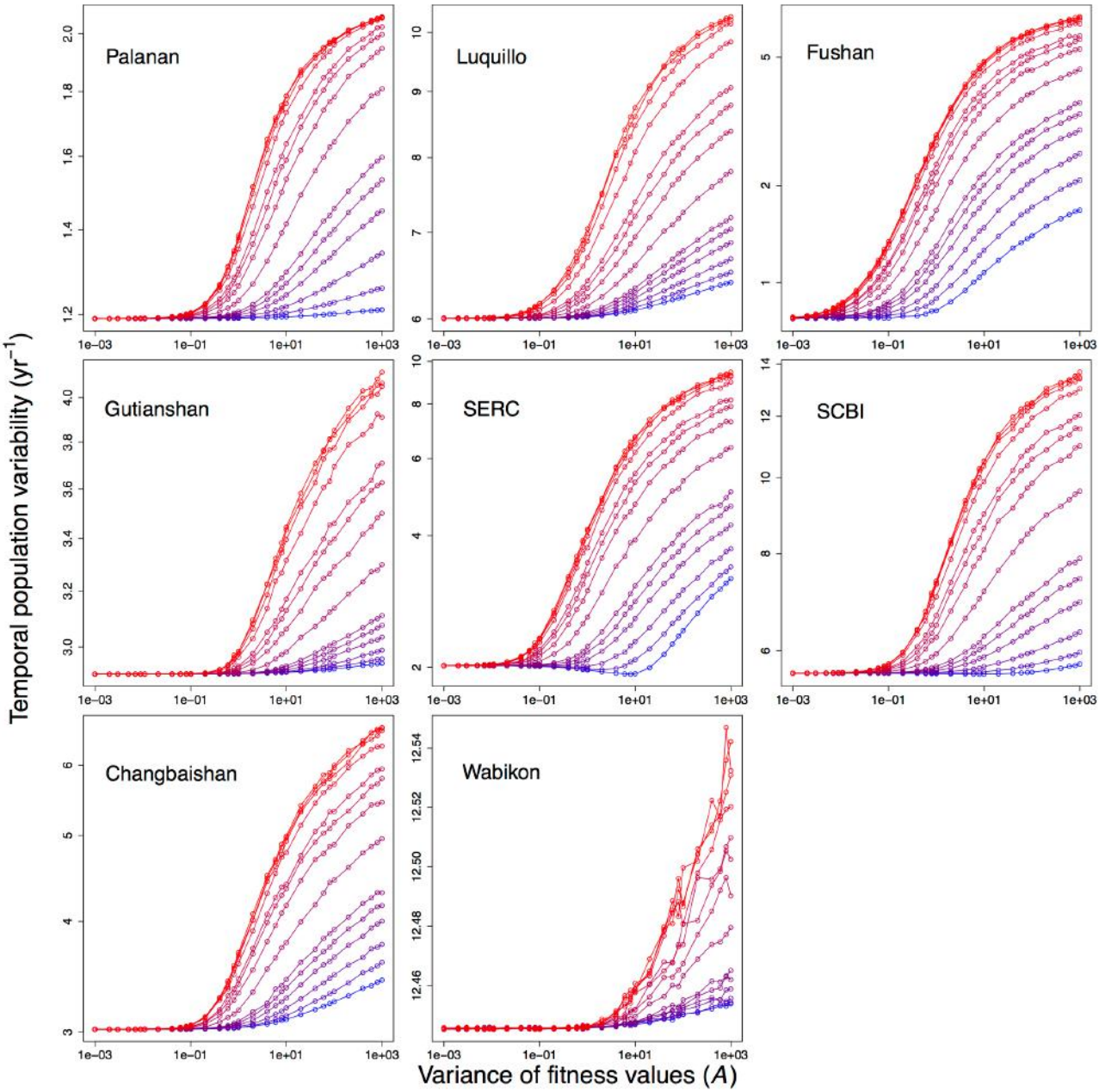


Figure S13. Values of temporal population variability ($\overline{\Delta N_{ra}}$) produced by the mechanistic model for each plot, across different combinations of two parameter values. The two parameters are: the variance of the lognormal distribution of possible fitness values for each model species (A) and the correlation time determining how frequently the fitness values of all species were redrawn due to changes in environmental conditions (τ, given in the figure as the absolute number of discrete model time-steps). The plots are ordered from left to right according to increasing absolute latitude.

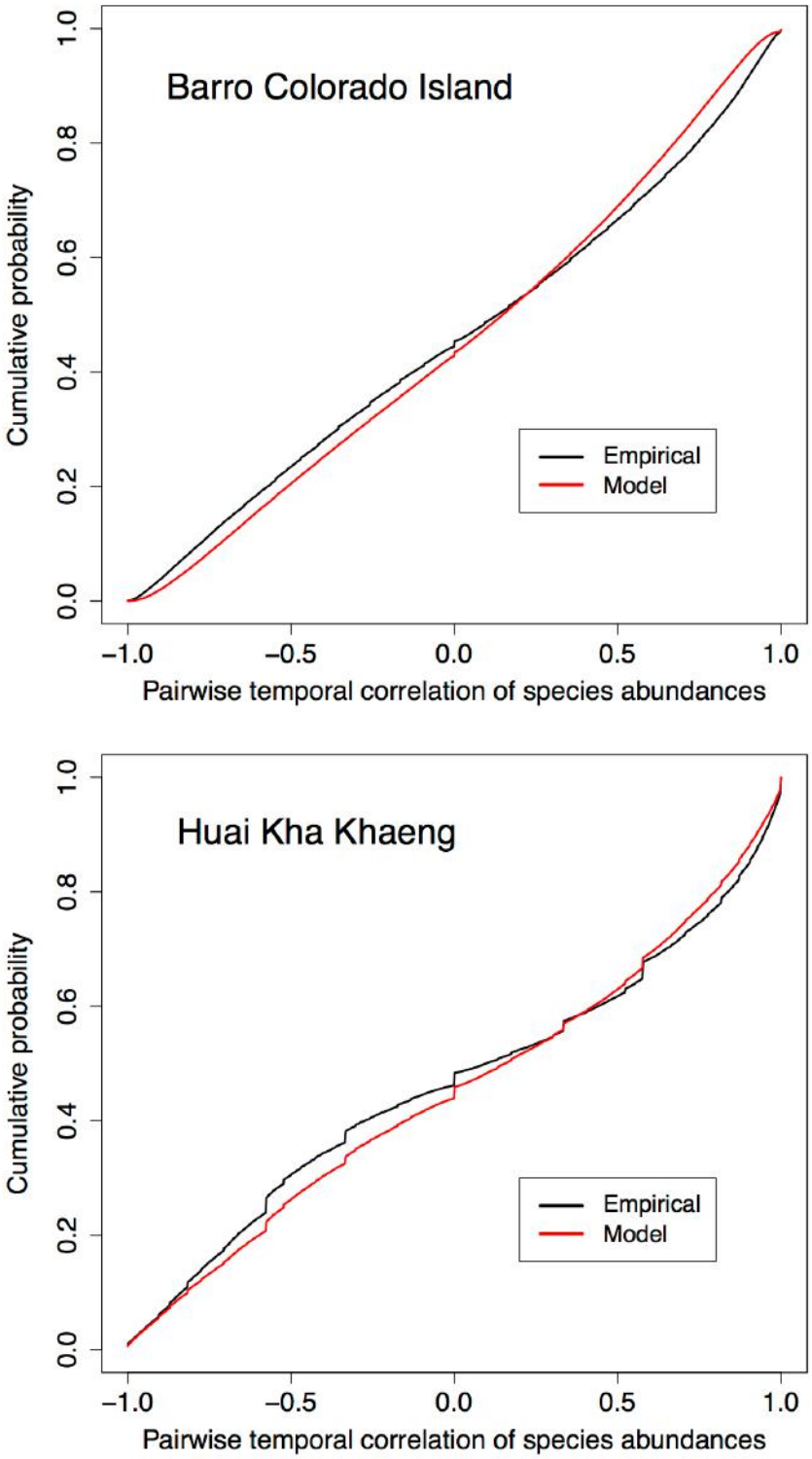


Figure S14. Cumulative distribution functions (cdfs) of temporal correlations of species abundances from the models for two plots, using the parameter values minimizing the error ($\varepsilon_{1,A,\tau}$; see text in Appendix S5 for details). For each plot, the model cdf shown was calculated as the average of 1,000 model cdfs derived by simulating the model using data from 1,000 rarefied datasets. The empirical cdf shown was calculated as the average of the 1,000 empirical cdfs from the 1,000 rarefied datasets.

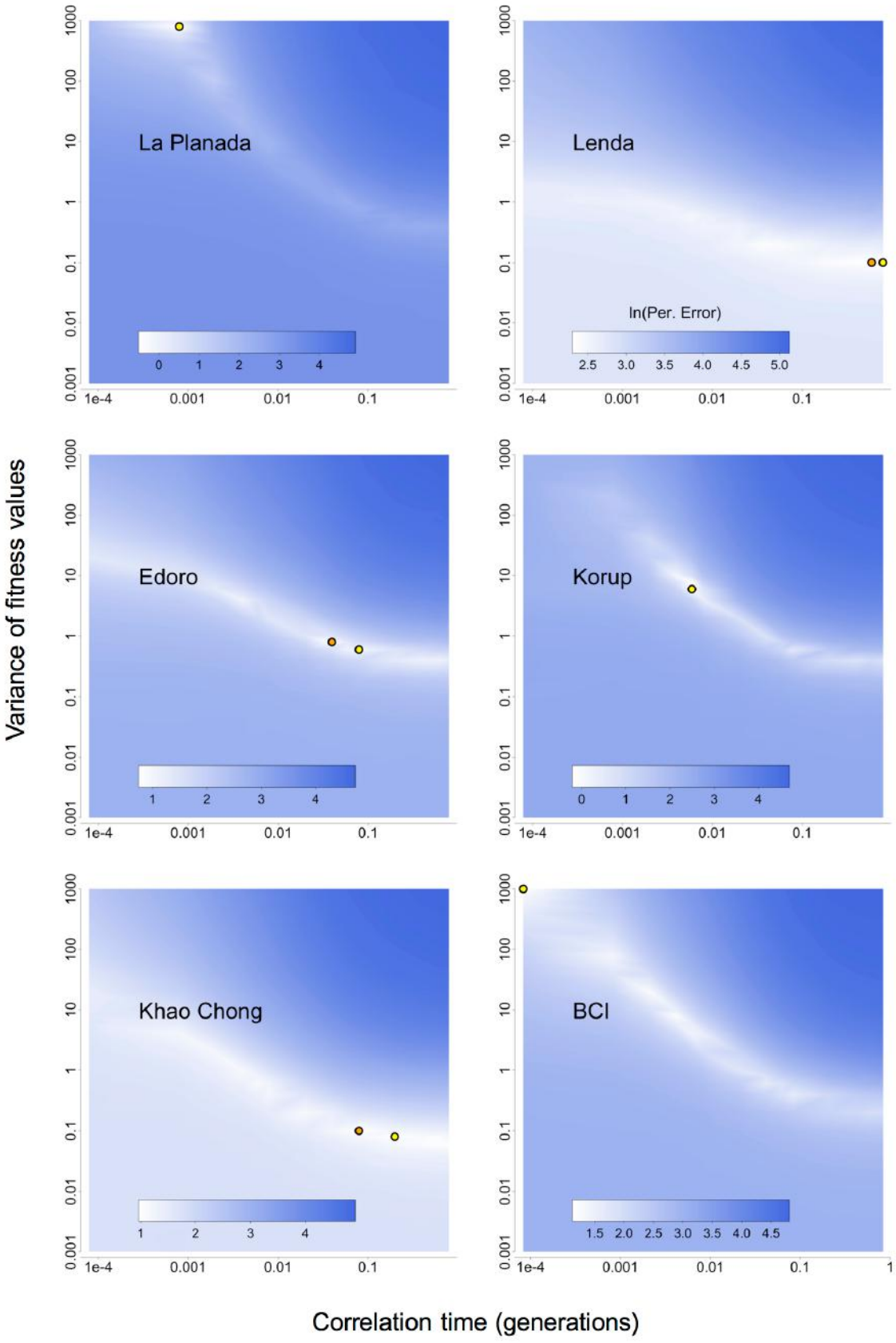
We assessed the error in our model according to two observed quantities, which were the observed temporal population variability and the observed cdf of temporal correlations. These two quantities were captured in the error metric that we used ($\varepsilon_{1,A,\tau}$ for the 17 plots with area greater than 16 ha and $\varepsilon_{j_{50},A,\tau}$ for the three plots with area 16 ha), as described above. Values of this error metric across the different combinations of A and τ indicated that for each of the 11 plots with good model fits (as defined above), there was a combination of A and τ that minimized the error, but that there were often multiple combinations giving similar errors (Figs. S15 and S16). This was partly because the model produced similar values of the mean temporal population variability for multiple combinations of A and τ (Fig. S13). Therefore, when examining which parameter regime the fitted model for a plot lies within, it was instructive to examine not only the combination of A and τ minimizing the error for a plot, but also the combinations of A and τ that gave errors close to the minimum. We determined such combinations of A and τ for the 11 plots for which the model gave good fits. An issue here is how to quantify the similarity of errors produced by two combinations of A and τ. If we had the likelihood of the model producing the observed quantities with a given combination of A and τ, we would be able to use the likelihood ratio to assess the relative error from two combinations of A and τ. A likelihood ratio of 1/4 corresponds to an absolute AIC difference of 2.77, which can be used to indicate that two models are considerably different (Burnham and Anderson, 2002). In principle, our model was able to produce the observed quantities for any combination of A and τ, but in practice this was extremely unlikely. Thus, it was not feasible to estimate this likelihood numerically. So instead of using the likelihood, which is the probability of the model producing zero error for a combination of A and τ, we used the probability of the model producing an error below a small non-zero threshold for a combination of A and τ; we denote this threshold by ξ.

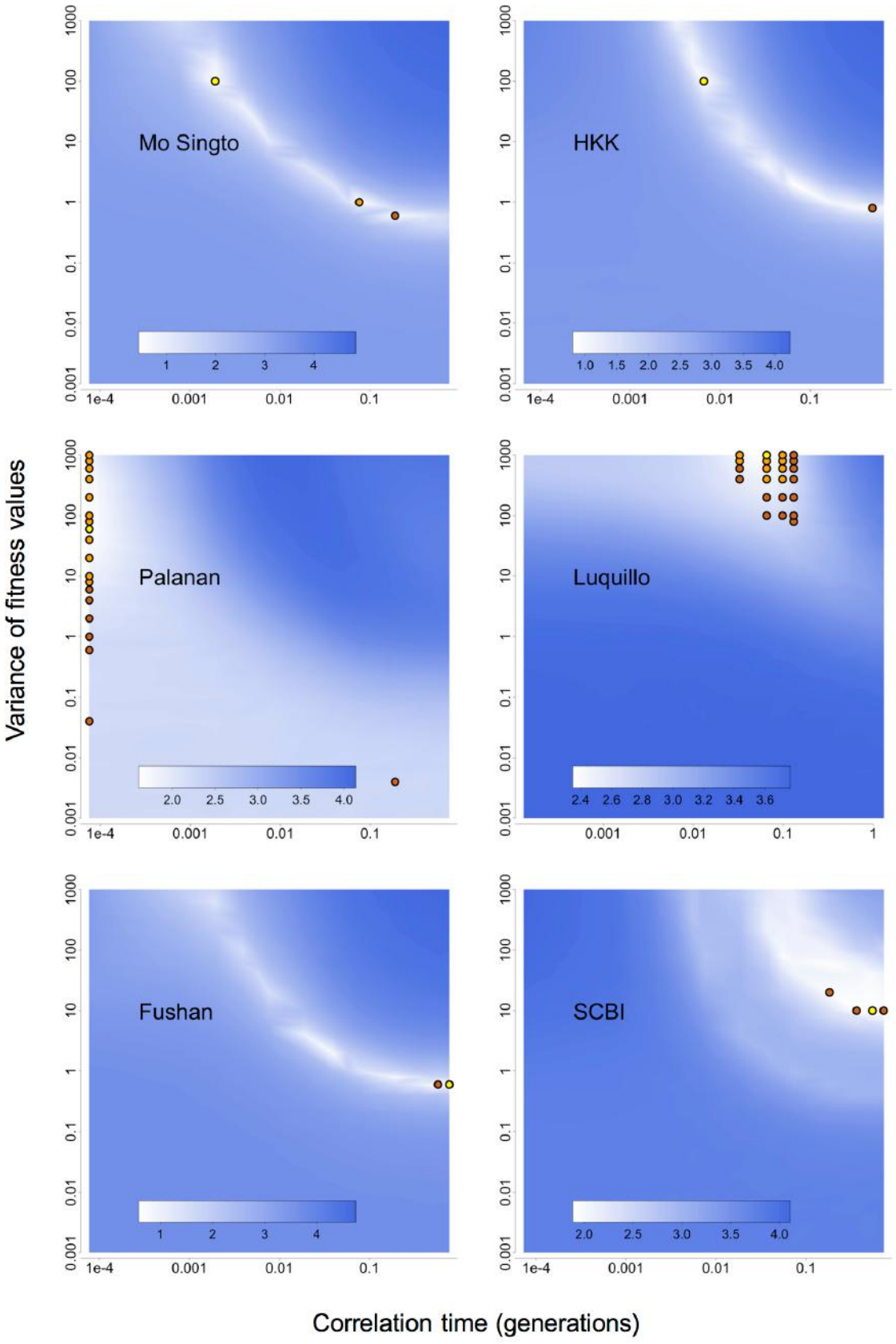
Specifically, 10 of the 11 plots for which the model gave good fits had an area greater than 16 ha, and for each of these 10 plots, we first extracted the 20 combinations of A and τ with the smallest errors, as measured by $\varepsilon_{1,A,\tau}$. For each of these 20 combinations, we ran a further 99 sets of 1,000 simulations, resulting in 100 sets of 1,000 simulations overall and a distribution of 100 realized errors (realized values of the random variable $\varepsilon_{A,\tau}$). Out of the 20 combinations, we determined the combination with the smallest median error, $\xi = \min_{A,\tau}\{\varepsilon_{j_{50},A,\tau}\}$. By definition, this combination produced errors that were below ξ 50% of the time (i.e., the probability of producing errors below ξ is 0.5). Then following the method described in the previous paragraph, any of the other 19 combinations that produced errors below ξ between 12.5% and 50% of the time (the lower limit of 12.5% produced a lower limit for the probability ratio of 1/4) was considered to produce errors that were "similar" to the combination with the smallest median error. We found that for each of the 10 plots, (1) at least one of the 20 combinations produced errors below ξ less than 12.5% of the time, that is, produced errors that were not similar to the combination with the smallest median error; and (2) there was small variation in the errors for each of the 20 parameter combinations – for the 10 plots, the average standard deviation of the 100 realized errors across the 20 parameter combinations was only 0.078–1.70%. Together, (1) and (2) indicated that the remaining 428 combinations of A and τ (which have larger $\varepsilon_{1,A,\tau}$) also produces errors that are not similar to the combination with the smallest median error. Thus, we regarded these 428 combinations as producing errors that are not similar to the combination with the smallest median error. The remaining plot for which the model gave good fits had an area of 16 ha. For this remaining plot, we considered all combinations of A and τ and determined the combination with the smallest median error, ξ. Afterwards, we again considered the combinations of A and τ that produced errors below ξ 12.5–50% of the time as giving errors similar to the combination

with the smallest median error. For the two plots (Lenda and Luquillo) with marginally good model fits, we repeated the analyses described for the 11 plots with good model fits.

For the 11 plots with good model fits, Fig. 5 in the main text and Fig. S17 in this Appendix show the mean number of extinctions predicted by the model, for all combinations of A and τ. For the 10 plots with area greater than 16 ha, the mean number of extinctions for each parameter combination was calculated by performing 100 simulations (each lasting a time period corresponding to the full census period plus an additional 2×10^5 time-steps) using 100 rarefied datasets, and then taking the mean number of extinctions over the 100 simulations. For the single plot with an area of 16 ha, there was only one rarefied dataset, so the mean number of extinctions for each parameter combination was calculated by performing 100 simulations with the same rarefied dataset, and then taking the mean number of extinctions over the 100 simulations. In Fig. S17, we also show the mean number of extinctions for the two plots with marginally good model fits. On Figs. 5 and S17, we highlight the combinations of A and τ that produced errors below ξ (smallest median error) 12.5–50% of the time. We considered all these "best-fit combinations" when determining which parameter regime a plot occupies, in the sense of how the mean number of extinctions changes with the two parameters when starting from a parameter combination. More specifically, the parameter regime a plot occupies was determined by considering how the mean number of extinctions generally changes with the two parameters when starting from the best-fit combinations, with each combination weighted by its probability of producing an error below ξ. So, for example, we determined the Palanan plot as occupying the parameter regime in the top-left quadrant of the parameter space explored, whereby the mean number of extinctions decreases with A but increases with τ. We note that the best-fit combinations of A and τ often lie near or at the boundary of the 2-D parameter space explored (Figs. 5 and S17). This suggests that if we extrapolate the parameter region of low model error (brightest regions in Fig. S15) to outside the parameter space explored, then we would find new parameter combinations that give similar or even lower errors than the combinations in the parameter space explored. However, such extrapolation would likely not change the parameter regime occupied by the best-fit combinations. For example, for the BCI plot, there was a line of low model error that extends to the top-left corner of the parameter space explored (Fig. 15), such that following the line outside of this parameter space would likely still give best-fit combinations occupying the parameter regime where increasing A would increase the mean number of extinctions and increasing τ would decrease the mean number of extinctions (Fig. S17). Therefore, we expect that our interpretation of which parameter regime a plot occupies is robust to exploration of a larger parameter space. We had already explored a large parameter space whereby A (the variance of the lognormal distribution of fitnesses) varies over three orders of magnitude and τ (the correlation time of environmental conditions) varies over four orders of magnitude. Therefore, parameter combinations outside of this space could be biologically infeasible. Future work is required to assess the biological feasibility of the parameter space for our model, which would help to constrain the best-fit combinations and further improve interpretation of model results.

Fig. S18 is the same as Figs. 5 and S17, except for the mean proportion of extinctions (proportion of initial number of species that go extinct) predicted by the model instead of the mean number of extinctions. Trends in the mean proportion of extinctions are the same as those for the mean number of extinctions, but on a standardized scale of 0–1.


La Planada
Lenda
Edoro
Korup
Khao Chong
BCI
ln(Per. Error)
Variance of fitness values
Correlation time (generations)


Mo Singto
HKK
Palanan
Luquillo
Fushan
SCBI
Variance of fitness values
Correlation time (generations)

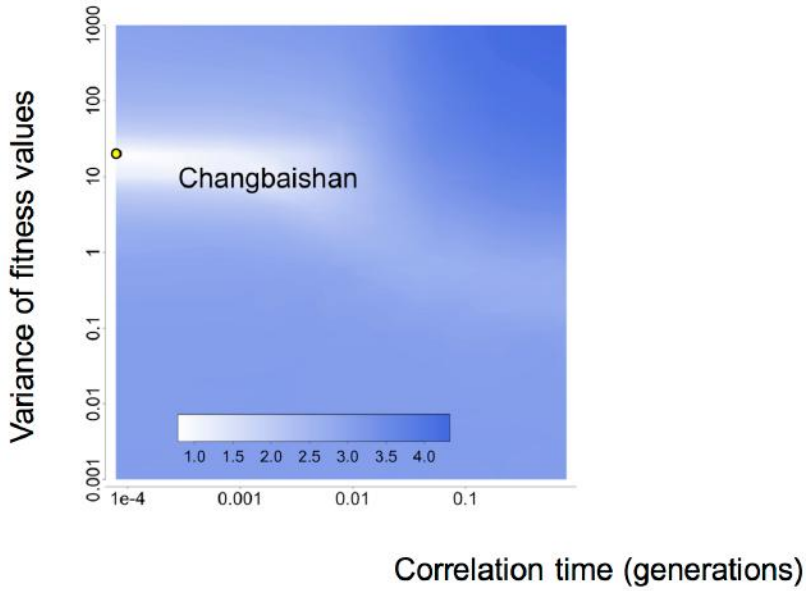
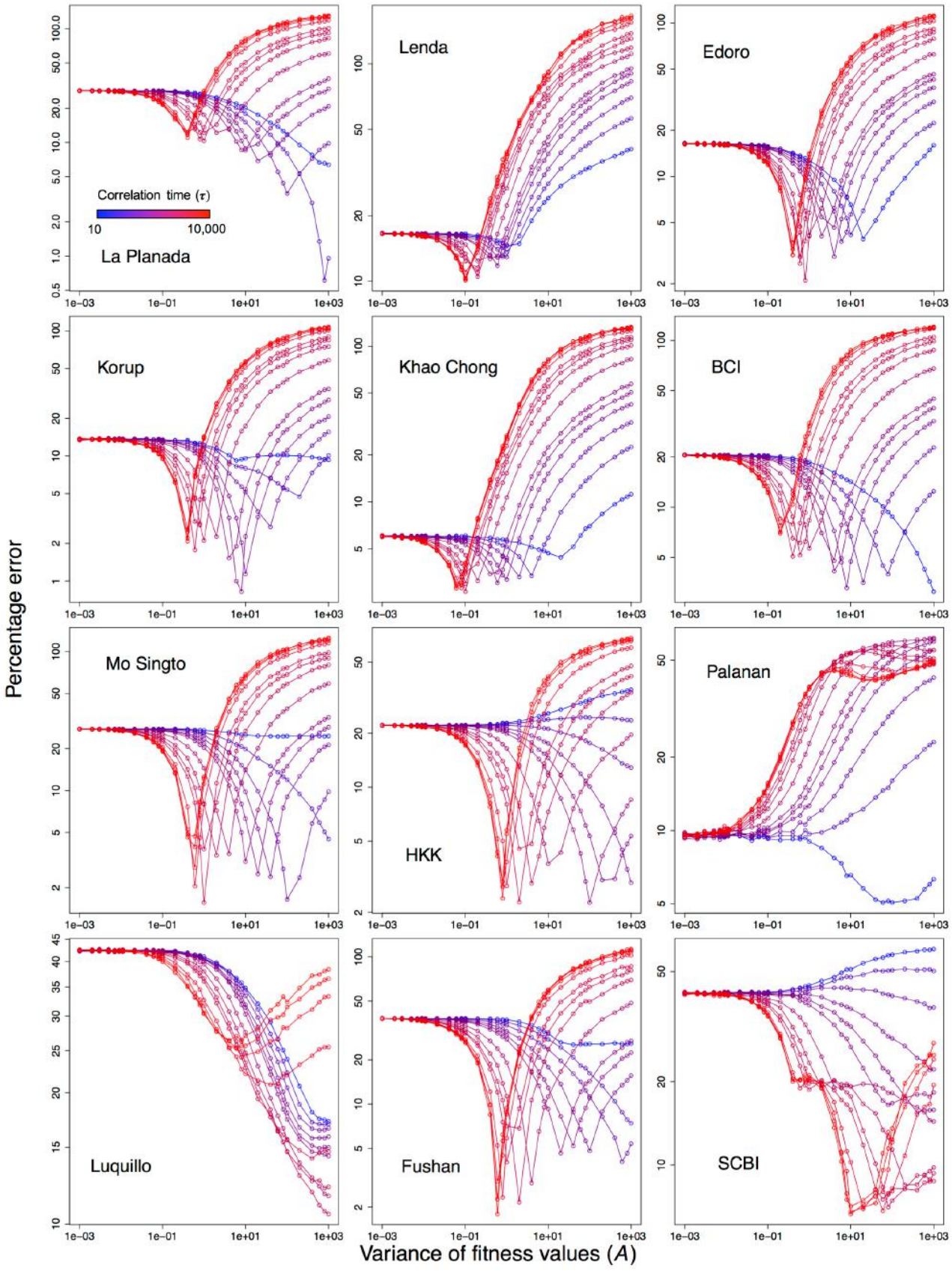


Figure S15. Error (with respect to temporal population variability and temporal correlations of species abundances) from simulations of a dynamic, mechanistic model for 13 of the CTFS–ForestGEO forest plots considered in this study. Each panel shows the logarithm of the error ($\varepsilon_{1,A,\tau}$ for the 11 plots with area greater than 16 ha and $\varepsilon_{j_{50},A,\tau}$ for the two plots with area 16 ha; see text in Appendix S5 for details) for different combinations of values of two key model parameters: the variance of the lognormal distribution of possible fitness values for each model species (A) and the correlation time determining how frequently the fitness values were redrawn due to changes in environmental conditions (τ) (see Fig. 3). 11 of the 13 plots have at least one parameter combination giving an error < 10% (plots with "good model fits"), whereas the remaining two plots (Lenda and Luquillo) have one parameter combination giving an error within 0.5% of 10% (plots with "marginally good model fits"). For each plot, the combination of parameter values giving the smallest median (typical) error, ξ, is marked with a yellow dot. Combinations of parameter values producing errors below ξ 25%–50% of the time are marked with orange dots, whereas combinations of parameter values producing errors below ξ 12.5%–25% of the time are marked with brown dots. Together, these are the "best-fit combinations" (see text in Appendix S5 for details). The plots are ordered (from left to right) in order of ascending absolute latitude: La Planada (1.16°N), Lenda (1.31°N), Edoro (1.56°N), Korup (5.07°N), Khao Chong (7.54°N), BCI (9.15°N), Mo Singto (14.4°N), HKK (Huai Kha Khaeng; 15.6°N), Palanan (17.0°N), Luquillo (18.3°N), Fushan (24.8°N), SCBI (38.9°N) and Changbaishan (42.5°N).


Percentage error
Variance of fitness values (A)
Correlation time (τ)
10
10,000
La Planada
Lenda
Edoro
Korup
Khao Chong
BCI
Mo Singto
HKK
Palanan
Luquillo
Fushan
SCBI

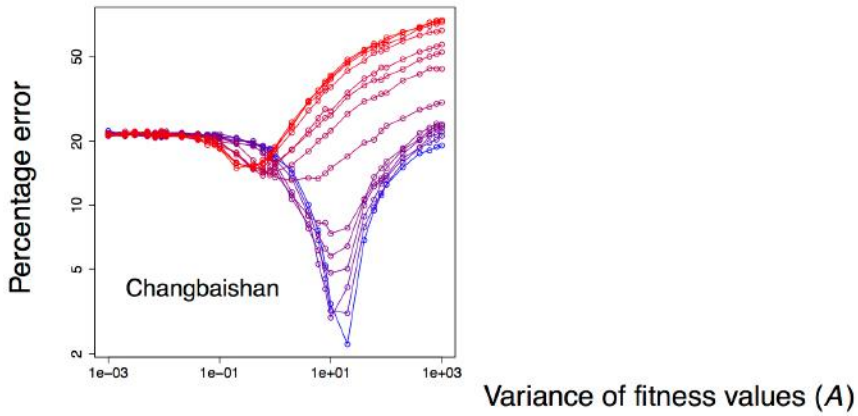
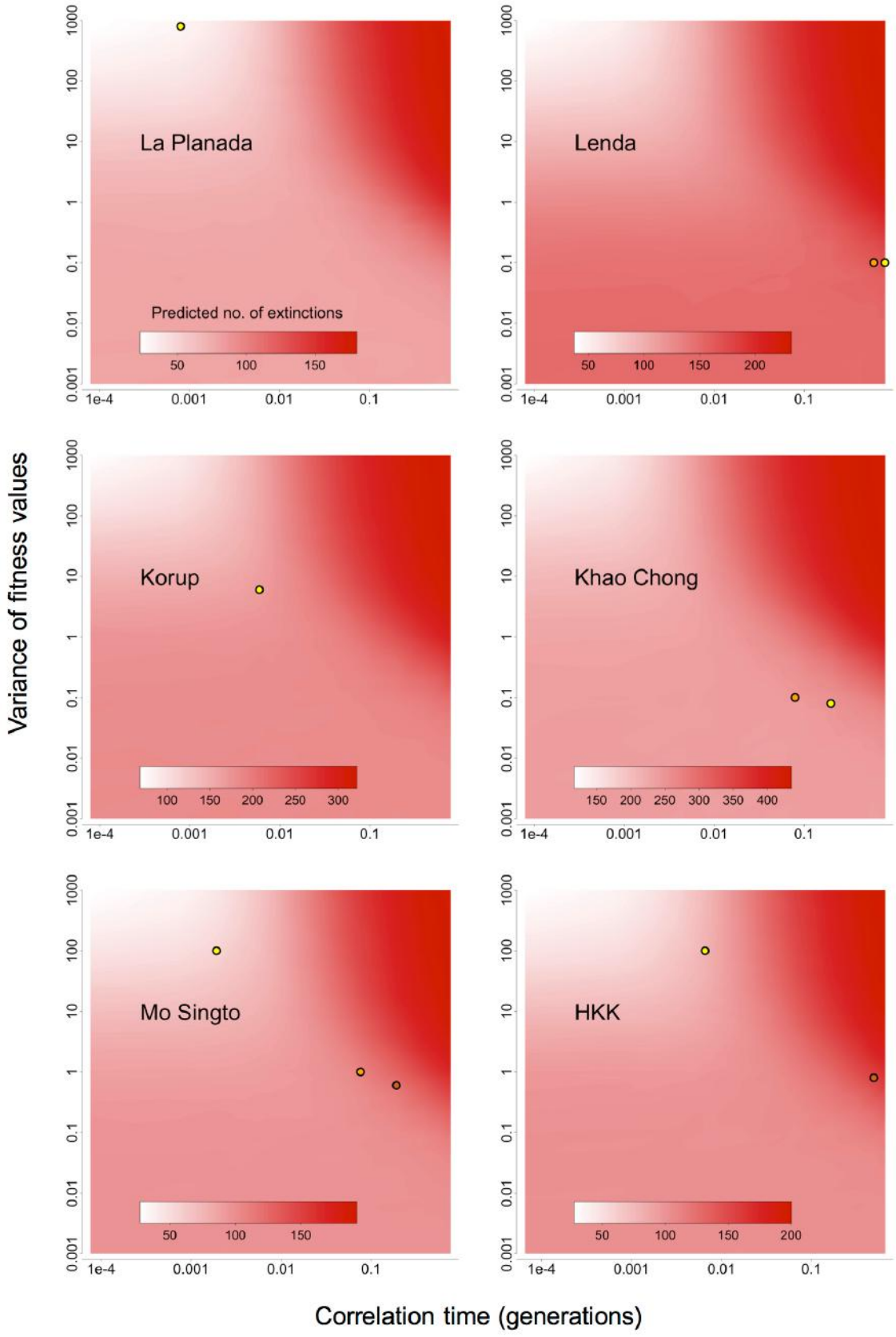


Figure S16. Same as Fig. S15, except using 2-D graphs instead of a heat map. The 2-D graphs make it easier to see that the error for each plot reaches a minimum value across the parameter combinations examined.


La Planada
Predicted no. of extinctions
50 100 150
Lenda
50 100 150 200
Korup
100 150 200 250 300
Khao Chong
150 200 250 300 350 400
Mo Singto
50 100 150
HKK
50 100 150 200
Variance of fitness values
Correlation time (generations)
1000
100
10
1
0.1
0.01
0.001
1e-4
0.001
0.01
0.1

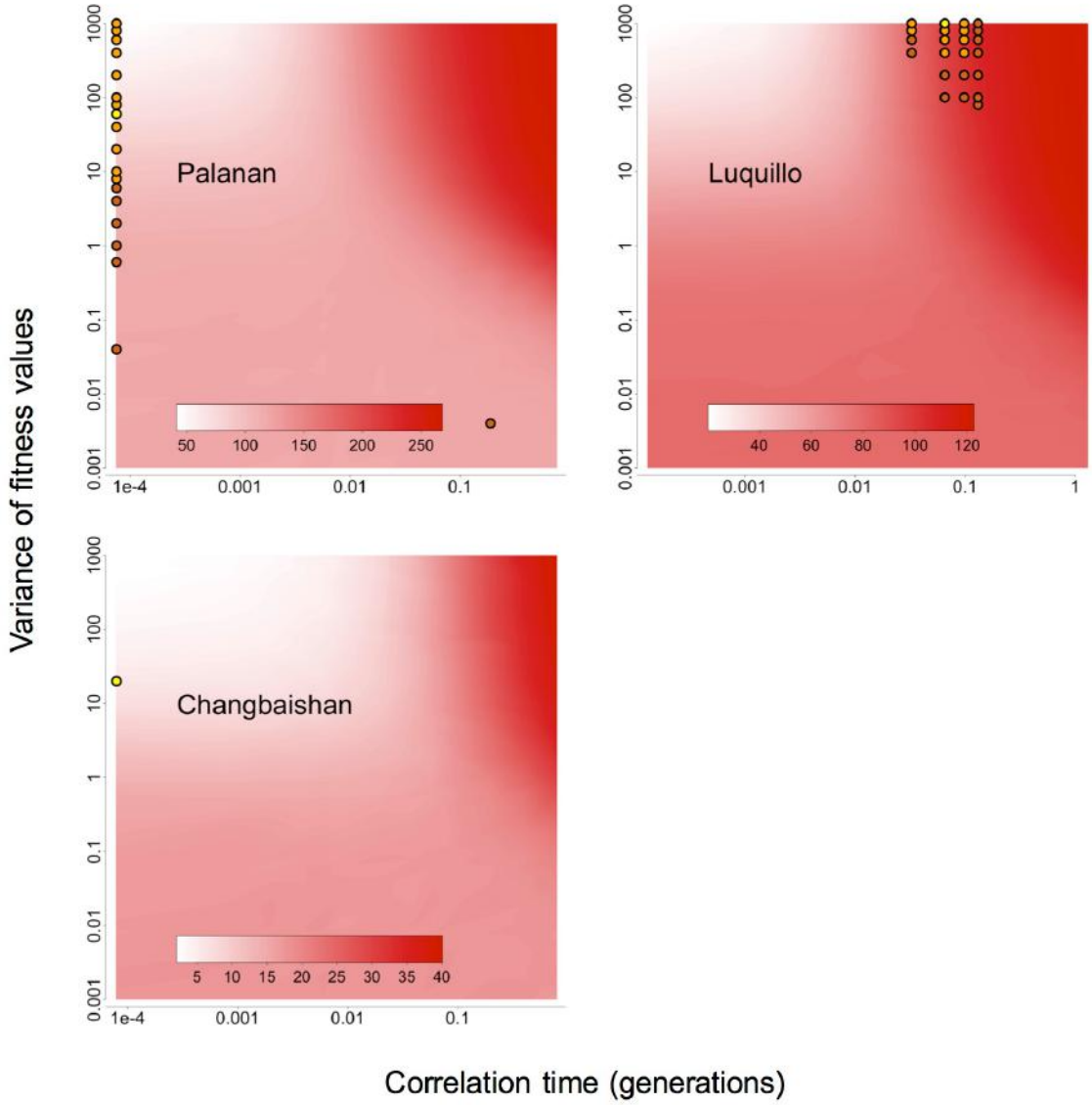
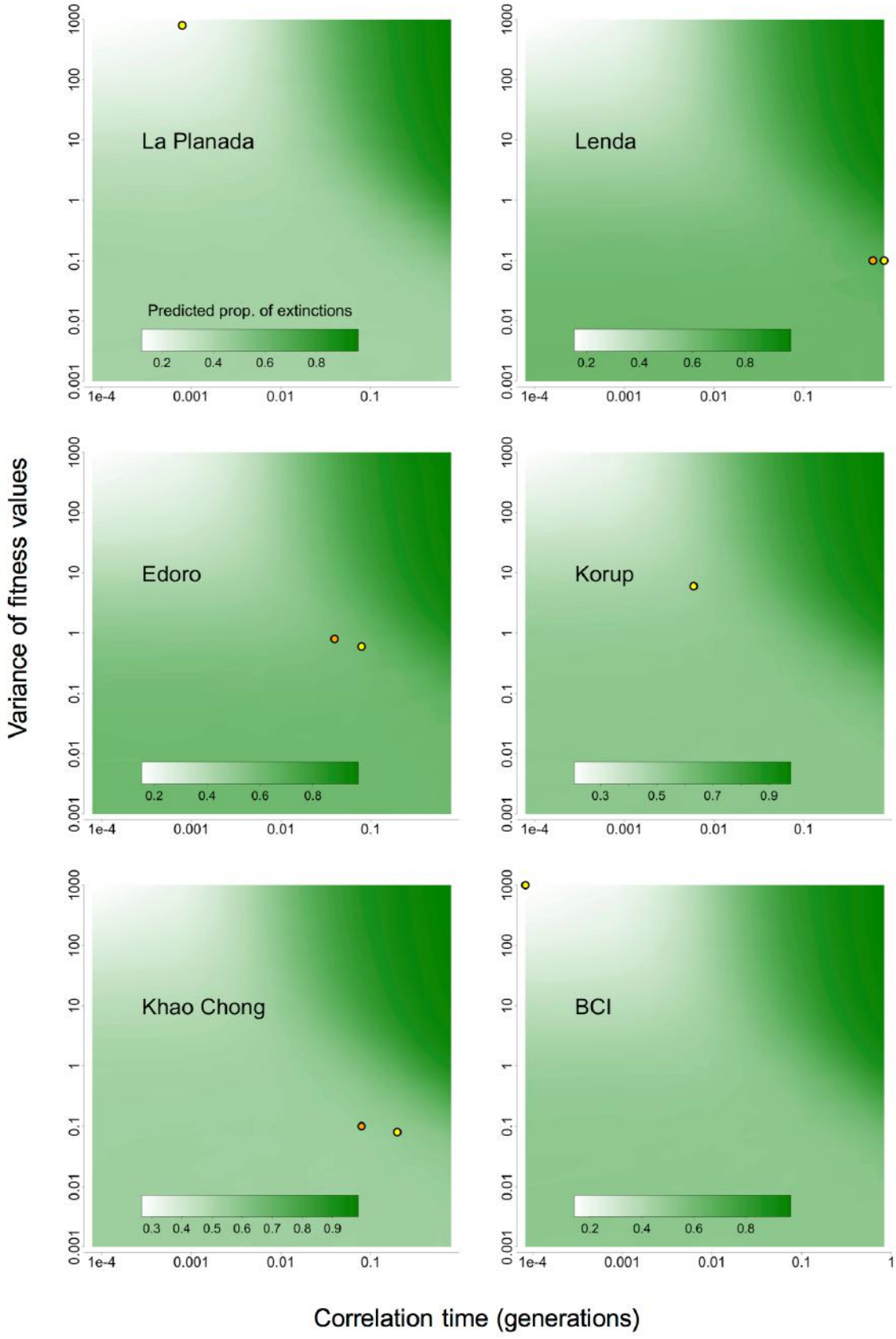
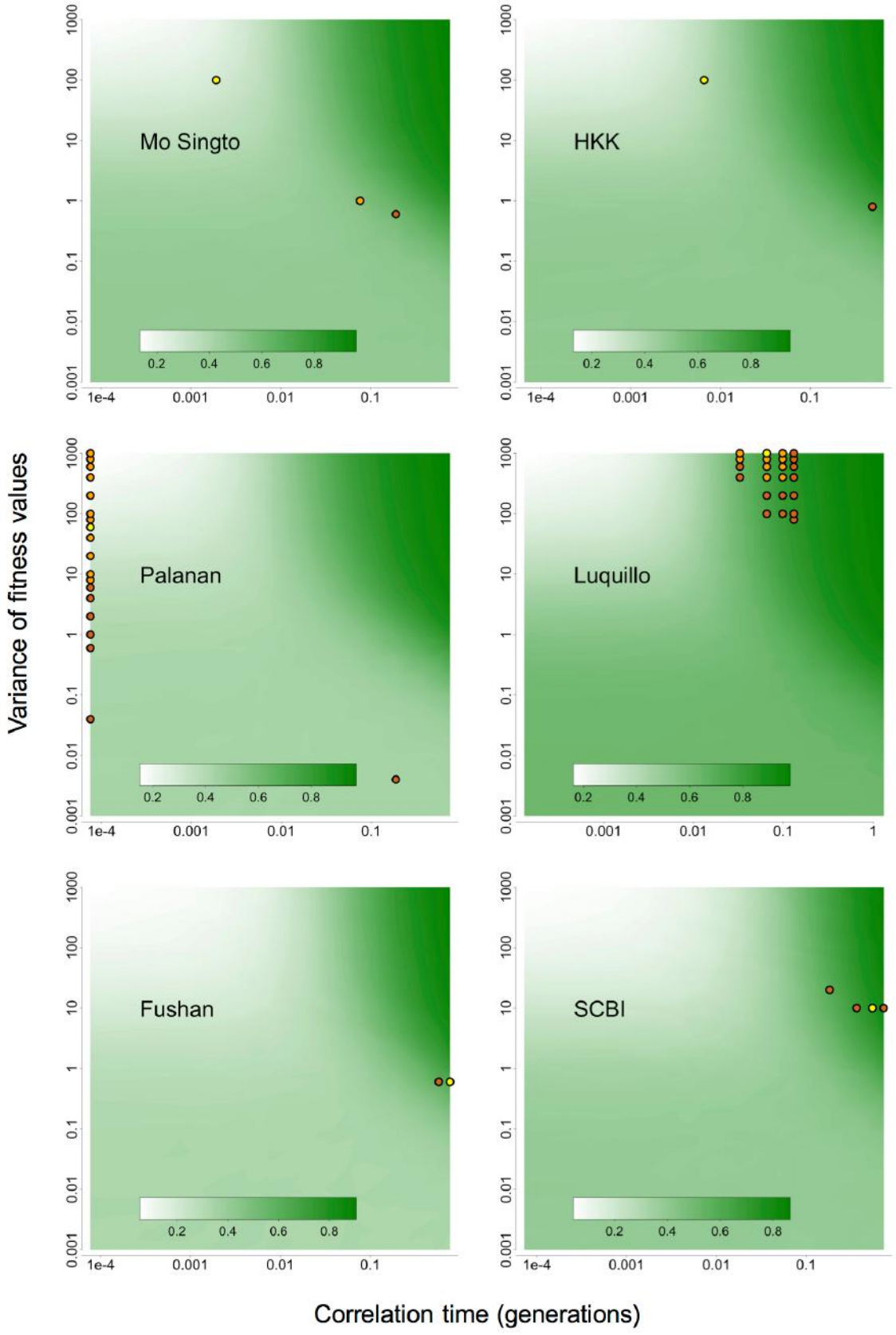


Figure S17. Predicted mean number of extinctions (different colors) from simulations of a dynamic, mechanistic model for nine of the CTFS–ForestGEO forest plots considered in this study. Each panel shows the predicted mean number of extinctions for different combinations of values of two key model parameters: the variance of the lognormal distribution of possible fitness values for each model species (A) and the correlation time determining how frequently the fitness values were redrawn due to changes in environmental conditions (τ) (see Fig. 3). For each plot, the combination of parameter values giving the smallest median (typical) error (with respect to the temporal population variability and temporal correlations of species abundances), ξ, is marked with a yellow dot. Combinations of parameter values producing errors below ξ 25%–50% of the time are marked with orange dots, whereas combinations of parameter values producing errors below ξ 12.5%–25% of the time are marked with brown dots. Together, these are the "best-fit combinations" (see text in Appendix S5 for details). The plots are ordered (from left to right) in order of ascending absolute latitude: La Planada (1.16°N), Lenda (1.31°N), Korup (5.07°N), Khao Chong (7.54°N), Mo Singto (14.4°N), HKK (Huai Kha Khaeng; 15.6°N), Palanan (17.0°N), Luquillo (18.3°N) and Changbaishan (42.5°N).


La Planada
Predicted prop. of extinctions
0.2 0.4 0.6 0.8
Lenda
0.2 0.4 0.6 0.8
Edoro
0.2 0.4 0.6 0.8
Korup
0.3 0.5 0.7 0.9
Khao Chong
0.3 0.4 0.5 0.6 0.7 0.8 0.9
BCI
0.2 0.4 0.6 0.8
Variance of fitness values
0.001 0.01 0.1 1 10 100 1000
1e-4 0.001 0.01 0.1 1
Correlation time (generations)


Mo Singto
HKK
Palanan
Luquillo
Fushan
SCBI
Variance of fitness values
Correlation time (generations)
0.001
0.01
0.1
1
10
100
1000
1e-4
0.2
0.4
0.6
0.8

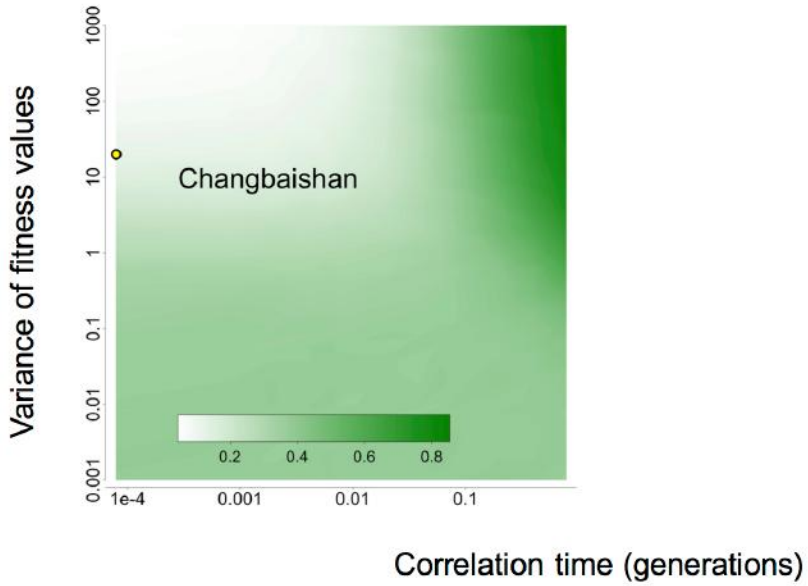


Figure S18. Same as Figs. 5 and S17, except for the predicted mean proportion of extinctions instead of the predicted mean number of extinctions.

References

Anderson-Teixeira, K.J., Davies, S.J., Bennett, A.C., Gonzalez-Akre, E.B., Muller-Landau, H.C., Wright, S.J. *et al*. (2015). CTFS–ForestGEO: a worldwide network monitoring forests in an era of global change. *Glob. Change Biol.*, 21, 528–549.

Burnham, K.P. & Anderson, D.R. (2002). *Model Selection and Multimodel Inference*. Springer-Verlag, New York, NY, USA.

Cayuela, L., Granzow-de la Cerda, Í., Albuquerque, F.S. & Golicher, D.J. (2012). Taxonstand: An R package for species names standardisation in vegetation databases. *Methods in Ecology and Evolution*, 3, 1078–1083.

Chao, A. & Jost, L. (2012). Coverage-based rarefaction and extrapolation: Standardizing samples by completeness rather than size. *Ecology*, 93, 2533–2547.

Chesson, P.L. & Warner, R.R. (1981). Environmental variability promotes coexistence in lottery competitive systems. *The American Naturalist*, 117, 923–943.

Chisholm, R.A., Condit, R., Abd. Rahman, K., Baker, P.J., Bunyavejchewin, S., Chen, Y.-Y. *et al*. (2014). Temporal variability of forest communities: empirical estimates of population change in 4000 tree species. *Ecol. Lett.* 17, 855–865.

Condit, R.C. (1998). *Tropical Forest Census Plots*. Springer-Verlag, Berlin, Germany.

Danino, M., Shnerb, N.M., Azaele, S., Kunin, W.E. & Kessler, D.A. (2016). The effect of environmental stochasticity on species richness in neutral communities. *J. Theor. Biol.*, 409, 155–164.

Danino, M. & Shnerb, N.M. (2018). Theory of time-averaged neutral dynamics with environmental stochasticity. *Phys. Rev. E*, 97, 042406.

Fung, T., O'Dwyer, J.P., Rahman, K.A., Fletcher, C.D. & Chisholm, R.A. (2016). Reproducing static and dynamic biodiversity patterns in tropical forests: The critical role of environmental variance. *Ecology*, 97, 1207–1217.

Harcombe, P.A. (1987). Tree life tables. *BioScience*, 37, 557–568.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., & Jarvis, A. (2005), Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.*, 25, 1965–1978.

Hubbell S.P. (1997). A unified theory of biogeography and relative species abundance and its application to tropical rainforests and coral reefs. *Coral Reefs*, 16(Suppl.), S9–S21.

Hubbell, S.P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton Univ. Press, Princeton, NJ, USA.

Hurlbert, S.H. (1971). The nonconcept of species diversity: A critique and alternative parameters. *Ecology*, 52, 577–586.

Kalyuzhny, M., Kadmon, R. & Shnerb, N.M. (2015). A neutral theory with environmental stochasticity explains static and dynamic properties of ecological communities. *Ecol. Lett.*, 18, 572–580.

Kitajima, K. & Augspurger, C.K. (1989). Seed and seedling ecology of a monocarpic tropical tree, *Tachigalia versicolor*. *Ecology*, 70, 1102–1114.

Ricklefs, R.E. & He, F. (2016). Region effects influence local tree species diversity. *P. Natl. Acad. Sci. USA*, 113, 674–679.

Smith, A.R., Pryer, K.M., Schuettpelz, E., Korall, P., Schneider, H. & Wolf, P.G. (2006). A classification for extant ferns. *Taxon*, 55, 705–731.

Van Valen, L. (1975). Life, death, and energy of a tree. *Biotropica*, 7, 259–269.

Volkov, I., Banavar, J.R., Hubbell, S.P. & Maritan, A. (2003). Neutral theory and relative species abundance in ecology. *Nature*, 424, 1035–1037.

Volkov, I., Banavar, J.R., Hubbell, S.P. & Maritan, A. (2007). Patterns of relative species abundance in rainforests and coral reefs. *Nature*, 450, 45–49.