

New Phytologist Supporting Information

Article title: Africa as an evolutionary arena for large fruits

Authors: Friederike J. R. Wölke, Andressa Cabral, Jun Ying Lim, W. Daniel Kissling, Renske E. Onstein

Article acceptance date: 14 May 2023

Methods S1: Supplementary Methods

Palm species that occur in Africa, Americas, Asia and outside

Twenty-two palm species were found in more than one regime (i.e., mainland Africa/Americas/Asia as well as outside). For those species, we manually checked the International Union for Conservation of Nature and Natural Resources (IUCN) range maps and Global Biodiversity Information Facility (GBIF) occurrence records for the species' predominant distribution and assigned species to 'Africa' or 'Elsewhere' (Table S1). Regimes were assigned following botanical continents as defined by the World Geographical Scheme for Recording Plant Distributions.

Evolutionary trait models

The method requires information on the ancestral states of the selective regimes (i.e., Africa vs. elsewhere; Americas vs. elsewhere; Asia vs. elsewhere), which was computed for both regimes using stochastic character mapping ('simmap') in the 'phytools' R package (Revell, 2012). This probabilistic approach has the advantage of not relying on parsimony to infer character changes, thus allowing for the modelling of reversions or multiple transitions in selective regime evolution, and allowing for high variance in trait evolution, which we consider important for adaptive processes. Furthermore, 'simmap' enables to integrate uncertainty about the reconstruction by repeating the reconstruction and evaluating emerging patterns in the posterior distribution of the result (Bollback, 2006; Revell, 2012).

Prior to the reconstruction of ancestral states, we fitted and compared two different Markov models for discrete character evolution to model transition rates between distribution in Africa/Americas/Asia and elsewhere through time (Harmon, 2019). We distinguished between an equal-rates model that assumes identical transition rates from Africa/Americas/Asia to elsewhere and from elsewhere to Africa/Americas/Asia and an all-rates-different model, that allows transition rates between both regimes to vary. The model assuming different rates had higher support in the analyses for Americas and Asia, while for Africa, the model with equal rates revealed a higher support, however, ΔAIC_c between both was < 2 . Thus, we did not reject the ARD model (Burnham & Anderson, 2004) and argued that it is more likely for transitioning from Africa to anywhere else to follow a different rate than transitioning from anywhere into Africa.

For the most complex OUwie models, which allow α and σ^2 , or both, to vary (OUMV, OUMA, OUMVA, see Table 1), the complexity of the model can often be too large for the information that is contained within the data, thus, parameters are poorly estimated and loglikelihood functions return suboptimal. This offers potential for poor model choice and wrong interpretations. We thus excluded all models which returned negative eigenvalues in the Hessian matrix, suboptimal likelihoods and AIC_c , and/or odd parameter estimates (Beaulieu *et al.*, 2012; Beaulieu *et al.*, 2016). Figure S1 in the Supporting Information provides an overview of all models that failed to converge.

Spatial autocorrelation

Spatial autocorrelation in the model residuals of the ordinary least square (linear) models and the SAR models (Table S7) were assessed using Moran's I (Figure S2). The spatial weight matrix reflected the minimal distance to connect each botanical country to at least one other country (i.e., 1423 km). As both linear and SAR models indicated similar effects of predictor variables on response, we focused on the standardized effects from the linear models in the main text, primarily because spatial autocorrelation in data can also be the result of biological processes, and the interpretation of spatial effects on the response is therefore complex.

Null expectation for palm fruit sizes under the absence of selection for trait evolution: Brownian motion trait simulations

We used simulations of trait states in an evolutionary-neutral scenario, where selection on the trait is absent, following a Brownian motion (BM) process. For this, a set of phylogenetic trees and empirical trait data for the species at the tips of the tree was needed to estimate the required parameters for the simulation based on the observed data.

First, we estimated the required starting parameters from the empirical data under a flat adaptive landscape using a likelihood approach with the ‘BBMV’ package (Boucher, 2018) in R Statistical Software (v4.1.3 and v4.0.0-2, R Core Team, 2022). These parameters include the trait value at the root of palm phylogeny (z_0), the stochastic rate at which the trait evolves (σ^2), and the bounds which sets limits to the trait. The method resembles commonly used models for evolution of discrete traits by requiring discretization of continuous values at the nodes between the two reflective bounds at a number of equally spaced points. To enhance the resolution of the discretization, we increased the number of points from the recommended default (N points = 50) to 100, with the expense of higher computational efforts and longer calculation times. It has been suggested that the minimum and maximum trait values at the tips are the maximum likelihood estimators of both bounds in a uniform distribution (Boucher & Démery, 2016). Thus, we included the log-transformed values for 0.3 cm and 45 cm, which represent the minimum and maximum observed palm fruit length values, as the bounds of this simulation. From this inference, we extracted the discretized trait value with the highest probability density at the root and the maximum likelihood estimate of the stochastic rate. These values varied between phylogenies but were $\log(z_0) = 0.982$ cm (i.e., 2.78 cm) and $\log(\sigma^2) = 0.049$ (i.e., 1.05 cm/million years) for the MCC tree. Note that the estimated z_0 value for fruit size at the root of palms was estimated to be larger than today’s average (i.e., $z_0 = 2.65$ cm, today’s average = 2.5 cm).

Next, these parameter values were integrated to simulate fruit length values following neutral/stochastic evolution with bounds. The expected distribution of trait values that evolves following a BBM process resembles a flat adaptive landscape with the explicit absence of a selective force. The resulting ‘neutrally’ simulated palm fruit lengths were used in subsequent analyses to assess differences in evolution and distribution with the empirical fruit length data. We calculated species averages for simulated traits across simulations from 100 phylogenetic trees to incorporate uncertainty about species placement in the phylogeny into the null expectation. Furthermore, we simulated fruit sizes 100 times across the same tree (the MCC tree), because tips may tend to show the root state after sufficient simulations across the same tree.

For the linear models, we calculated the 95th percentile fruit lengths in each assemblage (botanical country) by merging the simulated fruit length data for each species (averaged across the 100 simulations on the MCC tree) with the empirical distribution data. We expected that interactions between Africa and the environmental variables (canopy height, mammalian frugivore body mass change and the 95th percentile extant mammalian frugivore body mass) would not be supported under the null model.

Notes S1: Supplementary Results

Spatial autocorrelation

Accounting for autocorrelation in the spatial structure of the data increased the model fit (ΔAIC : 13.04) and reduced spatial non-independence of the data ($\Delta\text{Moran's I}$: 0.12; Figure S2). The SAR recovered the same patterns as the LM, yet with an even stronger influence from occurrence in Africa compared to elsewhere ($0.91, z = 4.53, P < 0.001$).

Null expectation for palm fruit sizes under the absence of selection for trait evolution: Brownian motion trait simulations

Results from linear models on simulated palm fruit lengths (Table S5, Figure 3c, Figure S7) suggested that the effects detected in the empirical palm data were different from those predicted under non-directional (neutral) evolution. Specifically, we detected a negative effect of annual precipitation on maximum fruit size. Furthermore, we detected a negative effect of maximum body mass of extant frugivores on fruit size, contrasting the positive effect detected in the empirical data. Similarly, the effect of Africa on maximum fruit size was negative, whereas a positive effect was detected in the empirical data. These results illustrate that under stochastic trait evolution, the largest fruits would have been expected outside Africa, and in places with relatively small-bodied mammalian frugivores.

Supplementary Tables

Table S1: Assignment of widespread palm species to realms. List of palm species that are distributed in Africa, Americas, Asia and outside. In order to assess the distribution in these three continents as a unique selective regime for palm fruit length, palm species had to be assigned either to ‘distribution in Africa, Americas or Asia’ or ‘distribution elsewhere’. Distributions were manually checked using range maps from the International Union for Conservation of Nature and Natural Resources (IUCN). Since *A. microcarpa*, *C. aruensis*, *C. vitiensis*, *C. warburgii*, *C. utan*, *H. costata*, *H. reptans*, *L. muelleri*, *M. salomonense*, *P. insignis*, *P. macarthurii*, and *R. taedigera* did not have any range map on IUCN, we checked the distribution of occurrence records on the Global Biodiversity Information Facility (GBIF). Species distributions for the Africa dataset were checked on the 24.06.2022, and for the Americas and Asia datasets on the 21.09.2022.

Species	Current distribution	Assigned selective regime
<i>Arenga microcarpa</i>	Asia, Australasia	Asia
<i>Borassus aethiopum</i>	Africa, Madagascar	Africa
<i>Calamus aruensis</i>	Asia, Australasia	Asia
<i>Calamus vitiensis</i>	Asia, Australasia; Pacific	Asia
<i>Calamus warburgii</i>	Asia, Australasia	Asia
<i>Chamaerops humilis</i>	Africa, Europe	Elsewhere
<i>Corypha utan</i>	Asia, Australasia	Asia
<i>Heterospathe elata</i>	Asia, Pacific	Asia
<i>Hydriastele costata</i>	Asia, Australasia	Asia
<i>Hyphaene coriacea</i>	Africa, Madagascar	Africa
<i>Hyphaene reptans</i>	Africa, Asia	Asia
<i>Hyphaene thebaica</i>	Africa, Asia	Africa
<i>Livistona benthamii</i>	Asia, Australasia	Elsewhere
<i>Livistona carinensis</i>	Africa, Asia	Asia
<i>Livistona muelleri</i>	Asia, Australasia	Elsewhere
<i>Metroxylon salomonense</i>	Asia, Pacific	Asia
<i>Phoenix caespitosa</i>	Africa, Arabian Peninsula	Africa
<i>Phoenix reclinata</i>	Africa Arabian Peninsula	Africa
<i>Pinanga insignis</i>	Asia, Pacific	Asia
<i>Ptychosperma macarthurii</i>	Asia, Australasia	Asia
<i>Raphia farinifera</i>	Africa, Madagascar	Africa
<i>Raphia taedigera</i>	Africa, Americas	Elsewhere (Africa analysis); Americas (Americas analysis)

Table S2: Linear model selection for empirical palm data. Model selection to determine the variables that explain variation in the maximum 95th percentile palm fruit lengths across global assemblages. Only assemblages that include at least 3 palm species were included. The best (selected) model, is depicted in **bold**. ‘P’ refers to the p-value of the comparison (when < 0.05, the more complex model was preferred). Res = residual, Df = degrees of freedom, F-value = model statistic.

Model	Model description	Comparison to...	Res.Df	RSS	Df	Sum of Sq	F value	P	Conclusion
1	Full model without Africa effect		110	3.27					
2	Only significant effects from model 1	Model 1	113	3.29	-3	-0.24	0.27	0.85	Simpler model, does not perform significantly worse than model 1
3	Model 2, plus biogeography (Africa/elsewhere)	Model 2	112	2.48	1	0.81	36.43	<0.001	Performs significantly better than model 2, biogeography effect supported
4	Model 3, plus all interactions with Africa	Model 3	108	2.32	4	0.16	1.85	0.13	Does not perform significantly better than model 3, interactions not supported
5	Model 3, plus only interaction Africa/mammal body mass	Model 3	111	2.48	1	0.003	0.15	0.698	Does not perform significantly better than model 3, interactions not supported
6	Model 3, plus interaction Africa/canopy height	Model 3	111	2.36	1	0.12	5.72	0.0185	Performs significantly better than model 3, interaction should be included. Highest adj. Rsquare.
7	Model 3, plus interaction Africa/TempSeas	Model 3	111	2.39	1	0.09	4.05	0.047	Performs significantly better than model 3, interaction should be included
8	Model 3, plus interaction Africa/AnnualPrec	Model 3	111	2.44	1	0.05	2.05	0.155	Does not perform significantly better than model 3, interaction not supported
9	Model 3, plus both interaction Africa/canopy height and Africa/TempSeas	Models 6 and 7	110	2.34	1	0.06	2.60	0.11	Does not perform significantly better than models 6 or 7, double interactions not supported

Table S3: Spearman's rank correlation coefficients for independent variables that were included in the linear models. Correlations in bold were statistically significant. The p-values were adjusted following Holm (1979). Correlation coefficients $> |0.7|$ are regarded as too strongly correlated to be implemented in the same linear model. This was not the case for any variables.

	Mean Canopy height	Precipitation seasonality	Annual precipitation	Temperature seasonality	Mean annual temperature	Mean body size decrease
Current maximum body size	-0.20	-0.29	0.56	0.31	-0.02	0.07
Mean body size decrease	0.10	-0.29	0.20	0.41	-0.29	
Mean annual temperature	-0.45	0.10	0.13	-0.33		
Temperature seasonality	0.03	-0.66	0.49			
Annual precipitation	-0.24	-0.52				
Precipitation seasonality	0.42					

Table S4: Parameter estimates from OUwie evolutionary trait models. Parameter estimates for alpha (selective strength towards optimum), sigma (rate of stochastic evolution) and theta (trait evolutionary optimum) from the best fitting model averaged across trees from empirical data. ‘n’ indicates the number of trees that selected each model over the others based on AICc, ‘0’ refers to ‘elsewhere’, whereas ‘1’ refers to the respective region (Africa, Americas, Asia or simulated [and then assessed for African lineages]), ‘aicw’ (between 0 and 1) refers to the AIC weight of the respective model relative to the other six models considered. Models that were selected with the highest proportion across 100 trees are presented in bold.

Regime	Model	N	alpha0	alpha1	sigma0	sigma1	theta0	theta1	aicw
Africa	BM	0	-	-	-	-	-	-	0
Africa	BMS	0	-	-	-	-	-	-	0
Africa	OU1	0	-	-	-	-	-	-	0
Africa	OUM	82	1.289		1.134		0.542	1.191	0.712
Africa	OUMA	0	-	-	-	-	-	-	0
Africa	OUMV	16	0.380		0.330	0.509	0.575	1.195	0.773
Africa	OUMVA	2	0.108	0.102	0.097	0.166	0.620	1.227	0.666
Americas	BM	0	-	-	-	-	-	-	0
Americas	BMS	0	-	-	-	-	-	-	0
Americas	OU1	0	-	-	-	-	-	-	0
Americas	OUM	16	2.167		0.070		0.850	0.986	0.756
Americas	OUMA	1	0.000	0.240		0.006		0.856	0.905
Americas	OUMV	45	0.578		0.020	0.029	0.852	0.988	0.901
Americas	OUMVA	38	0.121	0.118	0.004	0.004	0.858	0.985	0.910
Asia	BM	0	-	-	-	-	-	-	0
Asia	BMS	0	-	-	-	-	-	-	0
Asia	OU1	0	-	-	-	-	-	-	0
Asia	OUM	6	1.752		0.067		0.939	0.849	0.823
Asia	OUMA	5	0.096	0.103		0.003		0.959	0.848
Asia	OUMV	40	0.871		0.043	0.031	0.944	0.850	0.946
Asia	OUMVA	49	0.108	0.111	0.004	0.002	0.952	0.838	0.971

Table S5: Linear model selection for empirical angiosperm (excluding palms) data. Model selection to determine the variables that explain variation in the maximum 95th percentile angiosperm fruit widths across global assemblages (only including those assemblages that also contain at least 3 palm species). The best (selected) model, is depicted in **bold**. ‘P’ refers to the p-value of the comparison (when < 0.05, the more complex model was preferred). Res = residual, Df = degrees of freedom, F-value = model statistic.

Model	Model description	Compared to...	Res.Df	RSS	Df	Sum of Sq	F value	P	Conclusion
1	Full model without Africa effect		110	1.55					
2	Only significant effects from model 1	Model 1	115	2.16	-5	-0.18	2.04	0.0786	Simpler model does not perform significantly worse than model 1
3	Model 2, plus biogeography (Africa/elsewhere)	Model 2	114	1.74	1	0.42	27.81	<0.001	Performs significantly better than model 2, biogeography effect supported
4	Model 3, plus all interactions with Africa	Model 3	112	1.71	2	0.03	0.98	0.3797	Interaction model does not perform significantly better than model 3, interactions not necessarily supported
5	Model 3, plus interaction Africa/canopy height	Model 3	113	1.74	1	0.0001	0.0071	0.933	Does not perform significantly better than model 3, interaction not supported
6	Model 3, plus interaction Africa/TempSeas	Model 3	113	1.72	1	0.02	1.51	0.221	Does not perform significantly better than model 3, interaction not supported

Table S6: Linear model selection for simulated palm fruit length under BBM (bounded Brownian motion). Model selection to determine the variables that explain variation in the maximum 95th percentile simulated palm fruit lengths across global assemblages. Only assemblages with at least three palm species were included. The best (selected) model, is depicted in bold. ‘P’ refers to the p-value of the comparison (when < 0.05, the more complex model was preferred). Res = residual, Df = degrees of freedom, F-value = model statistic.

Model	Model description	Comparison to...	Res.Df	RSS	Df	Sum of Sq	F value	P	Conclusion
1	Full model without Africa effect								
2	Only significant effects from model 1		126	3.06					
3	Model 2, plus biogeography (Africa/elsewhere)	Model 2	125	2.84	1	0.21	9.34	<0.05	Performs significantly better than model 2, biogeography effect supported
4	Model 3, plus all interactions with Africa	Model 3	122	2.79	3	0.04	0.69	0.55	Does not perform significantly better than model 3, interactions not supported
5	Model 3, plus only interaction Africa/mammal body mass	Model 3	124	2.82	1	0.02	0.9	0.34	Does not perform significantly better than model 3, interaction not supported
6	Model 2, plus interaction Africa/MAT	Model 3	124	2.84	2	0.002	0.09	0.77	Does not perform significantly better than model 3, interaction not supported
7	Model 2, plus interaction Africa/AnnualPrec	Model 3	124	2.8	2	0.05	2.01	0.16	Does not perform significantly better than model 3, interaction not supported

Table S7: Model results for spatial autoregressive (SAR) model. This model considers the spatial neighborhood of data points (botanical countries, $n = 118$) to estimate the effect of predictors that were used in the linear models, i.e., to explain the 95th percentile maximum palm fruit length across assemblages. ‘Pr’ refers to the p-value (when < 0.05 , the predictor effect explained significant variation in palm fruit length). z-value = model statistic, Std = standard. Sign. = significant, with *** <0.001 , ** <0.01 , * <0.05 , ns = not significant.

Predictor	Estimate	Std. Error	z value	Pr(> z)	Sign.
Distribution in Africa relative to elsewhere	0.91	0.2	4.53	0.00	***
Mean canopy height * Distribution in Africa	-0.32	0.134	-2.358	0.02	*
Current mammal frugivore body mass	0.29	0.078	3.796	0.00	***
Temperature seasonality	-0.31	0.074	-4.114	0.00	***
Mean canopy height	-0.01	0.084	-0.097	0.92	ns

Supplementary Figures

Figure S1: Overview of convergence issues during evolutionary trait modelling. For the Africa, Americas, and Asia models. Models that failed to converge were not considered in the model selection and parameter inferences.

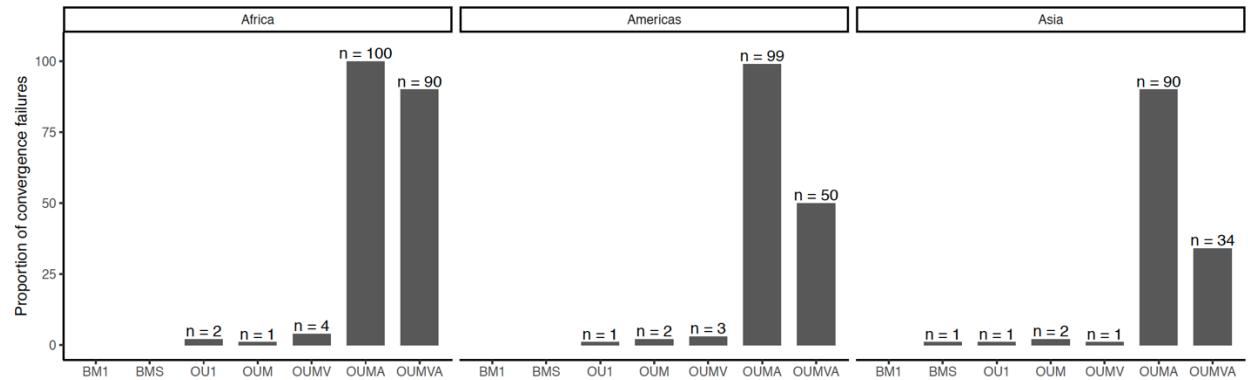


Figure S2: Spatial autocorrelation in the data. This is illustrated for the raw data (black dots), ordinary least square (linear model) residuals (white dots) and spatial error model residuals (black squares). Values close to zero indicate absence of spatial non-independence of the data, negative values indicate dispersion, and positive values indicate clustering.

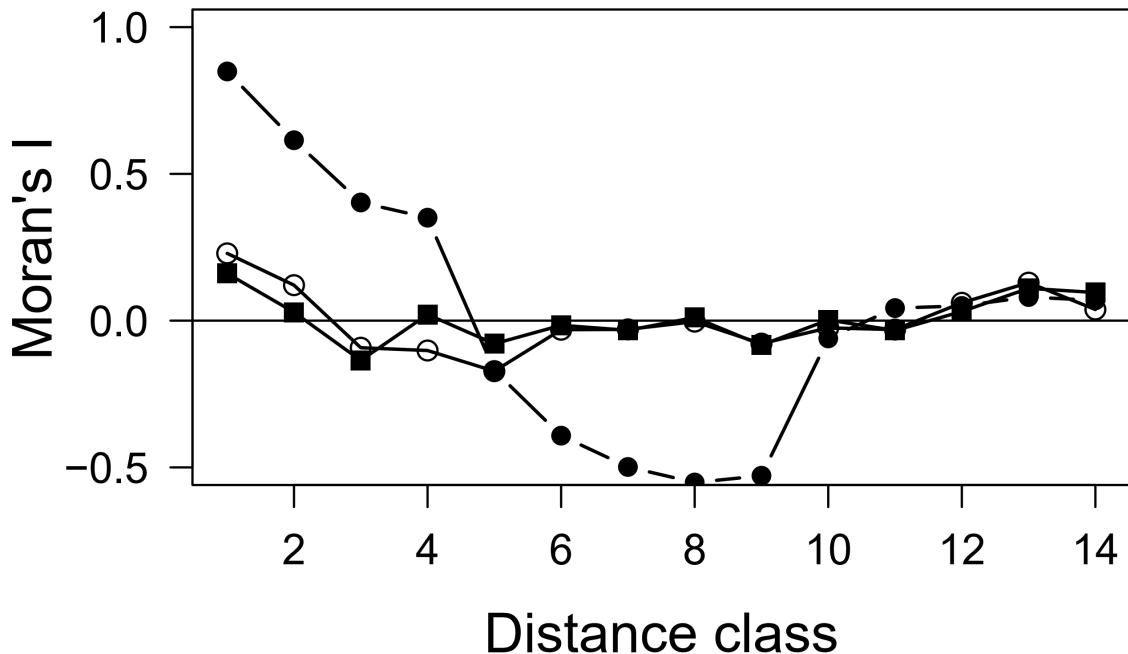
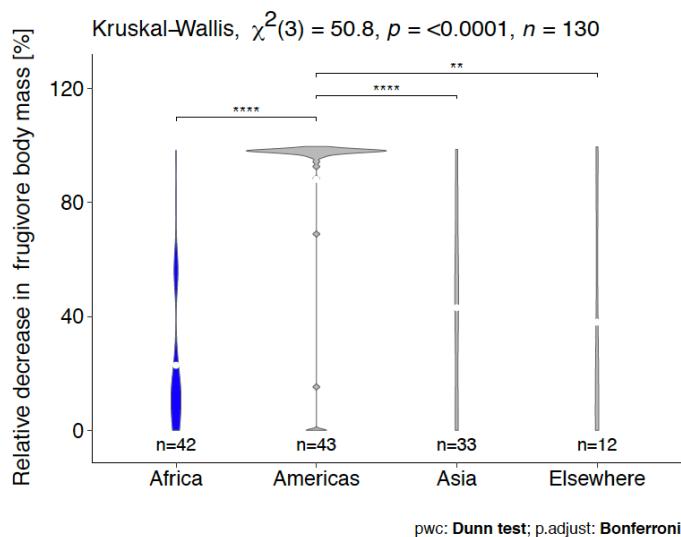


Figure S3: Global representation of relative decrease in mammalian frugivore assemblage body masses due to Late Quaternary extinctions and range contractions. a) the significant group differences between mammalian frugivore assemblages in Africa, Americas, Asia and elsewhere. Violin plot shapes illustrate the density of the raw data. Test statistic (χ^2), significance (p) sample sizes (n) (i.e., number of botanical countries), and significant differences between groups (**<0.001, **<0.01) are based on Dunn pairwise comparison tests. b) Spatial distribution of the relative degree (in percentage, %) of downsizing in kilograms (kg) that regional mammalian frugivore assemblages have experienced due to Late Quaternary extinctions and historic range contractions of still extant (persisting) species. We notice the spatial clustering of frugivore assemblages that experienced weak downsizing (in Africa) and those that were dramatically reduced (elsewhere, especially in the Americas, Australia and mainland Asia). This metric of relative downsizing is helpful to distinguish assemblages with generally very different body sizes among assemblages. Besides others, it better represents the turnover of mammalian frugivore assemblages, as it incorporates the overall body size of past assemblages. Relative body mass decrease was calculated for each assemblage as the percentage of 1 (average current body mass / average past body mass).

a)



b)

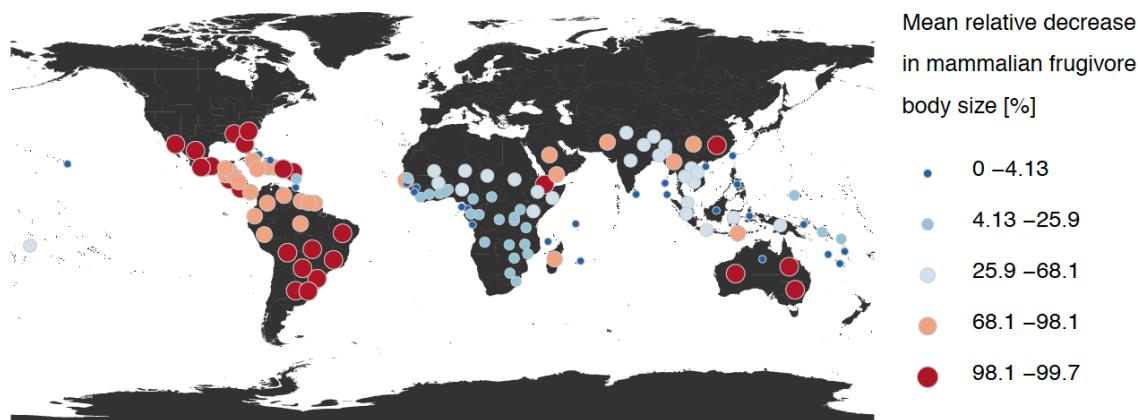


Figure S4: Overview of selected evolutionary trait models. Proportion of evolutionary models selected for the three regimes: Africa, Americas, Asia. OU = Ornstein Uhlenbeck, BM = Brownian Motion. For details on models, see Table 1. AICc weights reflect the support for each model compared to the other models (between 0 and 1).

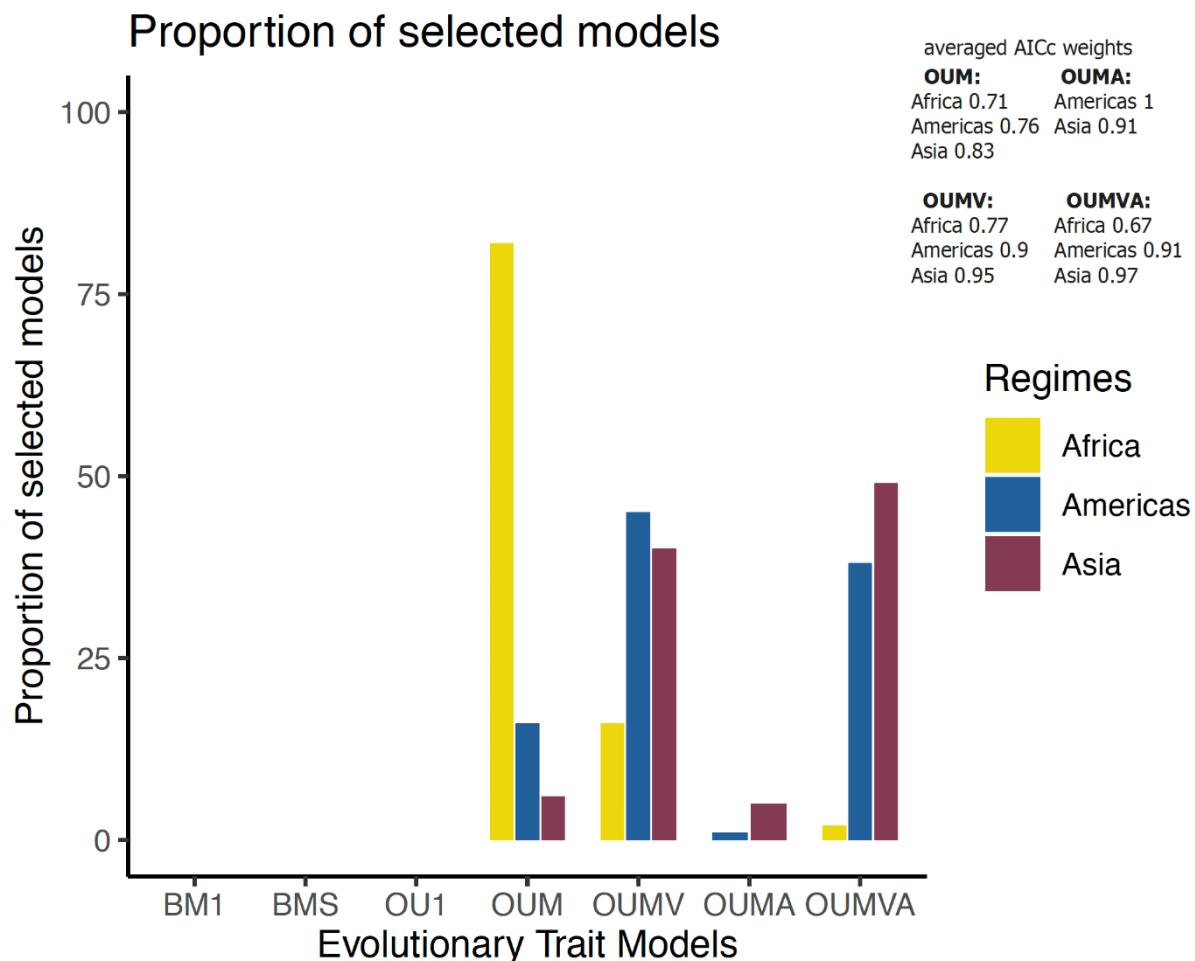


Figure S5: Extinction effect on evolutionary models: results from the sensitivity analysis. Phylogenetic linear regression models to assess the effect of biogeography (Africa vs. elsewhere, Americas vs. elsewhere, Asia vs. elsewhere) on interspecific variation in palm fruit size after removing 90%, 75%, 50%, 25%, 10% and 0% of species with large fruits (i.e., fruits ≥ 4 cm in length) from the analysis. These models were fitted under an Ornstein Uhlenbeck (OU) covariance structure (phylogenetic covariance in the residuals is accounted for in such models); models with OU covariance structure performed significantly better (based on AIC) than those with a Brownian motion covariance structure. The figure shows the ‘effect size’ of continent (Africa, Americas or Asia) on fruit size, defined as the difference from the intercept (elsewhere, i.e., species occurring outside the modelled region). Whiskers represent 95% confidence intervals. Only when removing 90% of African palm species with fruits ≥ 4 cm from the analysis, African palms had no longer significantly larger fruit sizes than palms elsewhere (i.e., 95% confidence interval overlaps with the 0 line). In contrast, American and Asian palms had consistently smaller fruit sizes than palms occurring elsewhere, but only significantly so for the Americas when 50% of American palms with large fruits were removed from the analysis. Overall, this result suggests that recent extinctions of species with large fruits have probably not affected our inference of distinct palm fruit size evolutionary optima in Africa vs. elsewhere.

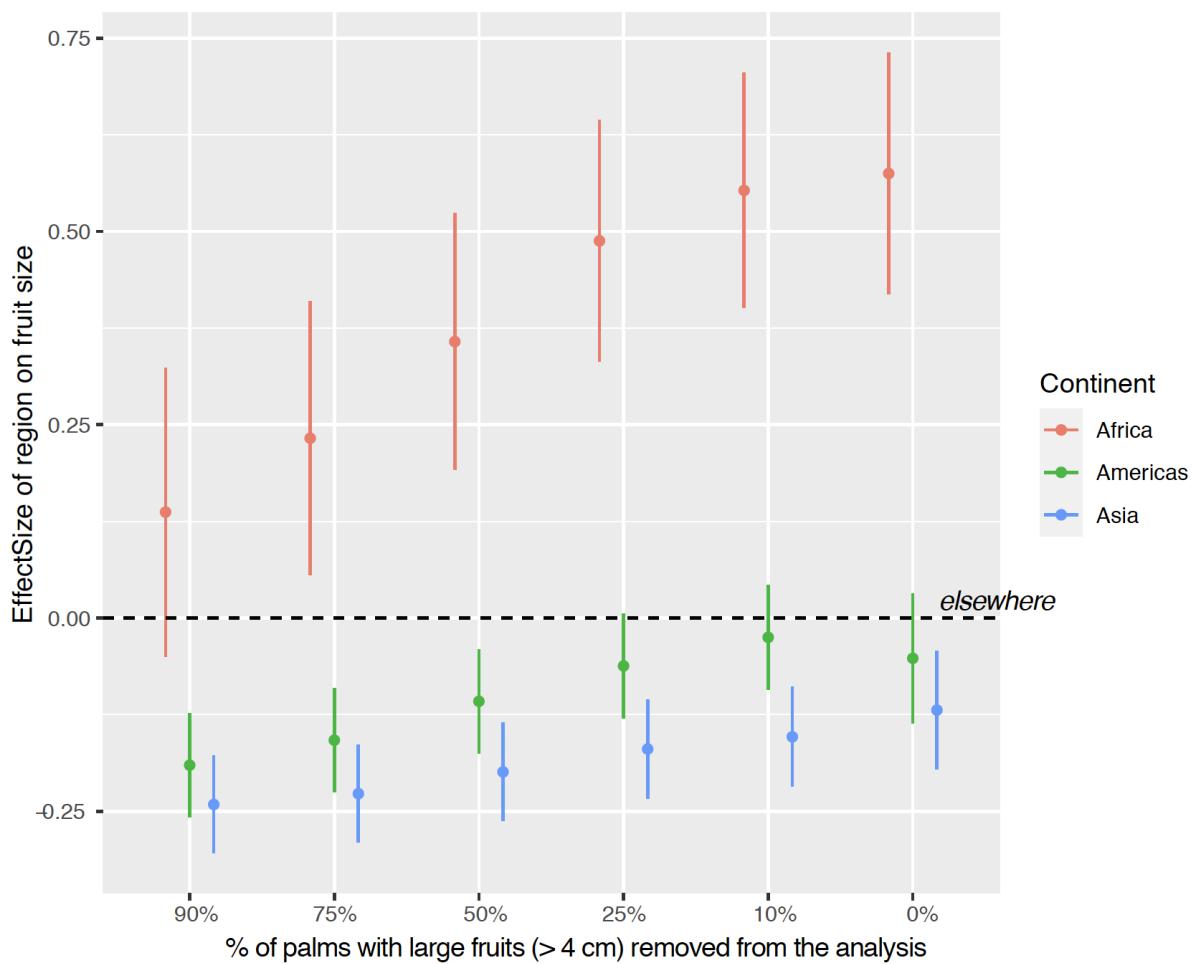
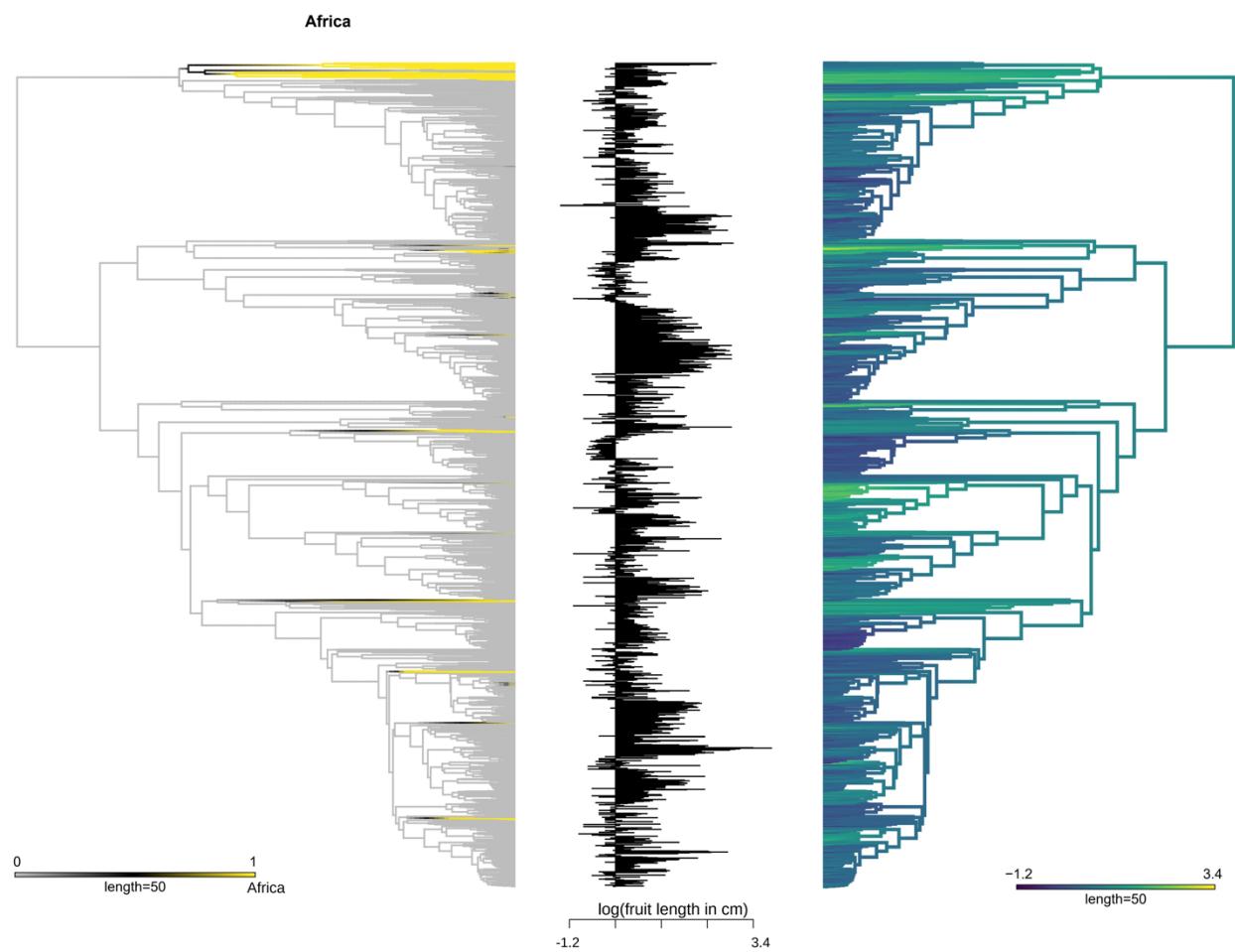
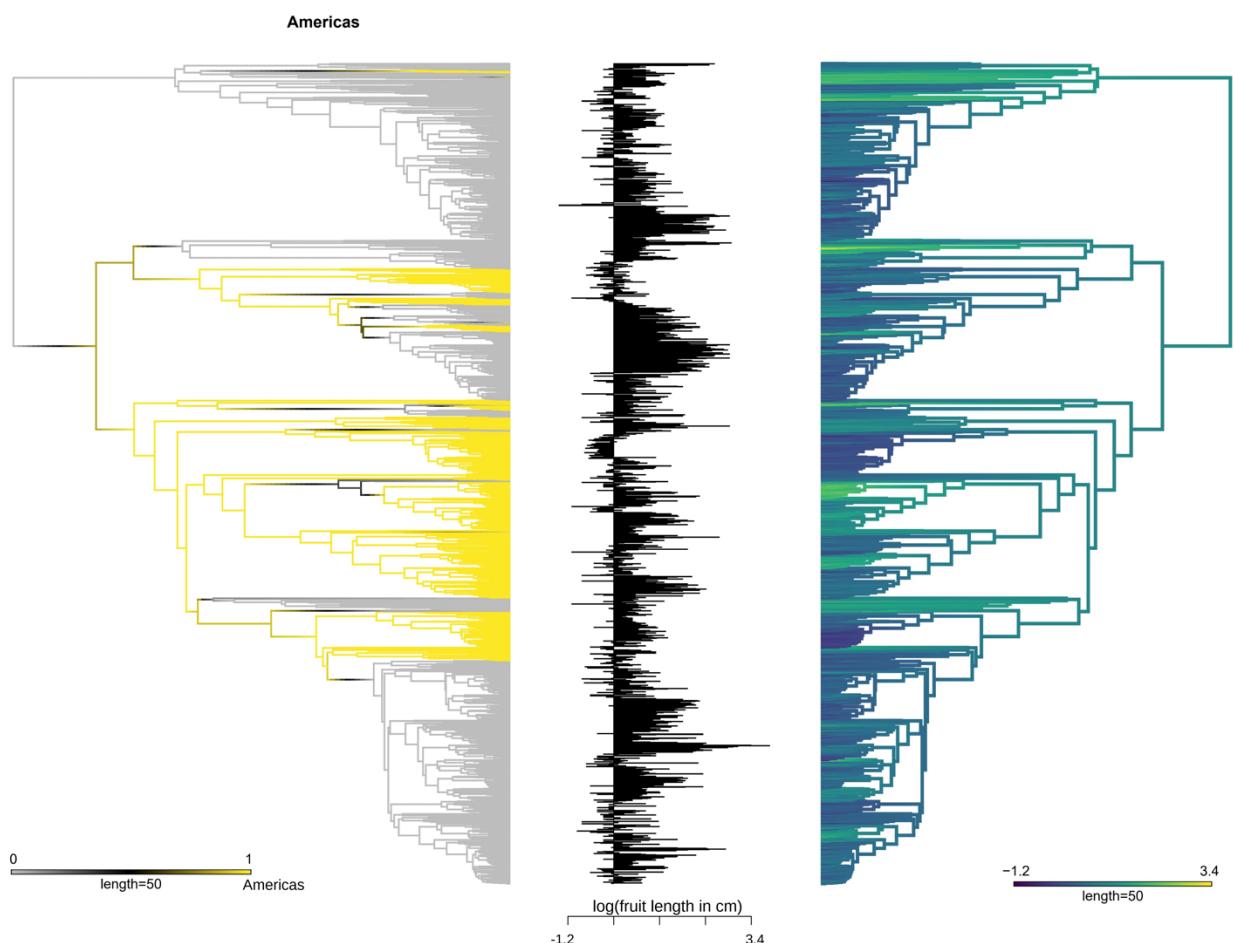


Figure S6: Ancestral state reconstruction of geographical area as a selective regime (left) and fruit size evolution (right) in palms (Areaceae) across Africa (a), the Americas (b), and Asia (c). Selective regimes were inferred through 500 simulations of stochastic character mapping on each of 100 phylogenetic trees, with colored branches denoting high (yellow), intermediate (black), or low (grey) posterior probabilities for each area as selective regime for a given lineage. The ancestral reconstruction of fruit size employed a maximum likelihood approach at the internal nodes, with states then interpolated along each edge using phylogenetically independent contrasts. The figure depicts the maximum clade credibility tree. Fruit size is represented on a log-scale, from -1.2 (i.e., 0.3 cm fruit length) to 3.4 (i.e., 30 cm fruit length), with yellow colored branches representing palm lineages with large fruits, green colored branches representing those lineages with intermediate fruit sizes, and blue/purple colored branches representing lineages with comparatively small fruits. The middle panel displays log-transformed fruit sizes of palm species today, i.e., at the tips of the tree. This comparative analysis of fruit size evolution and selective regimes suggests that large fruits evolved in lineages across various geographical regions.

a)



b)



c)

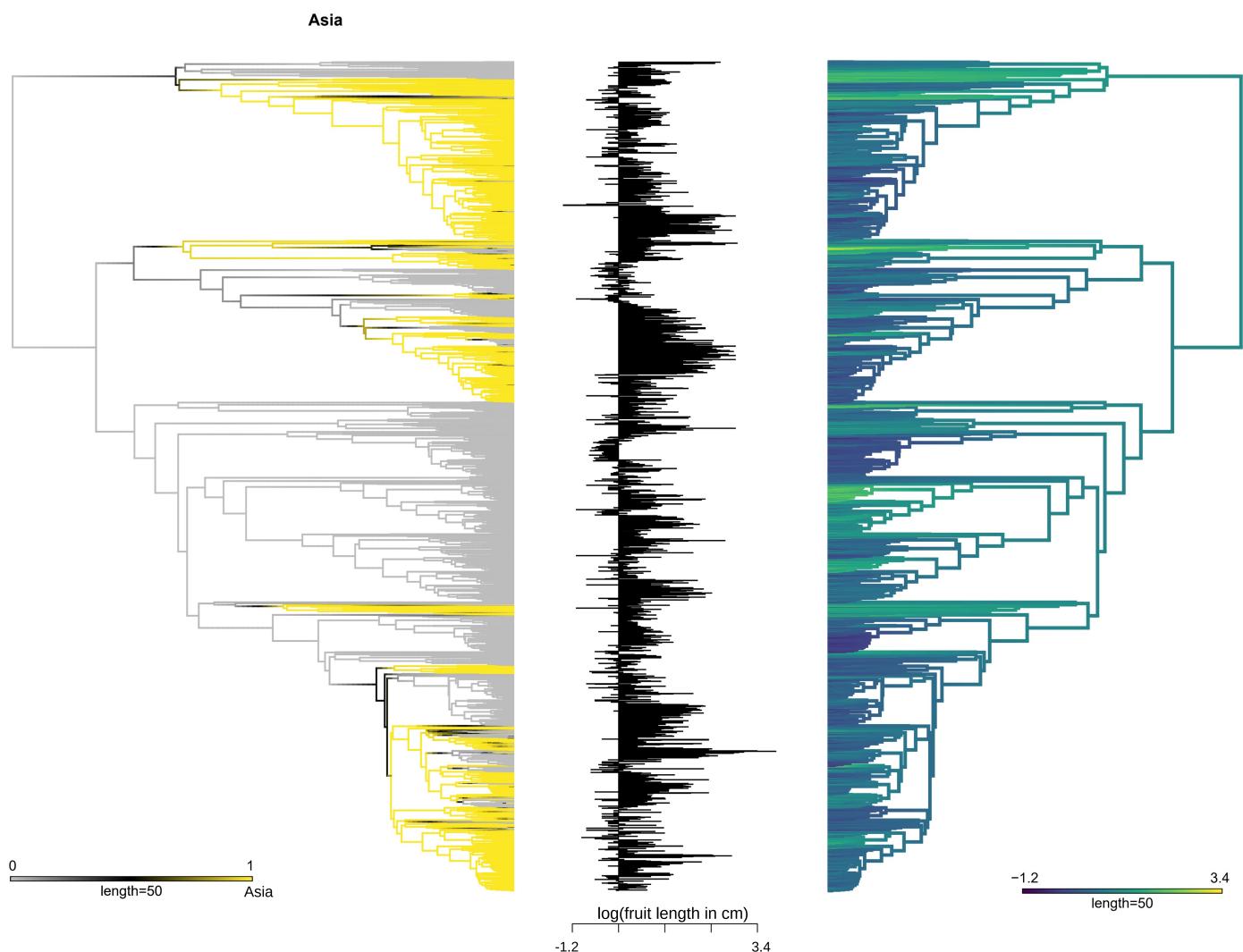
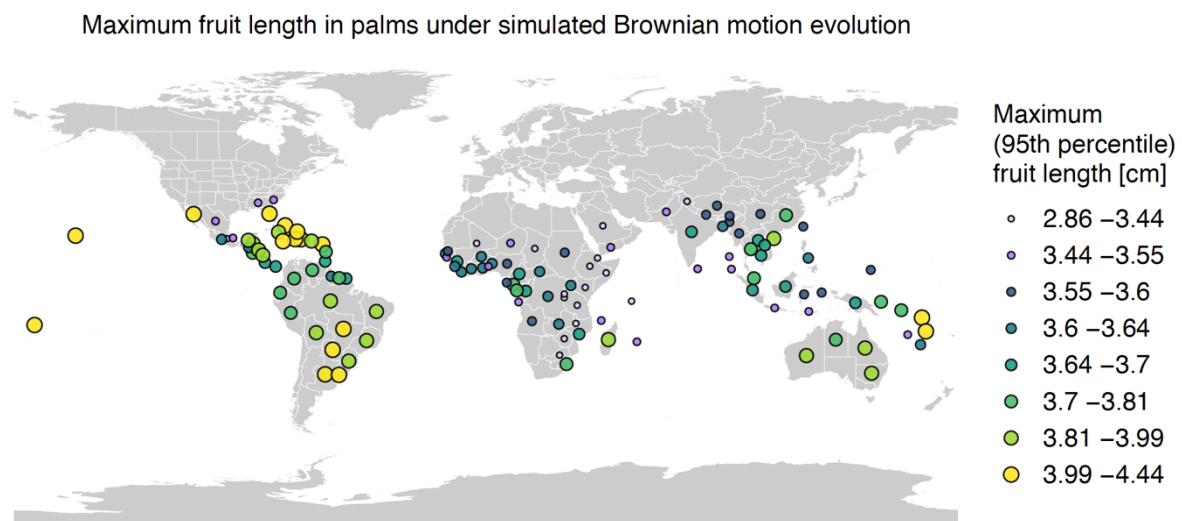


Figure S7: Global distribution of simulated palm fruit sizes. Simulations were repeated 100 times on the palm MCC tree, and averaged based on species presence in botanical countries. Here, we depict fruit size on a scale that includes eight levels (i.e., colors) whereas the map in the main text (Figure 3c) used the same levels as the empirical palm data (Figure 3a). The map below aims to better visualize contrasts between regions.



Supplementary references

- Bollback JP.** 2006. SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC bioinformatics* **7**: 88.
- Boucher FC.** 2018. BBMV: Models for Continuous Traits Evolving in Macroevolutionary Landscapes of any Shape.
- Boucher FC, Démery V.** 2016. Inferring Bounded Evolution in Phenotypic Characters from Phylogenetic Comparative Data. *Systematic biology* **65**: 651–61.
- Burnham KP, Anderson DR (Eds.).** 2004. *Model Selection and Multimodel Inference*. New York, NY: Springer.
- Harmon L.** 2019. *Phylogenetic Comparative Methods: Learning From Trees*. Center for Open Science.
- Holm S.** 1979. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* **6**: 65–70.
- R Core Team.** 2022. R: A Language and Environment for Statistical Computing.
- Revell LJ.** 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**: 217–223.