Global determinants of the distribution of insect genetic diversity

Authors: Connor M French[1,2*], Laura D Bertola[1,3], Ana C Carnaval[1,2], Evan P Economo[4], Jamie M Kass[4], David J Lohman[1,2,5], Katharine A Marske[6], Rudolf Meier[7,8], Isaac Overcast[2,9,10], Andrew J. Rominger[11,12], Phillip Staniczenko[13], and Michael J Hickerson [1,2,14]

[1] Department of Biology, City College of New York, New York, New York, USA
[2] Biology Ph.D. Program, The Graduate Center, City University of New York, New York, New York, USA
[3] Section for Computational and RNA Biology, Department of Biology, University of Copenhagen, Copenhagen N 2200, Denmark
[4] Biodiversity and Biocomplexity Unit, Okinawa Institute of Science and Technology, Onna, Okinawa, JAPAN, 904-0495
[5] Entomology Section, National Museum of Natural History, Manila, Philippines
[6] Department of Biology, University of Oklahoma, Norman, Oklahoma, USA
[7]Humboldt University, Berlin, Germany
[8] Museum für Naturkunde, Berlin, Germany
[9] Institut de Biologie de l'Ecole Normale Superieure, Paris, France
[10] Department of Vertebrate Zoology, American Museum of Natural History, New York, New York, USA
[11] School of Biology and Ecology, University of Maine, Orono, ME, USA
[12] Maine Center for Genetics in the Environment, University of Maine, Orono, ME, USA
[13] Department of Biology, Brooklyn College, Brooklyn, New York, USA
[14] Division of Invertebrate Zoology, American Museum of Natural History, New York, New York, USA

*email: cfrench@gradcenter.cuny.edu

Abstract

Understanding global patterns of genetic diversity (GD) is essential to describe, monitor, and preserve the processes giving rise to life on Earth. To date, efforts to map macrogenetic patterns have been restricted to vertebrate groups that comprise a small fraction of Earth's biodiversity. Here, we construct the first global map of predicted insect genetic diversity. We calculate the global distribution of GD mean (GDM) and evenness (GDE) of insect assemblages, identify the global environmental correlates of insect GD, and make predictions

for undersampled regions. Based on the largest and most species-rich single-locus genetic dataset assembled to date, we find that both GD metrics follow a bimodal latitudinal gradient, where GDM and GDE correlate with contemporary climate variation. Our models explain 1/4 and 1/3 of the observed variation in GDM and GDE in insects, respectively, making an important step towards describing global biodiversity patterns in the most diverse animal taxon.

Introduction

Describing global patterns of biodiversity is essential for understanding and protecting processes governing the distribution of life across the world. To date, such global-scale assessments have largely focused on species richness [1], phylogenetic diversity [2,3], species abundances [4,5], and functional trait diversity [6,7]. These macroecological metrics have long been used to inform conservation and gain insights into mechanisms underlying eco-evolutionary patterns. Only recently, however, have the advances in high-throughput DNA metabarcoding been utilized for global studies of biodiversity [8–11].

Large-scale georeferenced DNA barcode surveys [8,12] have great potential beyond their original use as a tool for identifying and delimiting species. They aid in identifying adaptive potential and ecosystem resilience to disturbance [13], and more generally, help to understand how intraspecific variation can help support critical ecological functions [14]. Along with being an important new component of the macroecological toolbox for conservation action [15,16], the promise of these eco-evolutionary insights is fueling the rise of the emerging field of "macrogenetics" [17,18]. Macrogenetic studies summarize the geographic distribution of average intraspecific genetic variation across species assemblages to find previously unknown patterns and processes underlying the generation and maintenance of biodiversity [19].

To date, global-scale macrogenetic studies have focused solely on vertebrate groups, uncovering links between aggregated genetic diversity, species richness, and phylogenetic

diversity [20,21], while documenting latitudinal gradients in aggregated genetic diversity [22–24]. Macrogenetic studies have also provided mixed support for the general influence of human disturbance on genetic diversity [20,22,25,26], while climate stability [20,26] and species' range sizes [27] have been shown to influence intraspecific genetic diversity on a global scale.

The existing bias toward vertebrate macrogenetics leaves undocumented the bulk of the planet's animal biodiversity: insects. Insects are vital for maintaining critical ecosystem services and functions [28,29], yet to date insect macrogenetic studies have been restricted to the regional scale due to the immense effort required to collect, identify, and sequence such a speciose taxon [30–33]. There also is little agreement as to what extent insect communities are resilient to global change [34,35], including biological invasions [36,37], habitat conversion [38], and climate change [39]. Here, we present the first global macrogenetic analysis of this large group, which is especially important given increasing evidence that many insect taxa may be in global decline with respect to occurrence, local richness, abundance and biomass [35,40–46].

Unlike for most terrestrial vertebrates, comprehensive knowledge of species diversity, distributions, and population dynamics are poorly known for most large insect groups [47,48]. These constraints on understanding broad-scale insect biodiversity patterns point to a need for a systematic global data synthesis [49]. One basic challenge for insects is the species identification bottleneck underpinning large-scale biodiversity surveys that use conventional morphological methods [50,51]. DNA barcoding and environmental DNA metabarcoding represent viable approaches for expedited, large-scale, global quantification of insect species diversity [52], despite some known limitations [53,54].

Most macrogenetic studies of animal taxa are based on mitochondrial DNA (mtDNA) sequence data, which represent the majority of available sequences [17]. Despite the limitations of using a single-locus marker [55–58], the pragmatic advantages of the ability to sample the genetic diversity of hundreds or thousands of taxa per locale potentially outweigh these considerations [59,60]. The Barcode of Life Consortium database (BOLD) is a rich source of single-locus mtDNA

that links quality-controlled genetic data with georeferenced metadata [8]. Leveraging this resource, we compiled the largest macrogenetic dataset ever assembled: over 2.3 million globally distributed and georeferenced mtDNA sequences (*cytochrome c oxidase 1 - COI*) for over 95,000 operational taxonomic units (OTUs). We use these data to generate the first global map of insect genetic variation using both the commonly used genetic diversity mean (GDM) and a new measure we introduce: genetic diversity evenness (GDE). While GDM describes the magnitude of average genetic diversity among species, GDE represents the shape of the distribution of individual GD measures for all focal taxa that co-occur in a given area. Considering both these values gives us the ability to discriminate between important processes underlying community assembly and structure [61]. As macrogenetic studies to date only describe average intraspecific genetic diversity (GDM), they are unable to determine whether high metrics of genetic diversity are due to high diversity within most community members or to the effects of a few taxa with extremely high diversity (Fig. 1; see Methods).

We focus our analyses on several questions about the macrogenetics of insects. First, we evaluate whether the magnitude and variability of GD generally follows latitudinal trends of increasing insect species richness in the tropics[62,63]. Explanations for this general latitudinal gradient are often explained in terms of the wet tropics being either “museums” or conversely “cradles”, with opposite predictions with regards to range sizes [64]. If geographic range size tends to correlate with GD [65], we might expect this gradient given the “museum” hypothesis that predicts that taxa in the tropics will be older, and have larger geographic range sizes [66,67], In contrast, species richness and GD may be decoupled due to the “cradle” hypothesis that predicts higher speciation rates, more population structure, and smaller range sizes leading to Rapoport’s Rule [68], the tendency for species’ range sizes to increase with increasing latitude [69,70]. Second, we might predict that the influence of Late Quaternary climate fluctuations to have an impact on GD through population demographic processes influenced by cyclical variation in precipitation, temperature, and glaciation patterns [71,72], where areas with more stable climatic

histories are predicted to have increased GD [72]. Finally, we consider the influence of human disturbance on patterns of assemblage-wide GD, which we expect to decrease in magnitude in areas of high human influence [42].

To answer these questions, we explore how well insect GDM and GDE are predicted by current and historical climate, habitat, and human disturbance. We first find environmental correlates of intraspecific insect genetic diversity globally using Bayesian generalized linear mixed models (GLMM), and then use these to predict patterns of insect GDM and GDE in undersampled regions, which includes most of the planet. In contrast to most global vertebrate biodiversity patterns documented to date, we find that insect GDM and GDE have bimodal latitudinal gradients that peak in mid-latitude regions and that both metrics are positively correlated with high temperature extremes.

Results

GDM and GDE were calculated from native-range insects sampled within raster grid cells at a 193 km x 193 km equal-area resolution. These cells were heterogeneously distributed across the globe on every continent except Antarctica (N = 187, Fig. 2). Regions with both high GDM and high GDE (above the 90th percentile) were found in eastern North America as well as in the North American desert southwest, in eastern Africa, and in southern China (Supplementary Fig. 1b). Areas with the lowest values of observed GDM and GDE were mostly distributed in northern North America and Europe (Supplementary Fig. 1d).

Insect genetic diversity correlates with latitude

While latitude did not significantly explain GDM or GDE across the entire planet (Table 1), it was correlated with GDE after removing cells above 60º latitude, which includes areas

covered by glaciers and tundra during the last glacial maximum (LGM; median $slope_{quadratic}$ = 0.002 [95% highest density interval (HDI): 0.001, 0.003], median R^2 = 0.103 [95% HDI: 0.022, 0.221]). In contrast to latitudinal diversity gradients for species richness in most taxa, insect GDE increased towards the poles (up to 60º latitude) and decreased towards the equator (Fig. 3). The same pattern is seen in GDM, although the pattern does not exhibit a strong statistical trend, where the 95% HDI of the predictor overlapped zero ($slope_{quadratic}$ = 2e-4 [-5e-5, 4e-4], Fig. 3).

Relationships between insect genetic diversity and the environment

Higher GDE values were mainly found in areas that rarely freeze. We divided the globe into areas above or below the global freeze-line (long-term minimum temperature of the coldest month (MTCM) above versus below 0º C) and found that GDE is significantly higher above this line than below it (Welch's unequal variance t-test; mean GDE_{above} - mean GDE_{below} = 0.042; t = -5.804, df = 184.690, $P < 0.001$), while GDM showed no significant correlative trend against this binary metric (P = 0.525).

We explored relationships among GDM, GDE, and environmental predictors within each cell using Bayesian generalized linear mixed models (GLMMs). Predictors included bioclimatic variables describing current climate, variables summarizing climate variation since the LGM ("historical climate"), a spatial habitat variation metric, a human habitat modification metric, and topographic variables. We found that GDM and GDE covary significantly with current climate and that both reach high values in the hottest regions of the planet. Notably, predictors describing human habitat modification, spatial habitat variation, and topography did not significantly predict either GD metric. After reducing the set of potential predictors to three current and historical climate variables for GDM (Supplementary Table 1) and six current and

historical climate variables for GDE (Supplementary Table 2), we were able to compare three hypotheses: H1) current climate; H2) historical climate; and H3) current + historical climate best explain the two GD metrics. We constructed models for these hypotheses using a Bayesian GLMM approach that accounts for spatial autocorrelation (SAC; see Methods) [73]. The most parsimonious models were selected based on an approximate leave-one-out cross-validation procedure (LOO; see Methods) that uses the expected log predictive density (ELPD; analogous to information criteria, i.e., Akaike's Information Criterion) as the utility function. Using this criterion, the best model for both GDM and GDE gave support to H1: current climate. The H1 models for GDM and GDE included the maximum temperature of the warmest month variable (MTWM). The best fit model (H1) for GDM additionally included precipitation seasonality, while the best fit model (H1) for GDE additionally included temperature seasonality and precipitation of the wettest month (PWM) (Fig. 4, Table 2).

For GDM, H1 (Table 2; Supplementary Fig. 2; R^2 = 0.234 [95% HDI: 0.088, 0.385]) had the most support, but was not statistically different from H3. We selected the simpler model (H1) for interpretation and prediction. Between the two current climate predictors, MTWM trends positively with GDM (slope = 0.002 [0.001, 0.003]; Fig. 4), and precipitation seasonality trends negatively (slope = -0.001 [-0.002, -1e-4]; Fig. 4).

For GDE, H1 had the most support (Table 2; Supplementary Fig. 2; R^2 = 0.327 [95% HDI: 0.152, 0.483]). MTWM was the only variable with a significant trend (Fig. 4). MTWM (slope = 0.014 [0.004, 0.025]) and temperature seasonality (slope = 0.008 [-0.004, 0.020]) both trend positively with GDE, whereas PWM trends negatively with GDE (-0.004 [-0.016, 0.008]).

Over 23% and 32% of the global observed variation in GDM and GDE, respectively, can be explained by current climate (H1). There was no residual spatial autocorrelation in the final models (Table 2). In addition, all parameter posterior distributions had less than 12% overlap with their prior distributions, indicating high identifiability (Supplementary Fig. 3).

Global predictions of insect genetic diversity

We then used the best-fit model (H1) to predict and map the global distribution of GDM and GDE individually and jointly at the original spatial scale used for modeling (193 km x 193 km resolution) across the globe, including unsampled areas (Fig. 2, Supplementary Fig. 4). We omitted predictions in all areas exposed to environmental conditions that fell outside the model training range, including Antarctica, a large portion of northern Africa, the Arabian Peninsula, parts of central Asia, and interior Greenland (Fig. 2; shown in gray; Supplementary Fig. 5).

Areas predicted to have high levels of both average genetic diversity and evenness (GDM and GDE; above the 90th percentile) include southeastern North America and southern India, while the regions with the very highest predicted values include southern and southwestern Australia, parts of the desert southwest of North America, and southern India (Supplementary Fig. 1a; Fig. 2). Areas predicted to have the lowest GDM and GDE values (below the 10th percentile for both) were found in Patagonia and the northern Nearctic and Palearctic (Supplementary Fig. 1c). When considered independently, GDM is predicted to be highest (above the 90th percentile) in the temperate forest regions of eastern North America (Supplementary Fig. 1e; Fig. 2), yet is predicted to be generally low for much of the Neotropics (Fig. 2). The very lowest GDM areas (below the 10th percentile) are distributed across the nearctic and palearctic tundra, the entire Andes mountains chain, and areas in the Himalayan mountains (Supplementary Fig. 1g). When GDE is considered independently, it is predicted to be the highest (above the 90th percentile) throughout Australia, the southeast and desert southwest of North America, as well as much of the Indian subcontinent and the outer fringes of Saharan Africa (Supplementary Fig. 1i). On the other hand, GDE is predicted to be lowest (below the 10th percentile) in northern Europe and parts of the nearctic and palearctic tundra as well as southern Patagonia in South America (Supplementary Fig. 1k; Fig. 2).

Taxon-specific patterns of GD

Six insect orders contain 97.2% of all OTUs in this study (Supplementary Fig. 6**;** Supplementary Table 3). In order of prevalence, they include Diptera, Lepidoptera, Hymenoptera, Coleoptera (the four mega-diverse orders that include *ca*. 80% of known insect species), Hemiptera, and Trichoptera. The remaining 2.8% of OTUs belong to 20 additional insect orders.

To investigate the influence of the three most prevalent orders (Diptera, Lepidoptera, and Hymenoptera, 84.4% of total) we removed these orders from the full dataset and reanalyzed patterns of GDE and GDM. Using Welch's unequal variance t-tests, we found no difference in GDE estimates between the full and reduced datasets (Supplementary Fig. 7**;** P = 0.335). However, GDM was slightly but significantly lower in the full dataset ($mean_{diff}$ = -0.003, df = 91.195, P = 0.002).

OTU sampling across the most abundant three orders varied geographically (Supplementary Fig. 8). We calculated OTU sampling as the number of OTUs per order within each cell. Diptera dominated OTU sampling towards the poles, while Lepidoptera dominated sampling towards the tropics and in some temperate locations, and Hymenoptera typically accounted for fewer than 50% of OTUs sampled, with overrepresented sampling in Madagascar (Supplementary Fig. 8).

Discussion

We found clear, and in some cases surprising, global biogeographic patterns of insect genetic diversity (Fig. 2; Supplementary Fig. 1). There is a reversed latitudinal gradient for GDM and GDE, with both significantly lower in the tropics than in temperate and subtropical regions

on either side of the equator. GDM and GDE have bimodal distributions which peak in areas that were unglaciated during the last glacial maximum (LGM, < 60º latitude, 18-21,000 years ago, [74] and are generally lower in the wet tropics and in temperate areas that were glaciated or tundra-like during the LGM (> 60º latitude). This suggests that the forces underlying intraspecific genetic diversity are inherently different from those driving the classical negative latitudinal gradients in species richness and phylogenetic diversity found in major arthropod groups such as ants and spiders [75–77], as well as plants [78], which are expected to be strongly linked to insect biogeographic patterns. Bees (order Hymenoptera) are one notable exception, showing a bimodal latitudinal gradient similar to what we find, with highest richness at mid-latitudes [79]. While neutral theories of biodiversity predict positive species genetic diversity correlations (SGDC) [31,80,81], there are also conditions for which one would expect weak or even negative correlations, e.g. neutral conditions paired with high mutation rates [82] or greater niche breadths result in higher genetic diversity, but fewer species in a community [83]. However, many confounding factors will affect how species diversity metrics relate to GDM and GDE, and these factors may have both positive and negative effects, leading to large variation in the sign and intensity of SGDCs [84], especially at a global scale in such a broad taxonomic group such as insects.

The bimodal latitudinal gradient we find also contrasts with recent macrogenetic studies of vertebrates, all of which find a negative latitudinal gradient of genetic diversity with average values peaking in the tropics and declining poleward, including mammals [20], amphibians [22], and fishes [21]. Our finding of lower genetic diversity in areas that were glaciated or tundra during the LGM is consistent with a gradient of lower haplotype richness in recently unglaciated areas found in European butterflies [30] based on the same *COI* data from BOLD used here. Similarly, aquatic insect species have lower intraspecific genetic diversities in recently unglaciated areas of Europe compared to Neotropical areas [32].

Why would GDM and GDE be lower in areas like the wet tropics when the species diversities of most insect groups reach their peaks in these habitats [87,88]? Rapoport's Rule, the tendency for species' range sizes to increase with increasing latitude [69,70], might explain this result because species with larger ranges tend to harbor greater genetic diversity [65,89]. Macrogenetic studies tend to calculate intraspecific genetic diversity at the grid-cell level, and extratropical cells are likely dominated by species with large ranges, while tropical cells are likely dominated by species with smaller ranges. Thus, coalescent times among sampled alleles within each species in each extratropical cell will tend to be older, yielding larger average pairwise distances, the metric we use for GD [90].

A few additional mechanisms may play a role in explaining higher insect extratropical GDM. Wide-ranging extratropical species can usually tolerate a broader range of climatic variation, whereas limited-range tropical species tend to have a narrow climatic niche, stronger habitat specializations, and narrower physiological tolerances ([91]). While peaks of GDM and GDE may be driven by larger range sizes and greater physiological tolerances of species in hot, seasonal areas, other studies have found that larger range sizes in temperate species can lead to greater population genetic structure [27]. This could in turn lead to lower levels of local GDM if the spatial resolution of sampling, *i.e.,* grid size, is smaller than the population range sizes of locally occurring species [92]. However, the coarse resolution of the cells we use (37,249 km^2) likely avoids this issue.

It is also possible that insect GDM and GDE patterns reflect their unique life history traits and responses to short-term environmental shifts. For instance, observed and predicted patterns of GDM and GDE (Fig. 2) are higher in areas where insect diapause, the temporary suspension of development during the life cycle, is more prevalent ([91]). Insect diapause is thought to provide adaptive tolerance to wider abiotic conditions and may result in larger and more uniform ranges [91,93]. Given the positive relationship between range size and GD [65,89], this provides a possible mechanistic relationship that connects Rapoport's Rule and the higher and

more uniform genetic diversities found in higher latitude regions with pronounced seasonality. Another important feature of GDE is that it is lower in areas that experience frequent freezing, *i.e.*, below the freeze-line, which may be related to the range of temperatures encountered by a species. Although this might seem counterintuitive, insect species that enter diapause in habitats that seasonally accumulate substantial snow are likely to encounter less extreme temperatures than those in more exposed habitats of temperate deserts, where large temperature oscillations are common [94]. In this light, we might predict more uniform and higher levels of GD in Australia, which both is above the freeze-line and has high temperature seasonality with little snow accumulation.

Since species diversity patterns of specialist insect herbivores correlate with their host plants' [95], some of the environmental correlates associated with high GDM and GDE in insects may be more directly tied to the climatic determinants underlying global plant diversity patterns [96,97]. Nearly half of all insect species are herbivorous, yet this varies across orders ranging from ~99% of Lepidoptera, 30-35% of Diptera and Coleoptera, yet only 10-15% of Hymenoptera [98]. Insects and angiosperms have species richness patterns with similar poleward gradients [78,99] and likely evolved in the tropics with subsequent adaptations for wider environmental tolerances associated with temperate radiations [100]. Although the diversities of insects and plants are tightly correlated [101], we found no significant correlations between habitat heterogeneity, a derived measure of variability in remotely sensed metrics of vegetation diversity, and GDM or GDE, suggesting that specific aspects of plant community composition, rather than just plant species richness, affects insect genetic diversity. Further, host plant intraspecific genetic diversity demonstrably influences herbivorous insect community assembly [102,103]. A potentially fruitful area of future investigation would be to search for links between the genetic diversity of local insect and plant assemblages.

While Rapoport's Rule may explain increasing extratropical genetic diversity, the observed bimodal latitudinal gradient emerges as genetic diversity begins to decline poleward in

the temperate regions. We hypothesize that the poleward spatial range expansions after the LGM resulted in founder effects that led to the low levels of genetic diversity we detect in these regions. This is consistent with many studies showing lower intraspecific genetic diversity in organisms inhabiting regions that were previously tundra, cold steppes, or glaciated [104], presumably due to post-glacial expansions [71].

The regions predicted to have high GDE and GDM correspond with known hotspots of insect biodiversity. For instance, North American southwestern deserts have the highest butterfly phylogenetic endemism in North America [79,105]. Southwestern Australian deserts also have exceptionally high arthropod endemism [106], and are among the original biodiversity hotspots identified by [107]. The high GDM observed in eastern North America also corresponds with high Odonata species diversity [108].

The bimodal latitudinal gradient in GDE is likely influenced by mechanisms at least partially independent of those that generate the same gradient in GDM. Higher levels of GDE, reflecting lower variability among genetic diversities, could partially result from histories of neutral community assembly processes [61]. Overcast et al. (2020) found in both simulated communities and empirical arthropod, annelid, and tree datasets that ecologically neutral communities tend to have higher GDE than non-neutral, *i.e.,* niche-structured communities. Here, the lower GDE in communities assembled via environmental filtering is likely caused by increased genetic diversity in species with stronger local ecological adaptation. While this suggests that equatorial insect communities may have stronger local niche-structured mechanisms than temperate insect communities, consistent with the idea of stronger niche conservatism in the tropics [109], this is one of many hypotheses emerging from our study.

While GDM and GDE can be informative about processes underlying biodiversity patterns, interpretation of GDM and GDE in isolation (or even together) without additional information about the study system can lead to erroneous conclusions. For example, if the rare and the least genetically diverse species go locally extinct, this could raise GDM for the

remaining species. Similarly, a disturbed community with high GDE may be composed of populations with low but similar genetic diversity, such that without also measuring GDM, the low genetic diversity of the community may go unnoticed. Conversely, relatively high GDM in a community may be driven by a few hyper-dominant taxa with high GD while a remaining majority of low GD taxa. Alternatively, this same scenario of high GDM and low GDE could be found in areas that have a mix of wide-ranging endemic taxa and several invasive species with low genetic diversity [110,111]. Low GDE may highlight this driver when considered in context with *a priori* hypotheses derived from other evidence. We recommend future macrogenetic studies at regional scales to include metrics of the average and shape of GD distributions in addition to ancillary information to effectively summarize and interpret the genetic diversity of assemblages.

Genetic diversity is critical to the survival of both the insects themselves, and the complex networks of interactions to which many insects belong. High genetic diversity may facilitate adaptation to changing climates and emerging diseases, two (of many) potential drivers of the “insect apocalypse” [44]. In addition, genetic diversity contributes to the diversity and stability of species interaction networks by affecting niche space and competition [112], community structure [113], and network complexity [114]. At larger ecological scales, insect genetic diversity may reflect ecosystem function and structure as reliably as other traditional macroecological metrics such as species richness [115]. It can augment the resilience of ecosystems that provide continuing services for humankind [14], such as disease management, curbing the spread of invasive plants, aiding sustainable agriculture, pollinating food crops, and controlling pests [13]. While the metric of global human modification we considered did not significantly correlate with GDM or GDE, there are many facets of anthropogenic disturbance acting at different spatial scales that are difficult to summarize in a single metric [116]. The spatiotemporal resolution of genetic sampling currently available does not permit rigorous assessment of how humanity affects insect GD at a global scale, but a concerted increase in sampling effort, especially in the data-poor regions we identify, will likely make this feasible in the not-to-distant future.

By modeling relationships between environmental data and our two complementary measures of intraspecific genetic diversity, GDE and GDM, we also make predictions about assemblage-level genetic diversity in data-poor regions of the planet. This has the potential to fill a knowledge gap that far exceeds the undersampling and taxonomic uncertainties underlying vertebrate and plant macroecological studies [117,118]. We provide targets for future efforts that will fulfill global commitments to monitor and conserve genetic diversity, a biodiversity component that has rarely been assessed or used to guide conservation decisions [15], while focusing attention on a data deficient group with evidence of global population declines [119]. While there have been recent arguments that the value of putatively neutral intraspecific genetic diversity is overstated in the context of conservation [120], a large body of literature indicates otherwise [121], especially if neutral genetic diversity correlates with adaptive potential [122,123]. Taken together, GDM and GDE are fundamental biodiversity metrics for documenting and understanding how "the little things that run the world" can change, persist, and potentially adapt in the face of global change [124].

Methods

Aligning and filtering sequence data

We downloaded *cytochrome c oxidase 1* (*COI*) mitochondrial sequence data for insects directly from the BOLD webpage using the application programming interface (API) (http://www.boldsystems.org/index.php/resources/api; downloaded 19 Nov 2019). Our initial

database comprised 3,301,025 complete insect records before applying a series of quality filters. We used the BOLD database's OTU assignments (termed barcode identification numbers; BINs), which cluster similar sequences algorithmically and map them against the BOLD database [125]. After trimming end gaps from sequences, we removed exceptionally long sequences (>800 base pairs, bp) which contained a large proportion of gaps that negatively impacted alignments and the calculation of summary statistics. In addition, we removed shorter sequences (<400 bp) that the BOLD database uses for BIN identification, but which may downwardly bias GD estimates. We also only used COI sequences when georeferenced metadata with geographic coordinates were available. Sequence alignments were independently performed for each OTU within single sampled geographic raster cells, *i.e.*, grid cells. We used default settings in Clustal Omega (v1.2.3) to align the sequences and visually assessed a subset of alignments to check for alignment errors [126].

To reduce the potential impact of invasive species on our analyses, we removed trans-continental invasive species from the dataset using a list of invasive insect species compiled from seven resources: Global Insect Species Database, [http://www.issg.org/database; accessed 23 May, 2020]; Invasive Species Compendium [https://www.cabi.org/isc/; accessed 24 May, 2020]; Center for Invasive Species and Ecosystem Health [https://www.invasive.org/; accessed 24 May, 2020]; Invasive Alien species in South-Southeast Asia [127]; Japan Ministry of the Environment [https://www.env.go.jp/en/nature/as.html; accessed 24 May, 2020]; European Alien Species Information Network [https://easin.jrc.ec.europa.eu/easin/Home; accessed 24 May, 2020]. We identified all species and OTUs present on multiple continents and removed those on the invasive species list from our dataset. While some invasive species may be restricted to single continents, removal of such taxa was not possible given the lack of information on changes in insect range boundaries and species assignments.

Calculating the evenness and mean of genetic diversity (GDE and GDM)

Global macrogenetic studies have so far focused on spatially defined metrics that summarize the genetic diversities calculated across all species sampled from an area of arbitrary spatial resolution [19,128]. This is most commonly the average genetic diversity or, alternatively, a measure of the allelic richness derived from the total number of unique and/or common alleles of a genetic locus across all taxa within an area [129]. We used two distinct summaries of genetic diversity - the mean genetic diversity (GDM) and the evenness of this GD per unit of area (GDE). To obtain the GD for each OTU per grid cell, we calculated nucleotide diversity as the average number of nucleotide differences across all pairwise sequence comparisons per OTU per base pair [90,130]. Aggregated across OTUs within each grid cell, GDM is then defined as the average GD among OTUs in each grid cell, following [20]. Because the distribution of GDM at the grid cell scale was highly skewed towards zero, we performed a square-root transformation to achieve a more normal distribution. All subsequent statistical analyses of GDM at the grid cell scale were based on the transformed GDM.

While GDM is a standard metric in the macrogenetic toolbox, GDE is derived from a set of metrics known as Hill numbers that permit direct comparisons of diversity across scales and data types [131–133]. GDE is then defined as the first-order Hill number of GD across OTUs per grid cell, corrected by sampled OTU richness [61]:

$$\frac{exp(\sum_{i=0}^{N} - \pi_i \ln(\pi_i))}{N}$$

Where N is the number of OTUs in the assemblage and π_i is the GD for a single OTU. Correcting for sampled OTU richness allows for comparison across assemblages of different numbers of OTUs. The numerator of this metric is the exponential of Shannon's diversity index,

which is also referred to as Shannon's information measure or Shannon's entropy in the literature [134]. It is commonly used to describe evenness and variability of species abundances [135,136], and here we adapt it to do the same for genetic diversities calculated from all species sampled from a particular area.

A higher GDE indicates where most OTUs have a similar GD (Fig. 1), whereas a lower GDE arises when GD values across the community diverge considerably [135]. Low GDE can take a variety of shapes, but the most common shape for low GDE cells in our observed data is markedly L-shaped **(**Fig. 1**)**.

Spatial resolution and sampling decisions

To assess how the spatial scale and density of OTU sampling impacted our results and to establish a sampling strategy that maximizes the amount of information, we calculated both metrics at 1) three different spatial resolutions, and 2) three thresholds of minimum OTU sample sizes per grid cell. The spatial resolutions include 96.5 km x 96.5 km, 193 km x 193 km, and 385.9 km x 385.9 km equal-area grid cells using a Behrmann cylindrical equal-area projection, which are 1°, 2°, and 4° longitude at 30°N. At each of three spatial resolutions, we considered a minimum of 100, 150, or 200 unique OTUs per grid cell, as these approach the lower bounds of the sample size needed to effectively estimate the diversity of a community using Hill numbers [137]. We then selected the spatial resolution and minimum number of OTUs per grid cell that maximized the average number of OTUs per grid cell, the number of grid cells, and the average number of taxonomic orders per grid cell, while minimizing variation in the number of OTUs across grid cells. With respect to numbers of sampled allele copies per OTU, we used a minimum of three individuals per OTU per grid cell. This is a conservative approach to estimate GD while still maximizing data use given that many BOLD data submissions omit duplicate alleles and that coalescent theory suggests that using average pairwise distance from 5-10

451 samples per OTU provides estimates of genetic diversity that are as reliable as those obtained
452 from hundreds of samples [90].

453 We found that a grid cell resolution of 193 km x 193 km with a minimum of 150 OTUs
454 per grid cell minimized variation in the number of OTUs across grid cells, while maximizing the
455 number of grid cells, the average number of OTUs per grid cell, and the average number of
456 taxonomic orders per grid cell (Supplementary Fig. 9). This filtering criteria led to a final dataset
457 that included 2,362,636 COI sequences from 95,540 OTUs sampled across 187 globally
458 distributed grid cells. On average, each cell contained ten insect orders, 460 OTUs, and 4,013
459 individuals. We calculated variation in the number of OTUs per cell as the difference between
460 the upper and lower 90% highest density interval (HDI) of OTUs across cells yielding a
461 difference of 1779 OTUs.

462 Because 97.2% of OTUs are represented by six taxonomic orders (Supplementary Fig.
463 6; Supplementary Table 3), with ~85% represented by three (Diptera, Lepidoptera, and
464 Hymenoptera), we investigated whether and to what degree over-represented orders might be
465 driving the signal of GDE and GDM. To examine whether any of these three most dominant
466 orders deviate from global patterns of genetic diversity, we compared the global frequency
467 distributions of per-cell GDM and GDE from these three orders combined with the distribution of
468 these summary statistics for the entire data set. The distributions of per-cell GDE and GDM
469 between these filtered data sets and the original data set were compared using Welch's unequal
470 variance t-tests [138].

471 Although coalescent theory predicts that the number of allele copies per OTU per grid
472 cell will have a limited impact on the per OTU genetic diversity [90], we examined whether this
473 assumption was met in the data by testing for Pearson's correlations between the per OTU GD
474 and number of individuals per OTU (r = 0.030, P < 0.001). Similarly, to investigate whether per
475 grid cell sampling, *i.e*., total number of individuals, number of individuals per OTU, and number
476 of OTUs per cell, had an effect on GDE or GDM, we tested for Pearson's correlations between

these quantities (all $P > 0.40$, Supplementary Table 4). In addition, we assessed sampling variation by taking the ten most sampled grid cells (2,748 to 13,300 OTUs per grid cell) and obtaining sampling distributions of GDM and GDE for each by resampling with replacement 150 OTUs per sample (N = 1000 resamples) and calculating the summary statistics for each resample (Supplementary Fig. 10).

Environmental variable selection

We aggregated a total of 47 abiotic, biotic, and anthropogenic variables that potentially influence intraspecific genetic diversity in insect communities (Supplementary Table 5). We removed highly correlated variables ($r > 0.75$), prioritizing variables that represent climate extremes, climate variability, habitat variability, last glacial maximum (LGM) climate stability, and human influence on the environment.

We retained a final data set of 11 ecologically relevant variables: five bioclimatic variables, habitat heterogeneity, global human modification, and four metrics of climate stability (temperature and precipitation) since the LGM (Supplementary Table 5). The five bioclimatic variables describe climate extremes and variability, and were obtained from the CHELSA database [139]. They include maximum temperature of the warmest month (MTWM), minimum temperature of the coldest month (MTCM), precipitation of the wettest month (PWM), precipitation of the driest month (PDM), temperature seasonality, and precipitation seasonality [139,140]. The habitat heterogeneity metric was calculated as the standard deviation of the Enhanced Vegetation Index, which was derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) (2.5 arc-min; [141]). The human modification variable is a cumulative measure of human modification to terrestrial areas [142]. Measures of both the historical trend and variability of temperature and precipitation over the last 21,000 years were obtained from [20]. The specific definitions of these derived metrics include “deep-time climate trend”, the change in

climate within each century, averaged across centuries, and “deep-time climate variability”, meaning the standard deviation around the change in climate, averaged across centuries. Low deep time trend values indicate regions with long-term climate stability, while low variability values indicate regions with short-term climate stability. Each variable was aggregated from its original resolution (see Supplementary Table 5) to 193 km by 193 km resolution bilinear interpolation.

In addition, we explored the relationship between GDE and GDM and a binary variable delineating the globe along the freeze-line. Here, areas with a MTCM above 0ºC are considered above the global freeze-line, while areas with a MTCM below 0ºC are considered below this line. These regions have been found to correspond with sharp community turnover in birds [143] and could correlate with critical life processes for insects.

Modeling approach

To identify the models that best explain the global distribution of GDM and GDE in insects, we applied multimodel inference using Bayesian approaches. We modeled a set of variables underlying environment, latitude, and the global freeze-line on the GDM and GDE of insect assemblages independently. The necessary complexity of the modeling procedure outlined below precluded the construction of a model including the combined effects of all variables.

We prioritized constructing a simple, interpretable linear model that predicts GDM and GDE robustly across the globe by conducting model selection in two steps. First, we reduced the number of potential predictor combinations from the set of 11 variables with low collinearity using Bayesian regression coupled with projective prediction feature selection. This approach minimizes the number of predictor variables in a simple model while retaining comparable predictive power to a model that includes the full suite of predictors [144]; [145]. For each model we

used weakly informative priors on all slope parameters (N(0, 0.1)) and the error term (N(0, 1)). We centered and scaled all predictors to standard deviation of 1 and mean of 0 prior to modeling. We aimed to retain single candidate models of GDM and GDE.

If residual spatial autocorrelation (SAC) is present, the assumption of independent and identically distributed residuals would be violated, resulting in potentially biased overprecision of parameter estimates [146]. We tested for SAC in the residuals of the resulting simplified models using Moran's *I* and 10,000 simulations implemented in the R package *spdep* v1.1-2 [147]. We detected significant levels of SAC in the residuals of our GDE model (Moran's $I = 0.149$, $P = 0.008$) and our GDM model (Moran's $I = 0.306$, $P < 0.001$).

Given this presence of SAC, we used a Bayesian generalized linear mixed-effects model (GLMM) implemented in the R package *glmmfields* v 0.1.4 to robustly model GDE and GDM while accounting for SAC [73]. SAC is modeled as a random effect with a multivariate t-distribution determining the shape of the covariance matrix. Model parameters were estimated from the posterior distribution using a No U-Turn Sampler [148,149]. We again tested for SAC in the residuals of these models using the same approach as above.

Since the covariance structure among predictors is modified by the spatial random effects, we performed additional model selection with the simplified candidate models. The simplified models allowed for the formulation of specific hypotheses on the relationship between GD and the environment. We compared three hypotheses (see Results) using the approximate LOO cross-validation procedure outlined above. After selecting a model, we used the percentage of prior-posterior overlap to assess the identifiability of parameter estimates relative to the information provided by their prior distributions [150]. Low overlap between the prior and posterior distribution of a parameter indicates that there is sufficient information in the data to overcome the influence of the prior.

We used a similar approach as above to test the effect of latitude on GDM and GDE. We constructed Bayesian GLMMs with latitude as a linear predictor and a quadratic term for GDM

551 and GDE. All priors were the same as those in the climate modeling procedure. We compared
552 the fit of the linear model, quadratic model, and an intercept model using approximate LOO
553 cross-validation. We used prior-posterior overlap to assess parameter identifiability.
554 We also independently tested the effect of the global freeze-line on GD using Welch's
555 unequal variance t-tests.

556 Global genetic diversity map generation

557 Using the final models of GDM and GDE, we created maps of the global distribution of
558 insect GD. We used 1000 draws from the posterior distribution to predict terrestrial
559 environments across the globe. We included all continents except Antarctica, which had no
560 observed data and included environments far more extreme than the observed data. We
561 created maps of the median predicted GDM, upper, lower, and range of the 95% HDI, as well as
562 for GDE. In addition, we created bivariate color maps of these prediction intervals for combined
563 GDM/GDE to highlight areas where GDM and GDE vary in similar and different directions.
564 Multivariate environmental similarity surface (MESS) maps were created to visualize
565 how environmentally similar or different areas across the globe are compared to the model
566 training data [151]. These maps aid in identifying areas of high extrapolation and thus where
567 uncertainty for predictions is also high. Decreasing negative MESS values represent
568 increasingly non-analogous environments, and increasing positive values indicate increasing
569 similarity. We used the MESS results to mask areas with non-analogous environmental space
570 (values less than 0) on our global prediction maps, indicating areas with high prediction
571 uncertainty.

Data availability

All geographic, environmental, and genetic data are available at **dryad_link (will make available upon acceptance)**.

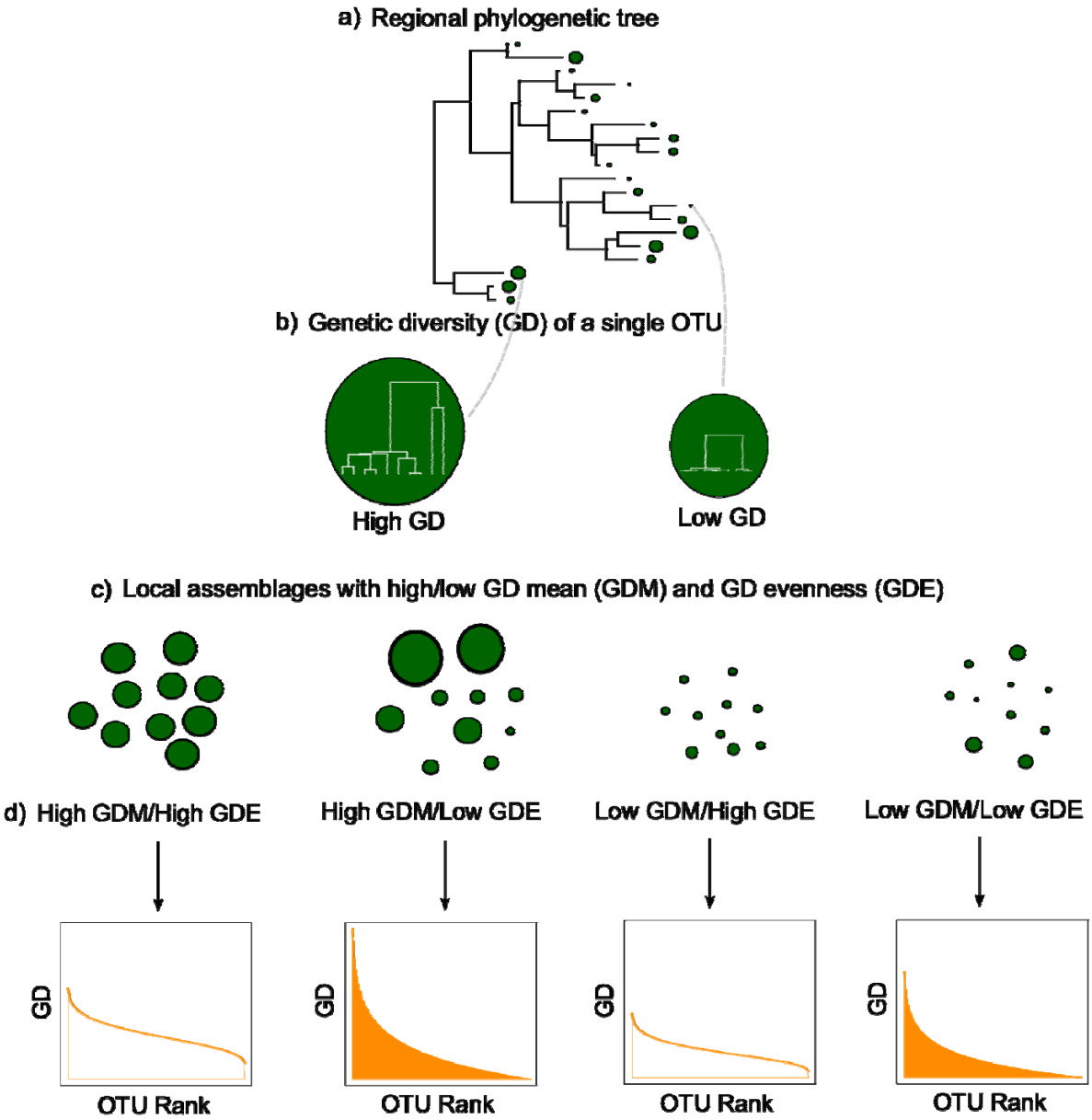
Computer code

All code used for data processing and analysis is available at https://github.com/connor-french/global-insect-macrogenetics.

Figure captions

Fig. 1

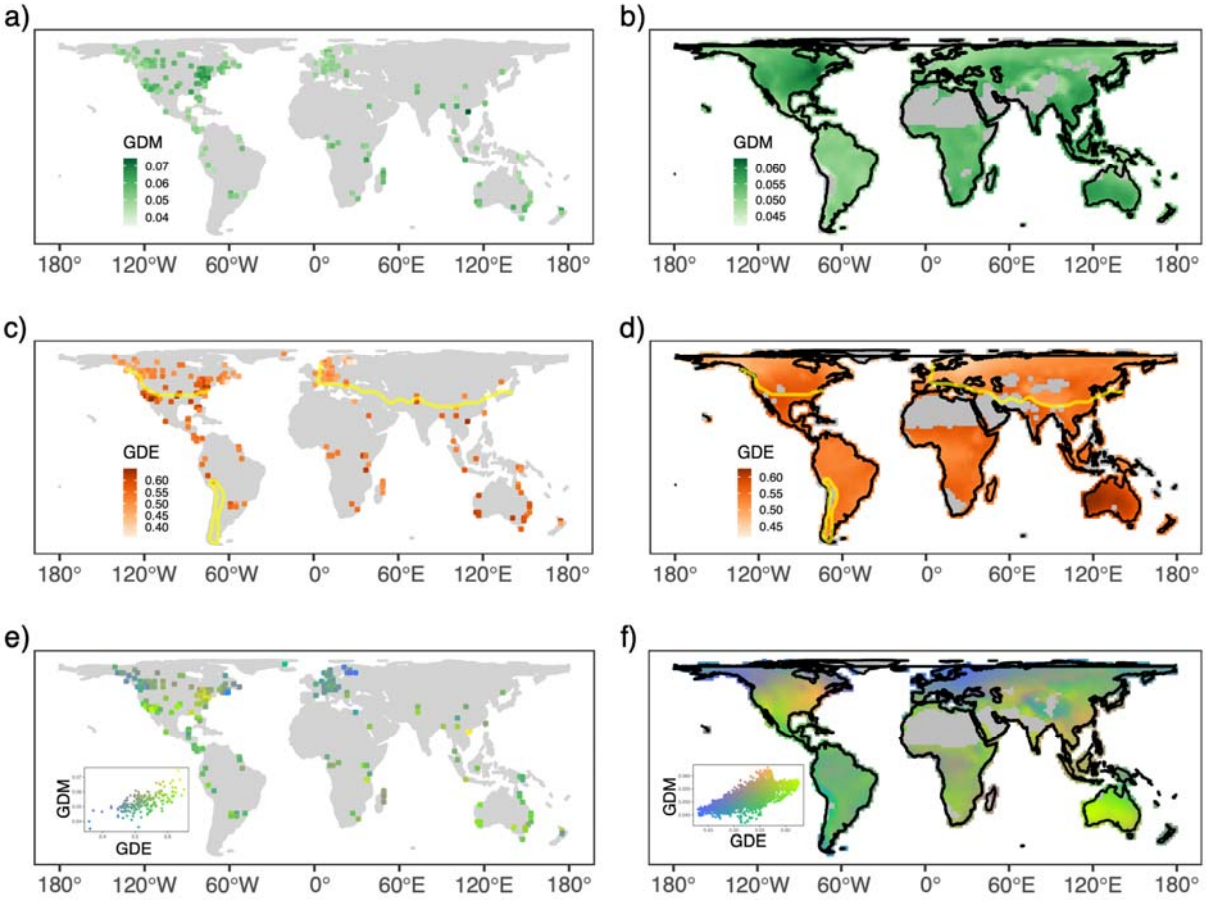
Diagram illustrating genetic diversity mean (GDM) and genetic diversity evenness (GDE). A local assemblage (c) is a set of operational taxonomic units (OTUs, analogous to species) sampled from a single grid cell that are a subset of a wider regional pool, whose evolutionary relationships are shown in (a). OTUs have varying amounts of genetic diversity (GD), represented by green circles with sizes corresponding to magnitude of GD. Longer branches among individuals within an OTU indicate a longer time to coalescence and therefore higher GD (b). Panel (c) illustrates four local assemblages sampled from four different grid cells from the same regional pool. The first local assemblage in (c) has high GDM and high GDE, represented by OTUs with high and similar GD and a corresponding relatively flat curve on the rank plot in (d). The second local assemblage in (c) has the same high GDM as the first assemblage in (c), but has lower GDE, indicated by dissimilar circle sizes and a steeper curve in the corresponding rank plot in (d). The third and fourth local assemblages in (c) have the same GDE as the first and second assemblages respectively, but have lower GDM, indicated by the smaller circle sizes and lower height curves on the rank plots in (d). This illustrates the complementary nature of the two metrics, where GDM describes the average magnitude of GD in a local assemblage, while GDE describes the distribution of GD in that same local assemblage.



596
597

598 Fig. 2

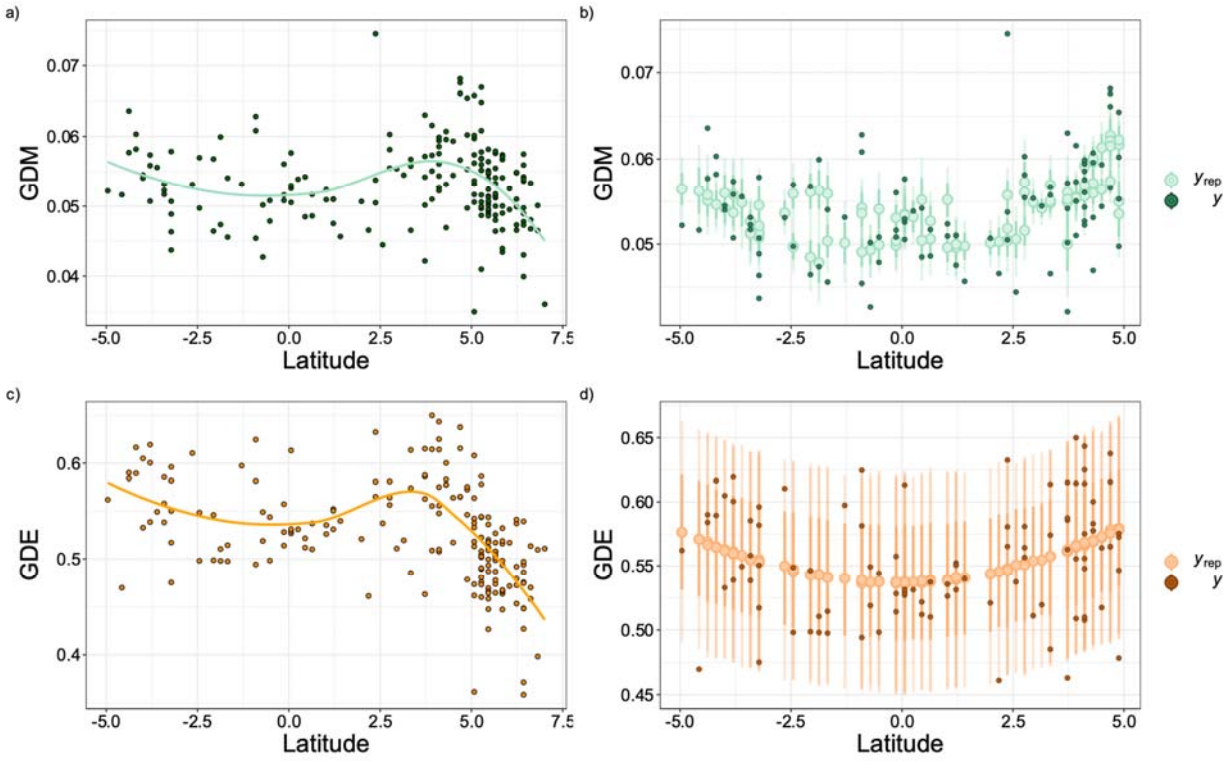
599 The observed (a, c, e) and projected (b, d, e) distributions of GDM (a, b), GDE (c, d), and their
600 composite (E, F) across the globe. Values for the projected maps were derived from a Bayesian
601 GLMM model with environmental predictors. For GDM (b), the best fit model included MTWM
602 and precipitation seasonality, while for GDE (d), the best fit model included MTWM, temperature
603 seasonality, and PWM. The yellow lines drawn across the maps of GDE (c, d) indicate the
604 global freeze-line, where areas north of the line and inside the polygon in South America have
605 minimum temperatures that dip below 0°C (below the global freeze-line), and areas south of the
606 line and outside the polygon have minimum temperatures that remain above 0° C year-around
607 (above the global freeze-line). Areas above the freeze-line on average have higher GDE than
608 those below the global freeze-line. We masked in gray areas with environments non-analogous
609 to the environments used for modeling. MTWM = maximum temperature of the warmest month;
610 PWM = precipitation of the wettest month.



611

612 Fig. 3

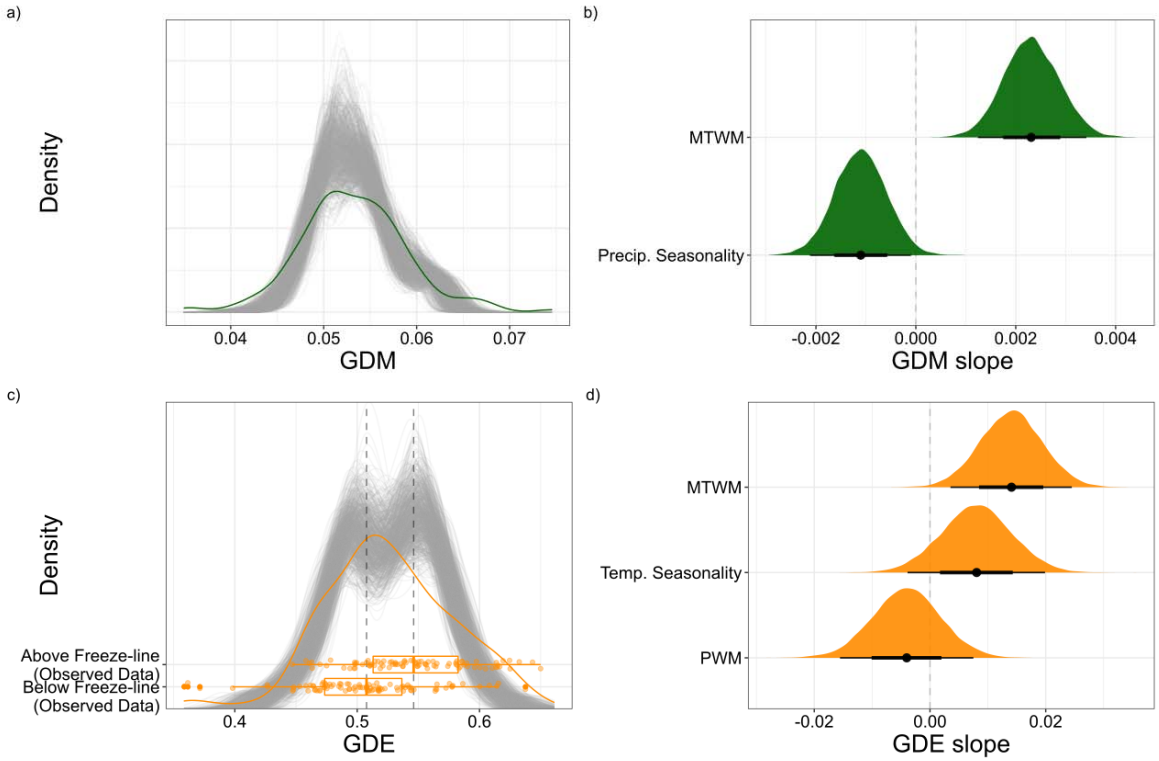
613 GDM and GDE increase towards the poles until 60º latitude, the southernmost extent of
614 glaciated and tundra sites. When removing areas that were glaciated or tundra during the LGM
615 (N = 100), GDE (c) showed a significant positive quadratic trend with latitude, while GDM
616 showed a qualitative positive quadratic trend (d) that was not statistically robust. The observed
617 data are plotted in dark green and orange, while the median, 70% HDI, and 95% HDI of
618 posterior predictions from Bayesian GLMM models are plotted in light green and orange. A
619 bimodal distribution of GDM (b) and GDE (d) is evident in the full observed data set (N = 187). A
620 Loess-smoothed trendline is overlaid to highlight this trend.



621

622 Fig. 4

623 Distributions of observed and predicted GDE and GDM. The gray lines in (a) and (c) are 1000
624 random samples from the posterior distribution of the GDM and GDE models. The green lines
625 are the observed distributions of GDM and GDE. The boxplot overlaid on (c) illustrates the
626 higher observed GDE above the global freeze-line (minimum temperature > 0° C) versus GDE
627 below the global freeze-line (minimum temperature <= 0°). The observed differences in GDE
628 above and below the global freeze-line are reflected in the posterior draws, which we highlight
629 with two gray, dashed lines drawn through the medians of the observed data. The posterior
630 distributions of the slopes for each predictor are shown in (b) and (d). The thin bars under each
631 density plot indicate the 95% HDI and the thick bars indicate the 70% HDI. MTWM = maximum
632 temperature of the warmest month; PWM = precipitation of the wettest month



633

Tables

Table 1

Results for models of GDM, GDE, and latitude, either across the entire data set or only including areas that were not glaciated or tundra during the LGM (indicated by model names with a subscript 60).

Model	Term	Median slope	Lower 95% HDI	Upper 95% HDI	Median R^2	Lower 95% HDI	Upper 95% HDI
GDM ~ latitude + latitude^2	linear	-0.001	-0.002	0.0001	0.186	0.063	0.327
GDM ~ latitude + latitude^2	quadratic	-0.0001	-0.0002	0.0001	0.186	0.063	0.327
GDE ~ latitude + latitude^2	linear	-0.006	-0.013	0.001	0.261	0.102	0.417
GDE ~ latitude + latitude^2	quadratic	-0.001	-0.002	0.001	0.261	0.102	0.417
GDM_{60} ~ latitude + latitude^2	linear	-0.0001	-0.001	0.0007	0.231	0.083	0.382
GDM_{60} ~ latitude + latitude^2	quadratic	0.0002	-0.0001	-0.0004	0.231	0.083	0.382
GDE_{60} ~ latitude + latitude^2	**linear**	**0.0004**	**-0.0024**	**0.0031**	**0.103**	**0.022**	**0.221**
GDE_{60} ~ latitude + latitude^2	**quadratic**	**0.002**	**0.001**	**0.003**	**0.103**	**0.022**	**0.221**

The bolded model has at least one predictor with a 95% highest density interval (HDI) that does not overlap with zero, indicating a significant statistical association.

642 Table 2

643 Results of the spatial linear modeling of environmental correlates for GDM and GDE.

Model	Predictors	Diff. in ELPD	SE in diff.	ELPD	SE ELPD	Median R^2	Lower 95% HDI	Upper 95% HDI	Moran's *I*	*P*-value (Moran's
GDM ~ Current climate	MTWM, precip. seasonality	0	0	746.215	16.704	0.234	0.088	0.385	0.040	0.2
GDM ~ Current climate + historical climate	MTWM, precip. seasonality, precip. trend	-0.711	1.056	745.505	16.683	0.250	0.100	0.398	0.047	0.1
GDM ~ Historical climate	Precip. trend	-7.149	5.368	739.066	16.646	0.202	0.072	0.342	0.045	0.2
GDE ~ Current climate	MTWM, PWM, temp. seasonality	0	0	346.465	11.935	0.327	0.152	0.483	-0.018	0.5
GDE ~ Current climate + historical climate	MTWM, PWM, temp. seasonality, temp. trend, temp. variability, precip. trend	-3.753	1.805	342.713	12.227	0.387	0.195	0.535	0.002	0.4
Historical climate	Temp. trend, temp. variability, precip. trend	-5.099	4.359	341.366	12.076	0.291	0.119	0.454	0.018	0.3

644

645 Each hypothesis we tested (current climate, historical climate, and current climate + historical

646 climate) had a unique set of predictors for GDM and GDE. Columns 3-6 contain summary

647 statistics (expected log predictive density, ELPD) from approximate leave-one-out cross-

648 validation model selection for spatial Bayesian GLMMs. The “diff. in ELPD” column indicates

649 differences in ELPD from the best fit model. In addition, columns 7-9 contain a summary of the

650 Bayesian R^2 model fit statistic. Residual spatial autocorrelation for each model was calculated

651 using Moran’s *I* and 10,000 simulations were used to calculate a *P*-value. HDI = highest density

652 interval; MTWM = maximum temperature of the warmest month; PWM = precipitation of the

653 wettest month.

654

References

1. Jenkins, C. N., Pimm, S. L. & Joppa, L. N. Global patterns of terrestrial vertebrate diversity and conservation. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E2602–10 (2013).
2. Jarzyna, M. A., Quintero, I. & Jetz, W. Global functional and phylogenetic structure of avian assemblages across elevation and latitude. *Ecol. Lett.* **24**, 196–207 (2021).
3. Faith, D. P. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10 (1992).
4. Howard, C., Stephens, P. A., Pearce-Higgins, J. W., Gregory, R. D. & Willis, S. G. The drivers of avian abundance: patterns in the relative importance of climate and land use. *Glob. Ecol. Biogeogr.* **24**, 1249–1260 (2015).
5. Callaghan, C. T., Nakagawa, S. & Cornwell, W. K. Global abundance estimates for 9,700 bird species. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
6. Cardoso, P., Pekár, S., Jocqué, R. & Coddington, J. A. Global patterns of guild composition and functional diversity of spiders. *PLoS One* **6**, e21710 (2011).
7. Safi, K. *et al.* Understanding global patterns of mammalian functional and phylogenetic diversity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 2536–2544 (2011).
8. Ratnasingham, S. & Hebert, P. D. N. bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Mol. Ecol. Notes* **7**, 355–364 (2007).
9. Lawrence, E. R. *et al.* Geo-referenced population-specific microsatellite data across American continents, the MacroPopGen Database. *Sci Data* **6**, 14 (2019).
10. Benson, D. A. *et al.* GenBank. *Nucleic Acids Res.* **41**, D36–42 (2013).
11. Arribas, P., Andújar, C. & Salces□Castellano, A. The limited spatial scale of dispersal in soil arthropods revealed with whole□community haplotype□level metabarcoding. *Molecular Ecology* (2021).

12. Macher, J.-N., Macher, T.-H. & Leese, F. Combining NCBI and BOLD databases for OTU assignment in metabarcoding and metagenomic datasets: The BOLD NCBI _Merger. *Metabarcoding and Metagenomics* vol. 1 e22262 (2017).
13. Stange, M., Barrett, R. D. H. & Hendry, A. P. The importance of genomic variation for biodiversity, ecosystems and people. *Nat. Rev. Genet.* **22**, 89–105 (2021).
14. Des Roches, S., Pendleton, L. H., Shapiro, B. & Palkovacs, E. P. Conserving intraspecific variation for nature's contributions to people. *Nat Ecol Evol* (2021) doi:10.1038/s41559-021-01403-5.
15. Hoban, S. *et al.* Global commitments to conserving and monitoring genetic diversity are now necessary and feasible. *Bioscience* (2021) doi:10.1093/biosci/biab054.
16. Santini, L. *et al.* The interface between Macroecology and Conservation: existing links and untapped opportunities. *Frontiers of Biogeography* **13**, e53025 (2021).
17. Schmidt, C. & Garroway, C. J. The conservation utility of mitochondrial genetic diversity in macrogenetic research. *Conserv. Genet.* (2021) doi:10.1007/s10592-021-01333-6.
18. Blanchet, S., Prunier, J. G. & De Kort, H. Time to Go Bigger: Emerging Patterns in Macrogenetics. *Trends Genet.* **33**, 579–580 (2017).
19. Leigh, D. M. *et al.* Opportunities and challenges of macrogenetic studies. *Nat. Rev. Genet.* (2021) doi:10.1038/s41576-021-00394-0.
20. Theodoridis, S. *et al.* Evolutionary history and past climate change shape the distribution of genetic diversity in terrestrial mammals. *Nat. Commun.* **11**, 2557 (2020).
21. Manel, S. *et al.* Global determinants of freshwater and marine fish genetic diversity. *Nat. Commun.* **11**, 692 (2020).
22. Miraldo, A. *et al.* An Anthropocene map of genetic diversity. *Science* **353**, 1532–1535 (2016).
23. Gratton, P. *et al.* Which Latitudinal Gradients for Genetic Diversity? *Trends in ecology & evolution* **32**, 724–726 (2017).

24. Barrow, L. N., Masiero da Fonseca, E., Thompson, C. E. P. & Carstens, B. C. Predicting amphibian intraspecific diversity with machine learning: Challenges and prospects for integrating traits, geography, and genetic data. *Mol. Ecol. Resour.* (2020) doi:10.1111/1755-0998.13303.

25. Millette, K. L. *et al.* No consistent effects of humans on animal genetic diversity worldwide. *Ecol. Lett.* **23**, 55–67 (2020).

26. Theodoridis, S., Rahbek, C. & Nogues□Bravo, D. Exposure of mammal genetic diversity to mid□21st century global change. *Ecography* **44**, 817–831 (2021).

27. Pelletier, T. A. & Carstens, B. C. Geographical range size and latitude predict population genetic structure in a global survey. *Biol. Lett.* **14**, 20170566 (2018).

28. Losey, J. E. & Vaughan, M. The Economic Value of Ecological Services Provided by Insects. *Bioscience* **56**, 311–323 (2006).

29. Valiente-Banuet, A. *et al.* Beyond species loss: the extinction of ecological interactions in a changing world. *Funct. Ecol.* **29**, 299–307 (2015).

30. Dincă, V. *et al.* High resolution DNA barcode library for European butterflies reveals continental patterns of mitochondrial genetic diversity. *Commun Biol* **4**, 315 (2021).

31. Papadopoulou, A. *et al.* Testing the Species–Genetic Diversity Correlation in the Aegean Archipelago: Toward a Haplotype-Based Macroecology? *Am. Nat.* **178**, 241–255 (2011).

32. Salinas-Ivanenko, S. & Múrria, C. Macroecological trend of increasing values of intraspecific genetic diversity and population structure from temperate to tropical streams. *Glob. Ecol. Biogeogr.* **30**, 1685–1697 (2021).

33. Baselga, A. *et al.* Whole-community DNA barcoding reveals a spatio-temporal continuum of biodiversity at species and genetic levels. *Nat. Commun.* **4**, 1892 (2013).

34. Wilson, R. J. & Fox, R. Insect responses to global change offer signposts for biodiversity and conservation. *Ecol. Entomol.* **46**, 699–717 (2021).

35. Wagner, D. L., Grames, E. M., Forister, M. L., Berenbaum, M. R. & Stopak, D. Insect

731 decline in the Anthropocene: Death by a thousand cuts. *Proc. Natl. Acad. Sci. U. S. A.* **118**,
732 e2023989118 (2021).
733 36. Gallien, L. & Carboni, M. The community ecology of invasive species: where are we and
734 what's next? *Ecography* **40**, 335–352 (2017).
735 37. Smith-Ramesh, L. M., Moore, A. C. & Schmitz, O. J. Global synthesis suggests that food
736 web connectance correlates to invasion resistance. *Glob. Chang. Biol.* **23**, 465–473 (2017).
737 38. Raven, P. H. & Wagner, D. L. Agricultural intensification and climate change are rapidly
738 decreasing insect biodiversity. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
739 39. Halsch, C. A. *et al.* Insects and recent climate change. *Proc. Natl. Acad. Sci. U. S. A.* **118**,
740 (2021).
741 40. Powney, G. D. *et al.* Widespread losses of pollinating insects in Britain. *Nat. Commun.* **10**,
742 1018 (2019).
743 41. van Klink, R. *et al.* Meta-analysis reveals declines in terrestrial but increases in freshwater
744 insect abundances. *Science* **368**, 417–420 (2020).
745 42. Forister, M. L., Pelton, E. M. & Black, S. H. Declines in insect abundance and diversity: We
746 know enough to act now. *Conservat Sci and Prac* **1**, e80 (2019).
747 43. Didham, R. K. *et al.* Interpreting insect declines: seven challenges and a way forward.
748 *Insect Conserv. Divers.* **13**, 103–114 (2020).
749 44. Montgomery, G. A. *et al.* Is the insect apocalypse upon us? How to find out. *Biol. Conserv.*
750 **241**, e108327 (2020).
751 45. Crossley, M. S. *et al.* No net insect abundance and diversity declines across US Long Term
752 Ecological Research sites. *Nat Ecol Evol* **4**, 1368–1376 (2020).
753 46. Fox, R. *et al.* Insect population trends and the IUCN Red List process. *J. Insect Conserv.*
754 **23**, 269–278 (2019).
755 47. Cardoso, P., Erwin, T. L., Borges, P. A. V. & New, Tim R. The seven impediments in
756 invertebrate conservation and how to overcome them. *Biol. Conserv.* **144**, 2647–2655

(2011).

48. Diniz-Filho, J. A. F., de Marco, P., Jr & Hawkins, B. A. Defying the curse of ignorance: perspectives in insect macroecology and conservation biogeography. *Insect Conserv. Divers.* (2010) doi:10.1111/j.1752-4598.2010.00091.x.

49. Grames, E. M. *et al.* Trends in global insect abundance and biodiversity: A community-driven systematic map protocol. *OSF Registries* (2019).

50. Wheeler, Q. D., Raven, P. H. & Wilson, E. O. Taxonomy: Impediment or expedient? *Science* **303**, 285 (2004).

51. Stribling, J. B., Pavlik, K. L., Holdsworth, S. M. & Leppo, E. W. Data quality, performance, and uncertainty in taxonomic identification for biological assessments. *J. North Am. Benthol. Soc.* **27**, 906–919 (2008).

52. Hebert, P. D. N. *et al.* Counting animal species with DNA barcodes: Canadian insects. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **371**, 20150333 (2016).

53. Meier, R. *et al.* A re-analysis of the data in Sharkey et al.'s (2021) minimalist revision reveals that BINs do not deserve names, but BOLD Systems needs a stronger commitment to open science. *bioRxiv* 2021.04.28.441626 (2021) doi:10.1101/2021.04.28.441626.

54. Hickerson, M. J., Meyer, C. & Moritz, C. DNA-Barcoding will often fail to discover new animal species over broad parameter space. *Syst. Biol.* **55**, 729–739 (2006).

55. Allio, R., Donega, S., Galtier, N. & Nabholz, B. Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Mol. Biol. Evol.* **34**, 2762–2772 (2017).

56. Hudson, R. R. & Turelli, M. Stochasticity overrules the 'three-times rule': genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution* **57**, 182–190 (2003).

57. Meiklejohn, C. D., Montooth, K. L. & Rand, D. M. Positive and negative selection on the mitochondrial genome. *Trends Genet.* **23**, 259–263 (2007).

783 58. Hurst, G. D. D. & Jiggins, F. M. Problems with mitochondrial DNA as a marker in
784 population, phylogeographic and phylogenetic studies: the effects of inherited symbionts.
785 *Proc. Biol. Sci.* **272**, 1525–1534 (2005).
786 59. Deiner, K. *et al.* Environmental DNA metabarcoding: Transforming how we survey animal
787 and plant communities. *Mol. Ecol.* **26**, 5872–5895 (2017).
788 60. Sigsgaard, E. E. *et al.* Population-level inferences from environmental DNA-Current status
789 and future perspectives. *Evol. Appl.* **13**, 245–262 (2020).
790 61. Overcast, I. *et al.* A unified model of species abundance, genetic diversity, and functional
791 diversity reveals the mechanisms structuring ecological communities. *bioRxiv*
792 2020.01.30.927236 (2020) doi:10.1101/2020.01.30.927236.
793 62. Wallace, A. R. & Harvard University. *Tropical nature, and other essays*. (London, Macmillan
794 and co., 1878).
795 63. Janzen, D. H. Why Mountain Passes are Higher in the Tropics. *Am. Nat.* **101**, 233–249
796 (1967).
797 64. Moreau, C. S. & Bell, C. D. Testing the museum versus cradle tropical biological diversity
798 hypothesis: phylogeny, diversification, and ancestral biogeographic range evolution of the
799 ants. *Evolution* **67**, 2240–2257 (2013).
800 65. Buffalo, V. Quantifying the relationship between genetic diversity and population size
801 suggests natural selection cannot explain Lewontin's Paradox. *eLife* **10**, e67509 (2021).
802 66. Stebbins, G. L. *Flowering Plants: Evolution above the Species Level*. (Harvard University
803 Press, 1974). doi:10.4159/harvard.9780674864856.
804 67. Gaston, K. J. & Blackburn, T. M. The tropics as a museum of biological diversity: an
805 analysis of the New World avifauna. *Proceedings of the Royal Society of London. Series B:*
806 *Biological Sciences* **263**, 63–68 (1996).
807 68. Chown, S. L. & Gaston, K. J. Areas, cradles and museums: the latitudinal gradient in
808 species richness. *Trends Ecol. Evol.* **15**, 311–315 (2000).

809 69. Stevens, G. C. The Latitudinal Gradient in Geographical Range: How so Many Species
810 Coexist in the Tropics. *Am. Nat.* **133**, 240–256 (1989).
811 70. Ruggiero, A. & Werenkraut, V. One-dimensional analyses of Rapoport's rule reviewed
812 through meta-analysis. *Glob. Ecol. Biogeogr.* **16**, 401–414 (2007).
813 71. Hewitt, G. The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913 (2000).
814 72. Carnaval, A. C., Hickerson, M. J., Haddad, C. F. B., Rodrigues, M. T. & Moritz, C. Stability
815 predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science* **323**, 785–789
816 (2009).
817 73. Anderson, S. C. & Ward, E. J. Black swans in space: modeling spatiotemporal processes
818 with extremes. *Ecology* **100**, e02403 (2019).
819 74. Pielou, E. C. *After The Ice Age*. (The University of Chicago Press, 1991).
820 75. Dunn, R. R. *et al.* Climatic drivers of hemispheric asymmetry in global patterns of ant
821 species richness. *Ecol. Lett.* **12**, 324–333 (2009).
822 76. Economo, E. P., Narula, N., Friedman, N. R., Weiser, M. D. & Guénard, B. Macroecology
823 and macroevolution of the latitudinal diversity gradient in ants. *Nat. Commun.* **9**, 1778
824 (2018).
825 77. Privet, K. & Petillon, J. Differences in tropical vs. temperate diversity in arthropod predators
826 provide insights into causes of latitudinal gradients of species diversity. *bioRxiv* 283499
827 (2018) doi:10.1101/283499.
828 78. Kreft, H. & Jetz, W. Global patterns and determinants of vascular plant diversity. *Proc. Natl.*
829 *Acad. Sci. U. S. A.* **104**, 5925–5930 (2007).
830 79. Orr, M. C. *et al.* Global Patterns and Drivers of Bee Distribution. *Curr. Biol.* **31**, 451–458.e4
831 (2021).
832 80. Vellend, M. Island biogeography of genes and species. *Am. Nat.* **162**, 358–365 (2003).
833 81. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography*. (Princeton
834 University Press, 2001). doi:10.1515/9781400837526.

82. Laroche, F., Jarne, P., Lamy, T., David, P. & Massol, F. A neutral theory for interpreting correlations between species and genetic diversity in communities. *Am. Nat.* **185**, 59–69 (2015).

83. Vellend, M. & Geber, M. A. Connections between species diversity and genetic diversity. *Ecol. Lett.* **8**, 767–781 (2005).

84. Lamy, T., Laroche, F., David, P., Massol, F. & Jarne, P. The contribution of species-genetic diversity correlations to the understanding of community assembly rules. *Oikos* **126**, 759–771 (2017).

87. Labandeira, C. C. & Sepkoski, J. J., Jr. Insect diversity in the fossil record. *Science* **261**, 310–315 (1993).

88. Grimaldi, D., Engel, M. S., Engel, M. S. *Evolution of the Insects*. (Cambridge University Press, 2005).

89. Mitton, J. B. *Selection in Natural Populations*. (Oxford University Press, 2000).

90. Tajima, F. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437–460 (1983).

91. Addo-Bediako, A., Chown, S. L. & Gaston, K. J. Thermal tolerance, climatic variability and latitude. *Proc. Biol. Sci.* **267**, 739–745 (2000).

92. Lawrence, E. R. & Fraser, D. J. Latitudinal biodiversity gradients at three levels: Linking species richness, population richness and genetic diversity. *Glob. Ecol. Biogeogr.* **29**, 770–788 (2020).

93. Tougeron, K. Diapause research in insects: historical review and recent work perspectives. *Entomol. Exp. Appl.* **167**, 27–36 (2019).

94. Danks, H. V. & Others. The wider integration of studies on insect cold-hardiness. *Eur. J. Entomol.* **93**, 383–404 (1996).

95. Du, C., Chen, J., Jiang, L. & Qiao, G. High correlation of species diversity patterns between specialist herbivorous insects and their specific hosts. *J. Biogeogr.* **47**, 1232–1245 (2020).

96. Becerra, J. X. & Lawrence Venable, D. Macroevolution of insect–plant associations: The relevance of host biogeography to host affiliation. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 12626–12631 (1999).

97. Futuyma, D. J. & Agrawal, A. A. Macroevolution and the biological diversity of plants and herbivores. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 18054–18061 (2009).

98. Schoonhoven, L. M., Van Loon, B., van Loon, J. J. A. & Dicke, M. *Insect-Plant Biology*. (OUP Oxford, 2005).

99. Francis, A. P. & Currie, D. J. A globally consistent richness-climate relationship for angiosperms. *Am. Nat.* **161**, 523–536 (2003).

100. Kerkhoff, A. J., Moriarty, P. E. & Weiser, M. D. The latitudinal species richness gradient in New World woody angiosperms is consistent with the tropical conservatism hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8125–8130 (2014).

101. Basset, Y. *et al.* Arthropod diversity in a tropical forest. *Science* **338**, 1481–1484 (2012).

102. Barbour, M. A. *et al.* Multiple plant traits shape the genetic basis of herbivore community assembly. *Funct. Ecol.* **29**, 995–1006 (2015).

103. Barbour, M. A. *et al.* Genetic specificity of a plant-insect food web: Implications for linking genetic variation to network complexity. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 2128–2133 (2016).

104. Janská, V. *et al.* Palaeodistribution modelling of European vegetation types at the Last Glacial Maximum using modern analogues from Siberia: Prospects and limitations. *Quat. Sci. Rev.* **159**, 103–115 (2017).

105. Earl, C. *et al.* Spatial phylogenetics of butterflies in relation to environmental drivers and angiosperm diversity across North America. *iScience* **24**, 102239 (2021).

106. Rix, M. G. *et al.* Biogeography and speciation of terrestrial fauna in the south-western Australian biodiversity hotspot. *Biol. Rev. Camb. Philos. Soc.* **90**, 762–793 (2015).

107. Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J.

887 Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).

888 108. Corser, J. D., White, E. L. & Schlesinger, M. D. Odonata origins, biogeography, and
889 diversification in an Eastern North American hotspot: multiple pathways to high temperate
890 forest insect diversity. *Insect Conserv. Divers.* **7**, 393–404 (2014).

891 109. Hawkins, B. A. & DeVries, P. J. Tropical niche conservatism and the species richness
892 gradient of North American butterflies. *J. Biogeogr.* **36**, 1698–1711 (2009).

893 110. Tsutsui, N. D., Suarez, A. V., Holway, D. A. & Case, T. J. Reduced genetic variation and
894 the success of an invasive species. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5948–5953 (2000).

895 111. Frankham, R. Resolving the genetic paradox in invasive species. *Heredity* **94**, 385 (2005).

896 112. Keith, A. R., Bailey, J. K., Lau, M. K. & Whitham, T. G. Genetics-based interactions of
897 foundation species affect community diversity, stability and network structure. *Proc. Biol.*
898 *Sci.* **284**, (2017).

899 113. Shuster, S. M., Lonsdorf, E. V., Wimp, G. M., Bailey, J. K. & Whitham, T. G. Community
900 heritability measures the evolutionary consequences of indirect genetic effects on
901 community structure. *Evolution* **60**, 991–1003 (2006).

902 114. Barbour, M. A. & Fortuna, M. A. Genetic specificity of a plant–insect food web: Implications
903 for linking genetic variation to network complexity. *Proceedings of the* (2016).

904 115. Raffard, A., Santoul, F., Cucherousset, J. & Blanchet, S. The community and ecosystem
905 consequences of intraspecific diversity: a meta-analysis. *Biol. Rev. Camb. Philos. Soc.* **94**,
906 648–661 (2019).

907 116. Riggio, J. *et al.* Global human influence maps reveal clear opportunities in conserving
908 Earth's remaining intact terrestrial ecosystems. *Glob. Chang. Biol.* (2020)
909 doi:10.1111/gcb.15109.

910 117. Hortal, J. *et al.* Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annu.*
911 *Rev. Ecol. Evol. Syst.* **46**, 523–549 (2015).

912 118. Coddington, J. A., Agnarsson, I., Miller, J. A., Kuntner, M. & Hormiga, G. Undersampling

913 bias: the null hypothesis for singleton species in tropical arthropod surveys. *J. Anim. Ecol.*
914 **78**, 573–584 (2009).
915 119. Leigh, D. M., Hendry, A. P., Vázquez-Domínguez, E. & Friesen, V. L. Estimated six per
916 cent loss of genetic variation in wild populations since the industrial revolution. *Evol. Appl.*
917 **12**, 1505–1512 (2019).
918 120. Teixeira, J. C. & Huber, C. D. The inflated significance of neutral genetic diversity in
919 conservation genetics. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
920 121. DeWoody, J. A., Harder, A. M., Mathur, S. & Willoughby, J. R. The long-standing
921 significance of genetic diversity in conservation. *Molecular ecology* vol. 30 4147–4154
922 (2021).
923 122. García-Dorado, A. & Caballero, A. Neutral genetic diversity as a useful tool for conservation
924 biology. *Conserv. Genet.* **22**, 541–545 (2021).
925 123. Reed, D. H. & Frankham, R. Correlation between fitness and genetic diversity. *Conserv.*
926 *Biol.* **17**, 230–237 (2003).
927 124. Wilson, E. O. The little things that run the world* (the importance and conservation of
928 invertebrates). *Conserv. Biol.* **1**, 344–346 (1987).
929 125. Ratnasingham, S. & Hebert, P. D. N. A DNA-based registry for all animal species: the
930 barcode index number (BIN) system. *PLoS One* **8**, e66213 (2013).
931 126. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence
932 alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
933 127. Pallewatta, N., Reaser, J. K. & Gutierrez, A. T. Invasive alien species in South-Southeast
934 Asia: national reports and directory of resources. *The Global Invasive Species Programme.*
935 Cape Town, South Africa (2003).
936 128. Hoban, S. *et al.* Genetic diversity targets and indicators in the CBD post-2020 Global
937 Biodiversity Framework must be improved. *Biol. Conserv.* **248**, 108654 (2020).
938 129. Paz-Vinas, I. *et al.* Systematic conservation planning for intraspecific genetic diversity.

939 *Proc. Biol. Sci.* **285**, 20172746 (2018).

940 130. Nei, M. & Li, W. Mathematical model for studying variation in terms of restriction
941 endonucleases. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 5269–5273 (1979).

942 131. Gaggiotti, O. E. *et al.* Diversity from genes to ecosystems: A unifying framework to study
943 variation across biological metrics and scales. *Evol. Appl.* **11**, 1176–1193 (2018).

944 132. Chao, A., Chiu, C.-H. & Jost, L. Unifying Species Diversity, Phylogenetic Diversity,
945 Functional Diversity, and Related Similarity and Differentiation Measures Through Hill
946 Numbers. *Annu. Rev. Ecol. Evol. Syst.* (2014) doi:10.1146/annurev-ecolsys-120213-
947 091540.

948 133. Alberdi, A. & Gilbert, M. T. P. A guide to the application of Hill numbers to DNA-based
949 diversity analyses. *Mol. Ecol. Resour.* **19**, 804–817 (2019).

950 134. Shannon, C. E. A mathematical theory of communication. *The Bell System Technical*
951 *Journal* **27**, 379–423 (1948).

952 135. Hill, M. O. Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology* **54**,
953 427–432 (1973).

954 136. Maurer, B. A. & McGill, B. J. Measurement of species diversity. *Biological diversity: frontiers*
955 *in measurement and assessment* 55–65 (2011).

956 137. Chao, A. *et al.* Rarefaction and extrapolation with Hill numbers: a framework for sampling
957 and estimation in species diversity studies. *Ecol. Monogr.* **84**, 45–67 (2014).

958 138. Welch, B. L. The Significance of the Difference Between Two Means when the Population
959 Variances are Unequal. *Biometrika* **29**, 350–362 (1938).

960 139. Karger, D. N. *et al.* Climatologies at high resolution for the earth's land surface areas.
961 *Scientific Data* **4**, 170122 (2017).

962 140. Brown, J. L., Hill, D. J., Dolan, A. M., Carnaval, A. C. & Haywood, A. M. PaleoClim, high
963 spatial resolution paleoclimate surfaces for global land areas. *Sci Data* **5**, 180254 (2018).

964 141. Tuanmu, M.-N. & Jetz, W. A global, remote sensing-based characterization of terrestrial

habitat heterogeneity for biodiversity and ecosystem modelling. *Glob. Ecol. Biogeogr.* **24**, 1329–1339 (2015).

142. Kennedy, C. M., Oakleaf, J. R., Theobald, D. M., Baruch-Mordo, S. & Kiesecker, J. Managing the middle: A shift in conservation priorities based on the global human modification gradient. *Glob. Chang. Biol.* **25**, 811–826 (2019).

143. White, A. E., Dey, K. K., Mohan, D., Stephens, M. & Price, T. D. Regional influences on community structure across the tropical-temperate divide. *Nat. Commun.* **10**, 2646 (2019).

144. Goodrich, B., Gabry, J., Ali, I. & Brilleman, S. rstanarm: Bayesian applied regression modeling via Stan. *R package version* **2**, 1758 (2018).

145. Piironen, J., Paasiniemi, M. & Vehtari, A. Projective inference in high-dimensional problems: Prediction and feature selection. *EJSS* **14**, 2155–2197 (2020).

146. Legendre, P. & Fortin, M. J. Spatial pattern and ecological analysis. *Vegetatio* **80**, 107–138 (1989).

147. Bivand, R. S. & Wong, D. W. S. Comparing implementations of global and local indicators of spatial association. *Test* **27**, 716–748 (2018).

148. Carpenter, B. *et al.* Stan: A Probabilistic Programming Language. *Journal of Statistical Software, Articles* **76**, 1–32 (2017).

149. Stan Development Team. RStan: the R interface to Stan. (2020).

150. Gimenez, O., Morgan, B. J. T. & Brooks, S. P. Weak Identifiability in Models for Mark-Recapture-Recovery Data. in *Modeling Demographic Processes In Marked Populations* (eds. Thomson, D. L., Cooch, E. G. & Conroy, M. J.) 1055–1067 (Springer US, 2009). doi:10.1007/978-0-387-78151-8_48.

151. Elith, J., Kearney, M. & Phillips, S. The art of modelling range-shifting species. *Methods Ecol. Evol.* **1**, 330–342 (2010).

Acknowledgements

We thank Jason L Brown for his valuable feedback on early drafts of the manuscript, and the Hickerson and Carnaval labs for their feedback at every step of the project. CMF, MJH, ACC, IO and AR acknowledge support from NSF DBI 2104147 and the NSF RCN Cross-Scale Processes Impacting Biodiversity (DEB 1745562). DJL was funded by NSF DEB-1541557.

Author contributions

C.M.F., L.D.B., and M.J.H. conceived of the study. C.M.F., L.D.B., J.M.K., K.A.M., I.O., A.R., P.S., A.C.C., and M.J.H. framed the study. C.M.F., L.D.B., and M.J.H. carried out the analyses. C.M.F., and M.J.H. led the writing. All authors contributed to interpretation of the results and to the writing, and all have approved the submission. A.R. and P.S. contributed to Figure 1.

Competing interests

The authors declare no competing interests.