APPENDIX

A HABITAT-BASED PERSPECTIVE OF MARINE BIOGEOGRAPHY IN PASSIVE AND CONVERGENT TECTONIC SETTINGS

Bjørn-Gustaf J. Brooks & Cinzia Cervato

In this supplement we contribute details that further describe the methods used the main paper. This supplement is divided into five sections. Section is a description of our computational routines that describes how to reproduce our results from the raw data. This is followed by a protocol for converting lithological and depositional categories to numerical values in Sections and . Section describes the computation of neighborhood beta diversity, and section further explores the correlation between alpha diversity and the diversity of k-means habitat types. Finally, an extended table of diversity indexes appears at the end of this supplement.

COMPUTATIONAL ROUTINES

All of our data and scripts can be downloaded as compressed tar or zip file from our repository (www.climatemodeling.org/paleo/ecosystem_drift/). These files can be used to reproduce the working database and our analytical results– however, you will need to use your own k-means clustering algorithm to develop the k-means habitat types (see lines 73-80 of `build_mdiv.bash`). Our scripts were developed for use in a Linux environment using standard utilities (`bash`, `awk`, etc.), the PostgreSQL open source database package, and IDL. The system or database administrator must provide the user with privileges to create PostgreSQL databases. Our diversity indexes were calculated in IDL. All scripts were developed and tested under Fedora Core 6 (`x86_64` release) using PostgreSQL version 8.1.10.

To reproduce our results you will first need the following:

1. Occurrence data (`20080801-occs.csv`) from our repository (www.climatemodeling.org/paleo/ecosystem_drift/) or download your fossil occurrences from www.paleodb.org specifying under "Basic options" 60 and 0 as the "Oldest and youngest intervals" and the following "Collection fields*":

   - "species" (genus names are included automatically)
   - "latitude/longitude in decimal"
   - "midpoint age (Ma)"
   - "lithology 1"
   - "lithology 2"
   - "environment"

2. The scripts contained in the main directory (`mdiv/`) our online repository (www.climatemodeling.org/paleo/ecosystem_drift/).

3. *k*-means clustering software.

*Genus/species names are included by default.

4. IDL data analysis software and the IDL scripts from our repository (`mdiv/idl/`), or you can rewrite these scripts (`gstring.pro`, `mdiv.pro`, `mdiv_pca.pro`, `pca.pro`, `rs_test.pro`) for use in another language.

LITHOLOGICAL DATA

We downloaded 83,213 globally distributed fossil occurrences from the Paleobiology Database (PD) that spanned the most recent 60 million years, each of which were annotated with fields describing lithology and depositional environment. Two fields describing the lithology, primary and secondary, annotated each fossil occurrence. We followed the PD's data entry criteria (http://paleodb.org/public/tips/lithtips.html) in order to assign values of percent sand, silt, clay, and lime mud for each occurrence based in its lithological annotations. A subset of 58,878 occurrences were described only by a primary lithology field, which reflected the entire lithology of the observation. For example, the category claystone designates those occurrences derived from clay rocks that area entirely fine-grained siliciclastics whether lithified or not (excluding mudstone and shale). We assigned claystone-only occurrences a value of 100% in the clay sized grain field, and 0% in all other fields.

The remaining 24,335 occurrences were described by both a primary and secondary lithology. For these the primary field could have accounted for as little as 51% or as much as 99% of the total rock volume, while the secondary 1% - 49%. We took the median of these ranges (primary: 75%, secondary 25%) to proportionally assign values to these occurrences. An occurrence with a primary lithology of claystone (100% clay) and secondary of marl (50% clay, 50% lime mud[†]) would be assigned proportionate values of

[†] Marl consists of 35%-65% clay and 35%-65% lime mud, therefore equal proportions of clay and lime mud are taken (50%:50%)

clay and lime mud based on the 75%:25% ratio between primary and secondary lithologies, such that the occurrence is said to contain 87.5% clay and 12.5% lime mud. The sum of all fields for each occurrence was always equal to or less than 100%. In total our data contained 29 marine lithological categories, and when considering combinations of both primary and secondary fields there were 336 different combinations of the two. Table 1 gives a partial listing of these categories, their equivalent values for different grain sized particles and explanations for these values. The complete listing of all lithological category combinations (mdiv_lith.db) can be found in the data directory of the compressed tar file.

Table 1: Lithological categories. Value equivalents for 8 of 336 lithological categories. See online repository for a complete listing. SA indicates percent sand- SI, percent silt- CL, percent clay- and LM, percent lime mud.

PD environment	**SA**	**SI**	**CL**	**LM**	**Explanatory notes**
bafflestone only	10	10	10	0	Rocks whose chief characteristic is that they are bound by a matrix of organisms that acted as baffles during deposition, and contain smaller grains, therefore a small, but equal fraction of sand silt and clay are assigned for these rocks.
bindstone only	10	10	10	0	Rocks whose chief characteristic is that they are bound by a matrix of organisms, especially algae, and may contain smaller grains, therefore a small, but equal fraction of sand, silt, and clay are assigned for these rocks.
claystone only	0	0	100	0	Fine-grained siliciclastic rocks consisting entirely of clay composed of grains less than 2 mm.
grainstone only	65	0	0	0	Grain supported limestone with a higher proportion of grains than wackestone, which are in contact with one another, but with spar cement, not lime mud, filling interstitial pores. Therefore grainstone is given the same sand-sized grain proportions but with no proportion of lime mud. (Compare mudstone, wackestone, packstone, grainstone, floatstone, rudstone in Dunham 1962/ Embry and Klovan 1971 carbonate classification
lime mudstone only	5	0	0	95	A carbonate (limestone) composed almost exclusively of micrite, but with less than 10% sand grains, therefore a median of 5% sand is taken, leaving 95% lime mud
limestone only	0	0	0	0	This category contains only those specimens that could not be classified

(Continued on next page)

Table 1 (continued from previous page)

PD environment	**SA**	**SI**	**CL**	**LM**	**Explanatory notes**
					inot a more specific carbonate lithology, and includes chalk.
marl only	0	0	50	50	Consists of 35%-65% clay and 35%-65% lime mud, therefore equal proportions of clay and lime mud are taken (50%:50%)
mudstone only	5	47.5	47.5	0	Fine-grained siliciclastic rocks containing silt to clay sized particles, but not less than 33% of either, and less than 10% sand-sized grains. Therefore equal proportions are taken of clay and silt (47.5%:47.5%) and the median of 0% and 10% sand. (Compare mudstone, Wackestone, Packstone, Grainstone, Floatstone, rudstone in Dunham 1962/Embry and Klovan 1971 carbonate classification

DEPOSITIONAL ENVIRONMENT

Each occurrence from the PD included in our data set was annotated with an environment of deposition. These facies interpretation were made by the contributing paleontologist, drawing information from a combination of lithological features, sedimentary structures and geologic context (http://paleodb.org/public/tips/environtips.html). Our analysis was for the marine occurrences beginning with those that occurred in the intertidal zone. We assigned each depositional environment a numerical value for water depth and used this, along with our lithological values (% sand, etc.) to cluster the observations into habitat types. Table 2 provides a complete listing of each PD environment included in our analysis, its equivalent water depth, and our explanatory notes.

Table 2: Lithological facies. Value equivalents for depositional environment categories.

PD environment	Water Depth	Explanatory notes
delta plain	0	Upper delta plain is above high tide, and has no marine influence. The lower delta plain is below, therefore a median of sea level is taken.
interdistributary bay	0	Includes marshes and mudflats. Part of the lower delta plain and sits at sea level.
foreshore	0	Landward of shoreface. Water depth varies with the tide and often becomes emergent. Could include evaporites. The median point at high tide would be below water, but during low tide would be emergent. Therefore a depth of 0 is taken.
sand shoal	1	Shallow water influenced largely by tide/wave action and conducive to ooid formation.
paralic indet.	2	Paralic (cf. marginal marine) represents the longitudinal transition between aquatice continental aquatice waters and marine, and are quite shallow.
shallow subtidal indet.	10	Water depth is greater than transitional zones. May occasionally become emergent.
peritidal	5	May become emergent and could include evaporites. The median of high and low tide is taken.
lagoonal	7.5	Water depth is shallow, yet still below tidal base.
estuary/bay	10	cf. paralic, but ranging to greater depth.
lagoonal/restricted shallow subtidal	10	Below low tide.
shoreface	15	Seaward of foreshore. The Upper shoreface is above wave base and can be agitated by waves, while the lower shoreface is only affected by large storms. The median is taken to be the wave base, as well as subtidal and below the foreshore.
transition zone/lower shoreface	20	Deeper than shoreface and below wave base. Only affected by major

(Continued on next page)

Table 2 (continued from previous page)

PD environment	Water Depth	Explanatory notes
		storm events.
open shallow subtidal	50	Below low tide. Above offshore.
delta front	50	The seaward portion of a delta still near enough to shore in order to be dominated by sand sized clasts.
reef, buildup or bioherm	50	Within the photic zone (i.e. less than 200 m), however most are much closer to surface light.
deltaic indet.	100	Ranging from a few meters above sea level (e.g. upper delta plain) to 200 m below. The median of the range is taken.
coastal indet.	100	Ranges anywhere between 0 and 200 meters, therefore the median of that range is taken.
marginal marine indet.	100	Closer to shore than offshore shelf. Ranges anywhere between 0 and 200 meters, therefore the median of that range is taken.
perireef or subreef	100	Water depth is between 0 and 200 meters, therefore the median of that range is taken.
platform/shelf-margin reef	100	Water depth is between 0 and 200m, therefore the median of that range is taken.
prodelta	100	Water depth is between 0 and 200 meters, therefore the median of that range is taken.
intrashelf/intraplatform reef	100	
offshore shelf	150	Farther offshore than marginal environments, but landward of the slope.
slope/ramp reef	150	Water depth is on the verge of the photic zone.
deep subtidal ramp	200	On the verge of the continental shelf.
marine indet.	200	Ranging from marginal marine to open marine. A median of 200 m is taken.
offshore	200	
offshore indet.	200	
carbonate indet.	500	Water depth is unknown, therefore the median of the marine water depth range is taken.
offshore ramp	600	

(Continued on next page)

Table 2 (continued from previous page)

PD environment	Water Depth	Explanatory notes
slope	600	
submarine fan	600	
basinal (carbonate)	1000	
basinal (siliceous)	1000	
basinal (siliciclastic)	1000	
basin reef	1000	
deep subtidal indet.	1000	
deep subtidal shelf	1000	
deep-water indet.	1000	

NEIGHBORHOOD BETA DIVERSITY

To measure the difference in diversity between assemblages we computed a neighborhood beta diversity (see `supplement/idl/mdiv.pro` in our repo: http://www.climatemodeling.org/paleo/ecosystem_drift/). We calculated our beta diversity as the difference between assemblages following Whittaker's method (Whittaker 1972) with one modification. Our beta diversity is equivalent to the mean of the betas between the assemblage of interest, A_i, and the assemblages within the immediate neighborhood of A_i. For each A_i that was surrounded by a number of occupied assemblages N where $1 \leq N \leq 8$, beta diversity was calculated as $\frac{D_c}{\overline{D}} - 1$, where D_c is the total diversity of species (or genera) in the composite of two assemblages and $\overline{D}$ is the mean diversity of species (or genera) in both assemblages. Empty assemblages, those without fossil occurrences, were excluded (Table 3). Where $N = 1$, there calculation of beta diversity was straight forward, but for neighborhoods where $2 \leq N \leq 8$, neighborhood beta diversity was calculated as the mean of all betas. To complete the neighborhood for assemblages located at geographical edges (*e.g.*, longitude 179°W to 180°W), we wrapped these neighborhoods to include assemblages at the opposite geographic extreme (*e.g.*, longitude 180°E to 179°E). No occupied assemblages in our data set appeared at latitudes above 89° north or south, therefore it was not necessary to wrap over the poles.

PRINCIPAL COMPONENT ANALYSIS

To further explore our analytical methods, we compared the correlation outcome of alpha

Table 3: Neighborhood beta diversity map. The assemblage of interest, indicated here by A_i, lies at the center of the grid of nine. In the example blow beta diversity for A_i is calculated as the mean of all β between A_i and its occupied adjacent assemblages (0.80, 0.70), which is equivalent to 0.75 for A_i.

0.80	null	null
0.70	A_i	null
null	null	null

diversity to the means of each environmental parameter (water depth, percent sand, etc.) and to the number of habitat types after PCA rotation of the axes. First we performed PCA ordination on all 5 environmental parameters, clustered (k-means) them, then compared the results between assemblages just as described above. PCA rotation of the 5 environmental parameters resulted in essentially no correlations to alpha diversity (water depth: $r^2 = -0.027$, pct. sand: $r^2 = -0.034$, pct. silt: $r^2 = 0.003$, pct. clay: $r^2 = -0.011$, pct. lime mud: $r^2 = -0.141$), indicating that PCA does not optimize correlations in our data better than k-means clustering. PCA ordination and clustering of all 5 parameters into habitat types also resulted in a reduced correlation to alpha diversity ($r^2 = 0.677$), which indicates that PCA rotation does not improve the correlation (or reduce noise) better than k-means for our data. In the last method we performed PCA ordination, but clustered only the three most principal factor loadings, which accounted for 81% of the total variance in the data, and resulted in a correlation between habitat types and alpha diversity ($r^2 = 0.672$) less than that of k-means clustering alone. Data rotation and dimension reduction from PCA therefore, do not improve the correlation of any of our environmental parameters to alpha diversity.

Table 4: Genus level data composition. Generic abundances of select phylogenetic classes, total genera, and the number of habitat types within each sub-epoch are listed. Epoch and sub-epoch ages correspond to the ages of their base.

Interval	Time (Ma)	Duration (m.y.)	Bivalve Genera	Gastropod Genera	Total Genera	Habitat Patches
Holocene	0.01	0.01	147	183	437	15
Pleistocene	2.6	2.59	342	513	1450	29
Pliocene	5.3	2.7	345	510	1632	28
late Miocene	11.6	6.3	360	493	1400	33
middle Miocene	16.0	4.4	378	489	1702	33
early Miocene	23.0	7.0	353	572	1486	32
late Oligocene	28.4	5.4	200	232	733	30
early Oligocene	33.9	5.5	224	326	793	25
late Eocene	40.4	6.5	273	457	1423	34
middle Eocene	48.6	8.2	264	401	1332	32
early Eocene	55.8	7.2	161	249	762	19

Table 5: Local species diversity by region. Mean abundances per grid cell (10,000 km^2) for alpha, beta, habitat type, source pool diversity, and connectivity are given for each sampling interval. Asterisks indicate peak values, except for early Eocene European beta diversity, which was a local peak.

Interval	Region	α-Div.	β-Div.	Habitat Types	Source Pool Div.	Connectivity
Holocene	E. N.Am.	40.1	0.00	2.2	1.0	0.00
	Europe	52.9	0.00	6.3	1.2	0.00
	Japan	4.3	0.67	1.3	1.9	0.08
Pleistocene	E. N.Am.	77.8	0.83	5.0	208.0	0.40
	Europe	21.0	0.52	2.3	30.5	0.16
	Japan	16.1	0.75	2.2	16.5	0.14
Pliocene	E. N.Am.	75.3	0.77	4.5	202.3	0.35
	Europe	22.0	0.72	3.1	30.0	0.14
	Japan	26.2	0.70	3.3	26.7	0.12
late Miocene	E. N.Am.	50.2	0.89	5.1	85.9	0.21
	Europe	26.7	0.73	2.8	50.9	0.19
	Japan	50.8*	0.72	4.7*	94.6	0.21
middle Miocene	E. N.Am.	69.3	0.56	4.8	90.4	0.11
	Europe	48.0	0.67	4.5	102.3	0.23
	Japan	27.9	0.84*	2.8	43.5	0.20
early Miocene	E. N.Am.	81.2*	0.88	5.8*	153.2	0.30
	Europe	45.0	0.48	3.7	30.6	0.10
	Japan	14.3	0.58	2.7	12.1	0.11
late Oligocene	E. N.Am.	12.0	0.22	2.9	1.8	0.03
	Europe	42.8	0.24	5.1	9.0	0.04
	Japan	27.6	0.62	4.9	23.8	0.12
early Oligocene	E. N.Am.	73.0	0.72	4.7	131.6	0.30
	Europe	45.1	0.46	4.0	4.9	0.09
	Japan	38.1	0.12	2.5	9.3	0.02
late Eocene	E. N.Am.	61.0	0.80	4.7	127.6	0.21
	Europe	70.4*	0.55	7.0*	94.5	0.12
	Japan	18.5	0.57	3.4	21.4	0.10
middle Eocene	E. N.Am.	58.0	0.85	5.2	90.6	0.23
	Europe	61.4	0.55	4.5	126.9	0.16
	Japan	36.5	0.65	4.0	47.2	0.22
early Eocene	E. N.Am.	33.5	0.80	6.6	48.2	0.22
	Europe	41.4	0.66*	4.0	106.0	0.26
	Japan	22.8	0.00	1.1	1.1	0.00

Table 6: An extended table of genus level correlations. Each column lists two Pearson correlations (r) each for alpha and beta diversity that correspond to unstandardized and rarefied diversity (e.g. [unstandardized correlation,rarefied correlation]). All correlations are significant ($P > 0.001$, two-tailed test), except alpha diversity:mean depth, and beta diversity:mean depth,mean sand, silt, and clay pct. Note the substantial difference between unstandardized and rarefied correlations for connectivity (see Discussion).

Parameter	α-div.	β-div.
Hab. Type Div.	0.857 (0.762)	-0.312 (0.347)
Connectivity	0.310 (0.365)	0.294 (0.743)
Mean Depth	-0.036 (-0.009)	0.022 (-0.046)
Depth Range	0.573 (0.270)	-0.181 (0.213)
Mean Sand Pct.	0.182 (0.091)	0.033 (0.131)
Mean Silt Pct.	0.190 (0.012)	-0.087 (0.011)
	Europe	
Hab. Type Div.	0.783	0.040
Connectivity	0.110	0.692
	Japan	
Hab. Type Div.	0.768	0.000
Connectivity	0.176	0.014
	Eastern North America	
Hab. Type Div.	0.868	0.011
Connectivity	0.193	0.561

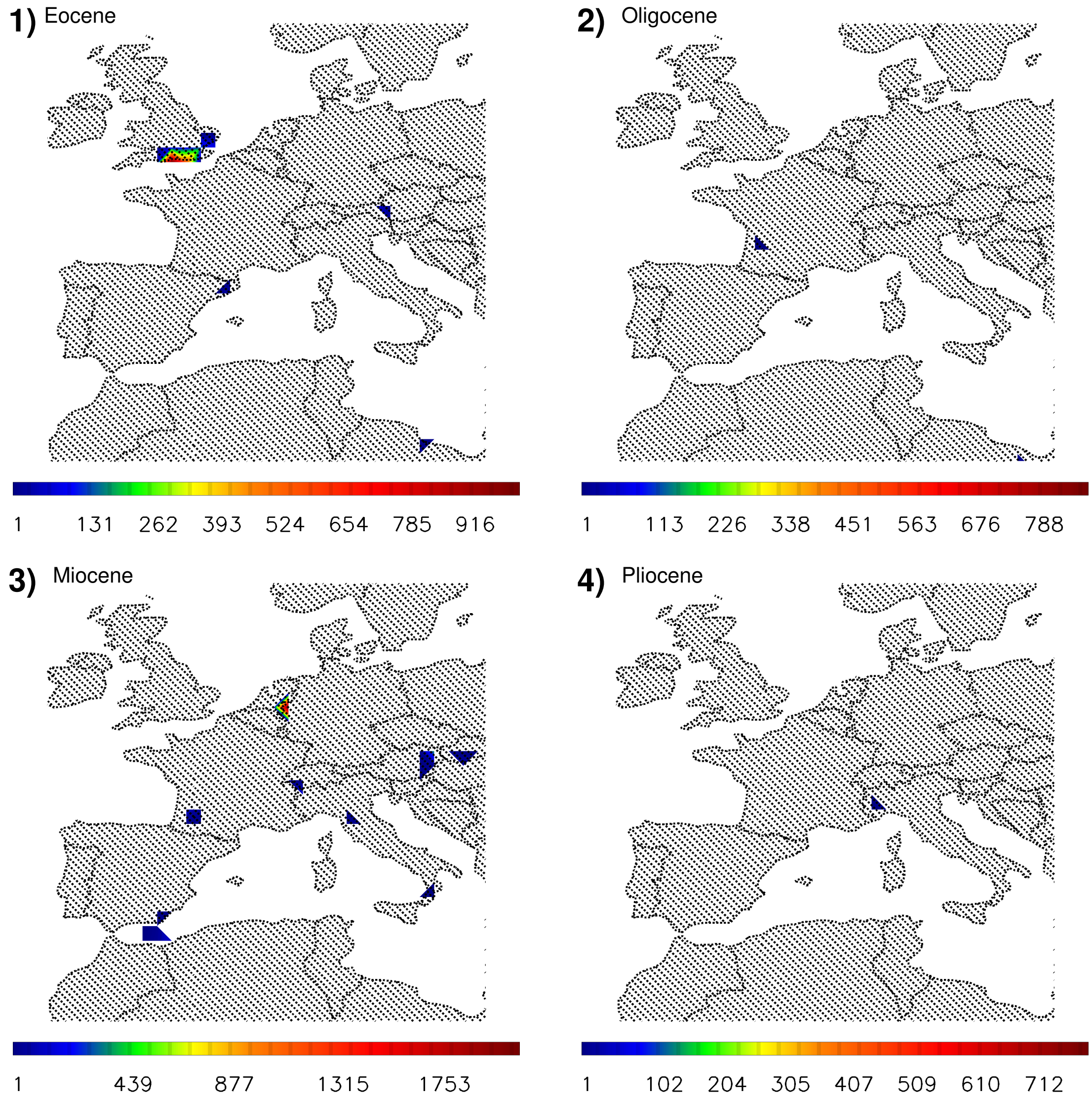

Figure 1: Alpha diversity of Europe by sub-epoch (1. Pliocene, 2. Miocene, 3. Oligocene, 4. Eocene). As in the main paper alpha diversity is indicated by the map legend colors.

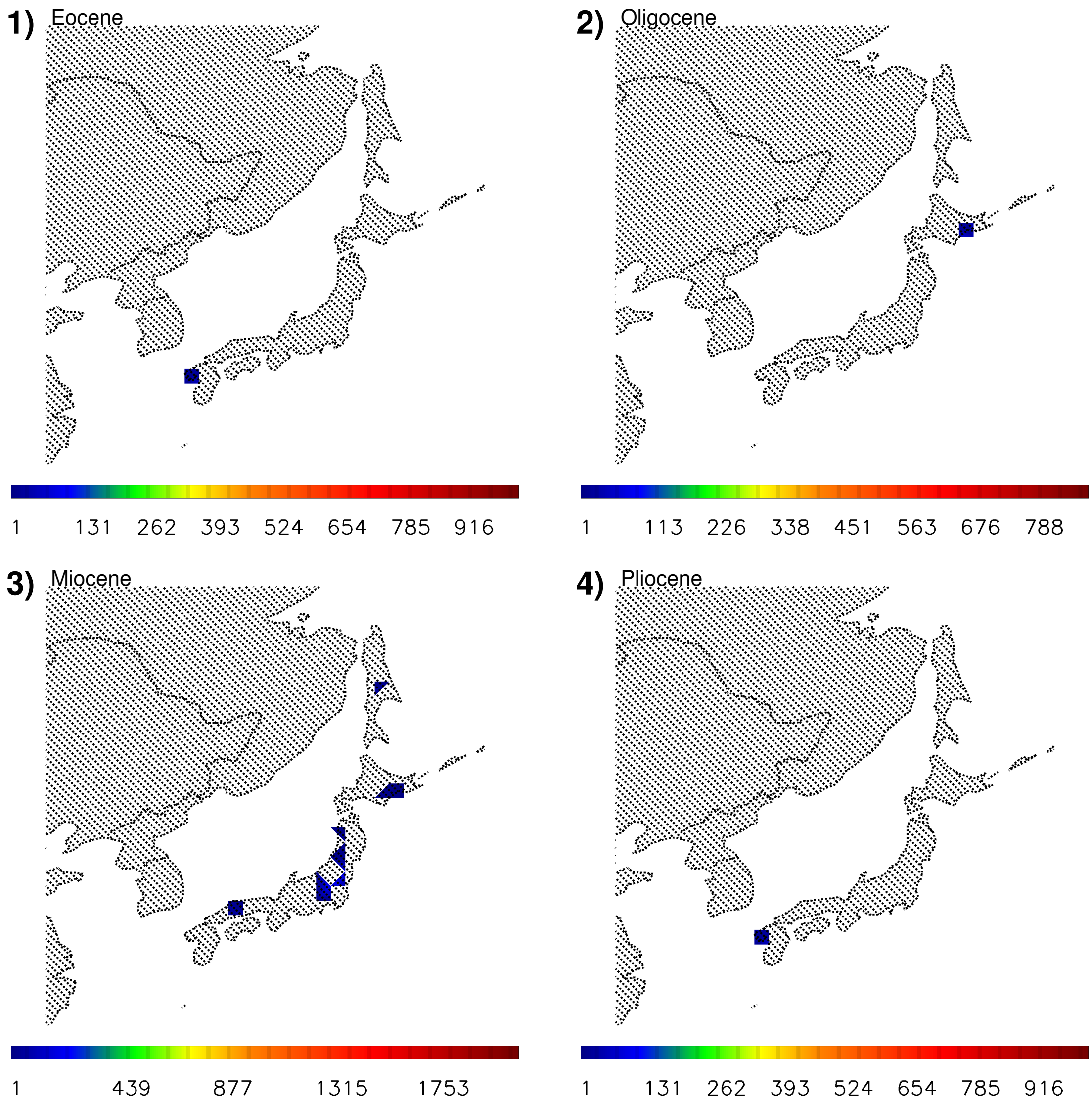

Figure 2: Alpha diversity of Japan by sub-epoch (1. Pliocene, 2. Miocene, 3. Oligocene, 4. Eocene).

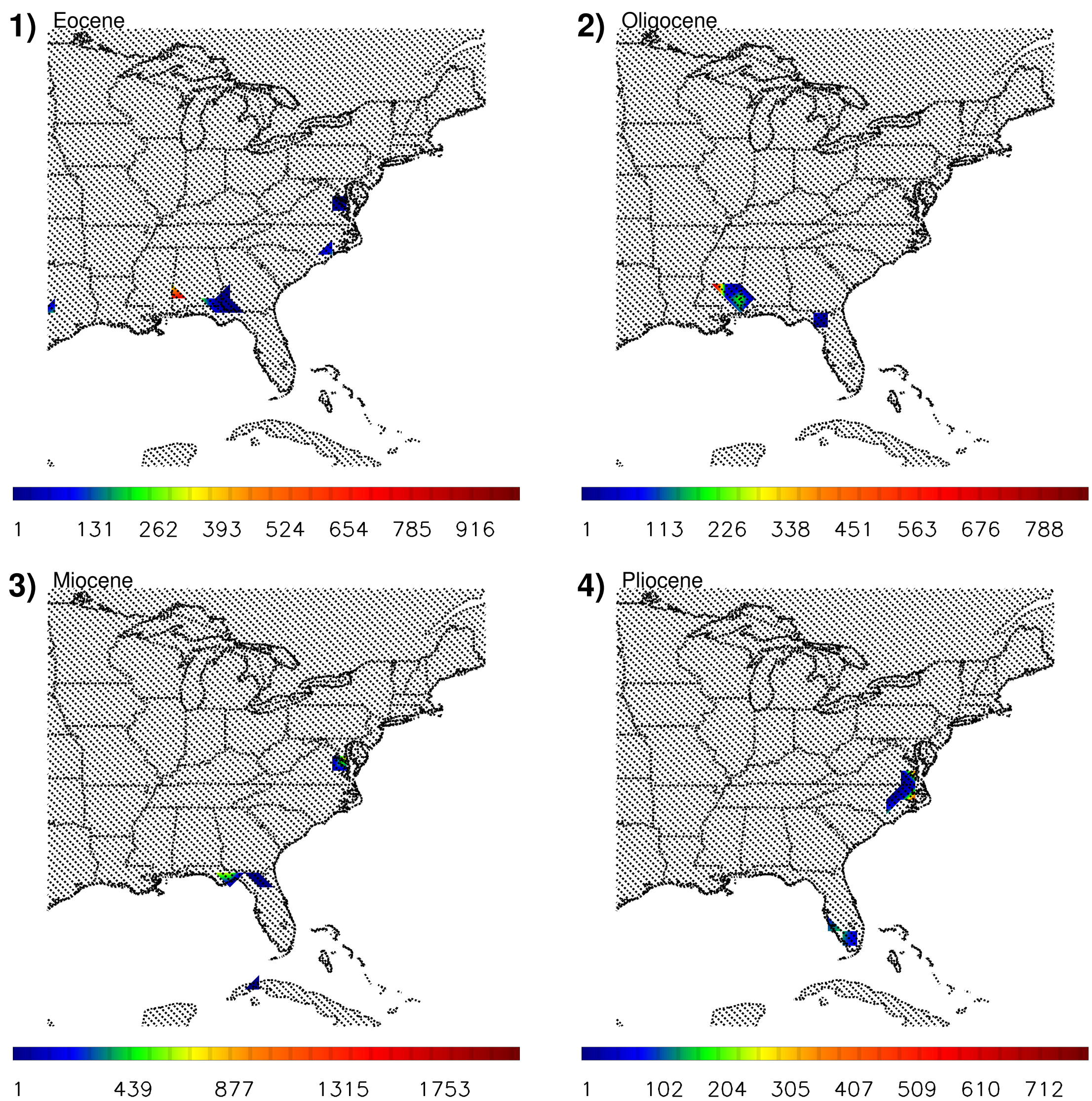

Figure 3: Alpha diversity of eastern North America by sub-epoch (1. Pliocene, 2. Miocene, 3. Oligocene, 4. Eocene).