Description of Modeling – Here, I provide an in-depth description of the global-scale simulation-based probabilistic community models used in the main manuscript, results of which are presented in figures 5a and 5b.

Framework – The foundations of the modeling can be summarized as randomly choosing 1000 locations on Earth and assembling communities at those locations from a global pool of 1000 species. Each community was represented as a vector (or list) of abundances. The set of communities was then represented as a site-by-species matrix, a common ecological data structure whereby sites correspond to rows and species identities correspond to columns. The site-by-species matrix allows straightforward calculations of pair-wise compositional similarity among sites.

Community assembly characteristics – Communities were assembled according combinations of the following 7 characteristics (see Table S1). Below, the subscript *i* pertains to the ith species, N_i is the abundance of the *i*th species, and p_i is the probability of the i^{th} species occurrence:

1. *Dispersal limitation* (DL): N_i and p_i were proportional to the distance (D) of a location from the center of a species geographical range. $p_i \propto 1/(1+D)$; $N_i = 10^4 \cdot p_i$.

2. *Differential growth* (DG): N_i was determined by a randomly chosen parameter (g_i) that limited species abundance independent of D, such that $N_i = g_i \cdot 10^4 \cdot p_i$.

3. *Differential dispersal* (DD): N_i and p_i were partly determined by a randomly chosen parameter (d_i), such that $p_i \propto d_i / (d_i + D)$.

4. *Randomly distributed geographical ranges* (RDR): Centers of species geographic ranges were chosen at random using the random-walk algorithm presented herein.

5. *Latitudinal Diversity Gradient* (LDG): Centers of species geographic ranges were chosen via a normally distributed probability with a mean at latitude 0 and probability of ~ 0 at latitudes of 90 and -90. Longitudinal values (-180, 180) were treated as equally likely.

6. *Environmental niche* (EN): N_i and p_i were proportional to the mean difference (D) between the environmental conditions of the current location and the environmental optimums of the i^{th} species. $p_i \propto 1/(1+D)$; $N_i = 10^4 \cdot p_i$.

7. *Aquatic or Terrestrial* (AT): Occurrence was determined by whether a species was aquatic or terrestrial and whether a given location was aquatic or terrestrial, using a 50m resolution map.

6 dispersal-driven community models – Here, I describe the 6 models constructed from assembly characteristics. These models were not meant to capture a factorial design.

Model 1. Dispersal limitation + randomly distributed geographical ranges. This model was ecologically neutral in that species did not differ in geographic range size, inherent growth or maximum abundance. Geography was homogeneous with no distinction between land and water and no accounting for the distribution of environmental variables. Dispersal was limited, which gave rise to restricted geographical ranges. Geographical ranges were circular and equivalent among species, with randomly distributed centers.

Model 2. Dispersal limitation + randomly distributed geographical ranges + latitudinal diversity gradient. This model only differed from model #1 in that it included a latitudinal diversity gradient (LDG), as described above.

Model 3. Dispersal limitation + randomly distributed geographical ranges + differential growth. This model only differed from model #1 in that it allowed species to differ in their inherent growth (or inherent maximum abundance), as described above.

Model 4. Dispersal limitation + randomly distributed geographical ranges + differential growth + latitudinal diversity gradient. This model only differed from model #3 in that it included a latitudinal diversity gradient (LDG), as described above.

Model 5. Dispersal limitation + randomly distributed geographical ranges + differential growth + differential dispersal. This model only differed from model #3 in that it allowed species to attain geographical ranges of different size, reflecting inherent differences in dispersal and/or colonization, as described above.

Model 6. Dispersal limitation + randomly distributed geographical ranges + differential growth + differential dispersal + latitudinal diversity gradient. This model only differed from model #5 in that it included a latitudinal diversity gradient (LDG), as described above.

6 environmentally-driven community models – The following models were constructed from niche-based characteristics. Each model used 12 environmental data layers downloaded from NASA's Earth Observatory (https://earthobservatory.nasa.gov/) as geotiffs. These included: Topography, land surface temperature, sea surface temperature, vegetation density, net primary productivity, chlorophyll concentration, total rainfall, sea ice and snow concentration, leaf area index, human population density, and permafrost. Rather than focus on the importance of these variables, I simply used each for its realistic geographical distribution. I converted each geotiff to grey scale which produced a gradient of values ranging between 0.0 and 1.0 for each environmental data later. Species optimums for each environmental variable were chosen randomly chosen within the range of 0 - 1. These procedures allowed me to simulate abundance and probabilities of occurrence for each species at each site according to environmental-driven niche dynamics resulting from the average difference between environmental conditions and species optimums, as described under "Community assembly characteristics".

Model 7. Environmental niche. The abundance and occurrence of species in this model were solely determined by the procedure described above. That is, the abundance and occurrence of each species at each site was determined by the average difference between the 12 environmental conditions and the 12 corresponding species optimums. There was no explicit distinction between land and water and no simulation of species geographical ranges (species could occur wherever the environment permitted).

Model 8. Environmental niche + latitudinal diversity gradient. This model was identical to model #7 except that it included a latitudinal diversity gradient, as described under "Community assembly characteristics".

Model 9. Environmental niche + dispersal limitation. This model was identical to model #7 except that it included randomly chosen geographical ranges and a probability of occurrence that was determined by the distance of a location from the center of the species' geographical ranges, as described under "Community assembly characteristics".

Model 10. Environmental niche + dispersal limitation + latitudinal diversity gradient. This model was identical to model #9, except that it included a latitudinal diversity gradient, as described under "Community assembly characteristics".

Model 11. Environmental niche + dispersal limitation + Aquatic or Terrestrial. This model was identical to model #9, except that it accounted for whether a species was designated as aquatic or terrestrial and whether an environment was aquatic or terrestrial, as described under "Community assembly characteristics". Determining whether a given location on Earth was located on land or in water was easily done using python-based geographical libraries (e.g., cartopy, geopy, shapely); see https://github.com/klocey/GlobalDef.

Model 12. Environmental niche + dispersal limitation + Aquatic or Terrestrial + Latitudinal diversity gradient. This model was identical to model # 11 except that it included a latitudinal diversity gradient, as described under "Community assembly characteristics".

Figure S1. 95% confidence hulls relating the slope of the distance-decay relationship (DDR) to average geographical distance among communities. Each of these 6 models excluded niche dynamics. Vertical dashed lines represent the average distance expected from a hemispheric scale distribution. **Top row**: Abundance and occurrence of species was determined solely by distance (*D*) of a location from the center of species ranges. **Center row**: Abundance and occurrence of species was determined by *D* and by specific growth rates (*g*). **Bottom row**: Abundance and occurrence were determined by *D*, *g*, and by specific dispersal rates.

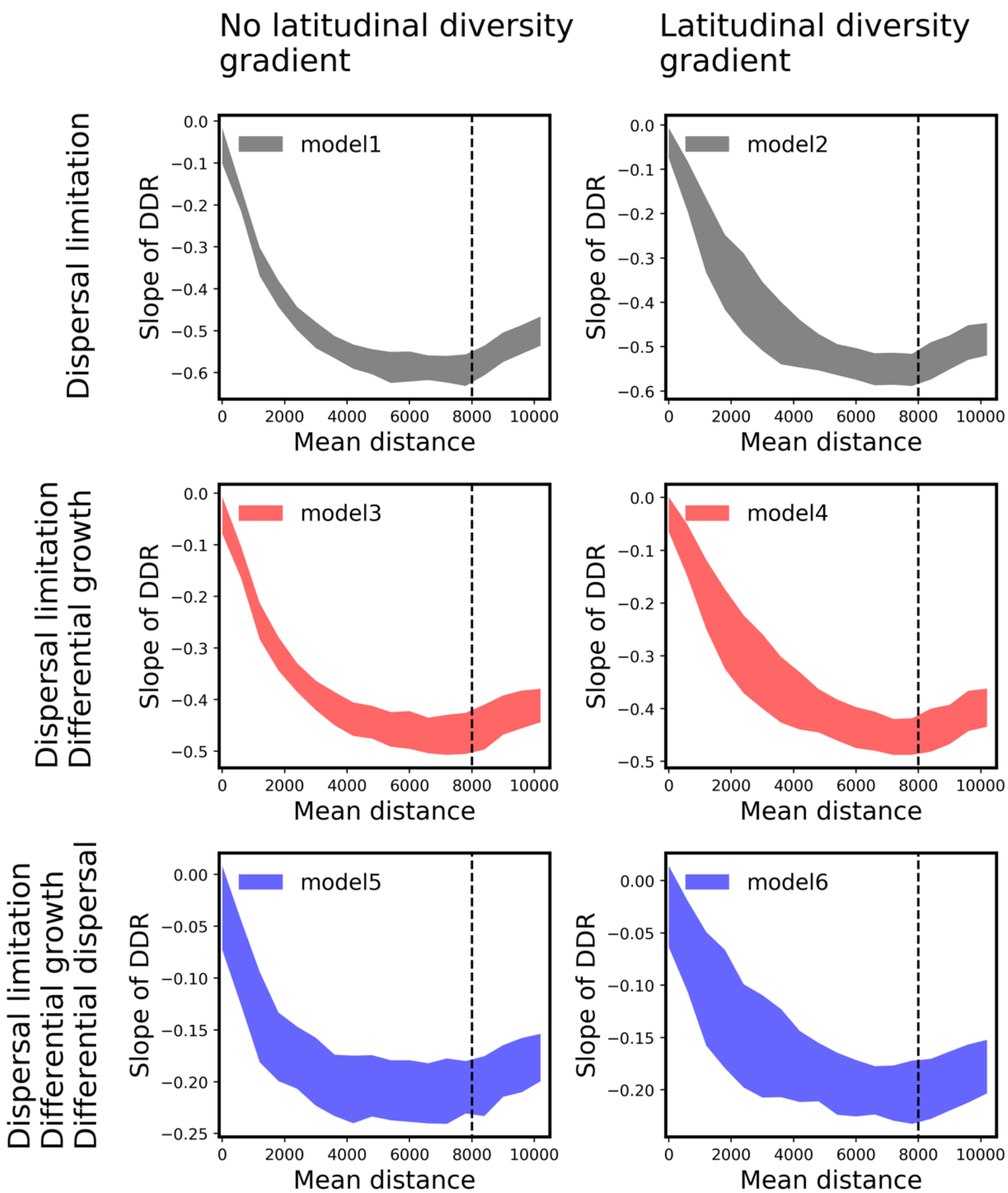

Figure S2. 95% confidence hulls relating the slope of the distance-decay relationship (DDR) to average geographical distance among communities. Each of these 6 models included niche dynamics. Vertical dashed lines represent the average distance expected from a hemispheric scale distribution. **Top row**: Abundance and occurrence of species was determined solely by the average difference (*D*) between environmental conditions of a location and species optimums. **Center row**: Abundance and occurrence of species was determined by *D* and by species-specific dispersal (*d*). **Bottom row**: Abundance and occurrence were determined by *D*, *d*, and by whether species and environments were aquatic or terrestrial.