OIKOS

Research

Robust metagenomic evidence that local assemblage richness increases with latitude in ground-active invertebrates of North America

Michael D. Weiser, Cameron D. Siler, Sierra N. Smith, Katie E. Marshall, Jessica F. McLaughlin, Matthew J. Miller and Michael Kaspari

M. D. Weiser (https://orcid.org/0000-0001-9080-0834) ✉ (michael.d.weiser@ou.edu), C. D. Siler, S. N. Smith, J. F. McLaughlin, M. J. Miller and M. Kaspari (https://orcid.org/0000-0002-9717-5768), Univ. of Oklahoma, Dept of Biology, Norman, OK, USA. MDW and MK also at: Univ. of Oklahoma, Geographical Ecology Group, Norman, OK, USA. – K. E. Marshall, Univ. of British Columbia, Dept of Zoology, Vancouver, BC, Canada. CDS, SNS and JFM also at: Sam Noble Oklahoma Museum of Natural History, Norman, OK, USA. MJM also at: Reneco International Wildlife Consultants, Abu Dhabi, UAE.

Oikos
2022: e08791
doi: 10.1111/oik.08791
Subject Editor: Matty Berg
Editor-in-Chief:
Gerlinde B. De Deyn
Accepted 31 March 2022

Biodiversity monitoring is imperative for understanding how changing climate may impact the distributions of taxa from single species to the spatial distribution of biological diversity. Large-scale and cross-taxa biodiversity monitoring also allows an empirical understanding of biogeographic patterns across taxa. One such pattern, where in taxonomic richness peaks at tropical latitudes are typically treated as a biogeographical rule with few notable exceptions. Here we leveraged the invertebrate pitfall collections of the National Ecological Observatory Network (NEON) across North America to describe patterns of local taxonomic richness across taxa and across taxonomic scale. We focused on Arthropoda, Annelida and Mollusca. Additionally, we estimated regional species richness using expert-identified samples of three NEON sentinel taxa: Carabidae, Culicidae and Ixodida. To sample pitfall animals, we filtered storage ethanol and employed environmental DNA-barcoding methodologies to amplify and sequence extracted DNA from the filtrate for two regions of a mitochondrial gene. We assigned taxonomic names to these sequences at 97% similarity to reference sequences and calculated local taxonomic richness at the levels of species, genus, family and order. We calculated local species richness for 12 common invertebrate taxa. We used generalized linear models to describe the relationships between taxonomic richness and spatial, climatic and abundance predictor variables. At four taxonomic scales, ranging from species to order, taxonomic richnessincreased significantly as a function of latitude. Of the twelve invertebrate taxa weexamined, seven mirrored this positive latitudinal gradient in species richness. At theregional scale, two of three NEON Sentinel Taxa showed positive latitudinal gradientsin species richness. Temperature, precipitation, abundance and sequence read numberplayed minor roles in explaining patterns of taxonomic richness. When considering these mostly temperate sites that span 46 degrees of latitude, we found no support for the expected negative latitudinal gradients across taxa and taxonomic scales. Instead, for many of these taxa and taxonomic scales, we observed significant, positive richness gradients with increasing latitude among ground-dwelling invertebrate communities. Thus, one of the most 'general' patterns in biogeography was not found for most invertebrate taxa across temperate latitudes.

NORDIC SOCIETY OIKOS
www.oikosjournal.org

Keywords: environmental DNA, latitudinal gradient, metagenomic barcoding, species richness, taxonomic richness, taxonomic scale

Introduction

The tendency for taxonomic richness, the number of taxa found in a given area, to decline with increasing latitude is considered one of the most pervasive patterns in biogeography (Hillebrand 2004, Willig and Presley 2018). Despite a number of processes put forth to explain such patterns (e.g. competition, predation, solar energy, productivity, etc.), there is no clear consensus on the mechanisms that generate and maintain latitudinal gradients in taxonomic richness (LGTRs) (Rohde 1992, Willig et al. 2003, Willig and Presley 2018). However, mechanisms invoked to explain what drives such large spatial patterns must operate, by necessity, across large spatial scales. Most of the hypotheses generated to explain LGTRs are post hoc and meant to explain observed negative relationships between latitude and taxon richness. As mechanisms that predict negative LGTRs should not explain the opposite pattern (Pyron and Burbrink 2009, Weiser et al. 2018), understanding the generality of this pattern is a vital step to testing the generality of hypothesized mechanisms.

Despite the numerous examples of peak taxonomic richness in the tropics, there are notable exceptions where species richness peaks outside in extra-tropical areas and shows a positive relationship with increasing latitude (Aphids in Dixon et al. 1987; ichneumonid wasps in Owen and Owen 1974 and Janzen 1981; sawflies in Kouki et al. 1994; arctic plants in Marshall and Baltzer 2015). Many of these positive LGTRs are observed at intermediate taxonomic scales (e.g. within families) from groups that at higher taxonomic scales (e.g. order, phylum) show a traditional negative gradient in species richness (e.g. the plant family Caryophyllaceae within the order Caryophyllales, Weiser et al. 2018). Taxon-scale-dependent patterns can be detected by 'deconstructing' the overall LGTR at finer taxonomic patterns to test for generality across sub-taxa. In addition to negative LGTRs being hidden by patterns at higher taxonomic scales, latitudinal gradients in species richness can be strongly impacted by one to few diverse taxa. For example, in a comparison of published floras (Weiser et al. 2018), two families (Asteraceae and Fabaceae) were responsible for more than a third of the difference in species richness across latitudes while more than half of plant families contributed nothing to the difference in species richness.

A difficulty of the cross-taxon approach is that published studies describing LGTRs also differ in spatial sampling grain and spatial extent, and both affect the slope of LGTRs (Hillebrand 2004, Kinlock et al. 2017). Large-scale 'polygon' studies typically collate species lists from the literature and/or use overlap of range maps to estimate species richness at the scale of large grid cells or political boundaries (Guenard et al. 2012). Local-scale community studies, where researchers typically collect samples at a fine spatial scale, are rarer, especially for terrestrial invertebrate animals. Local-scale studies that directly sample the environment may also be limited in taxonomic breadth to one to few taxa that the researchers know well enough to quantify taxonomic richness, for example, spiders (Pinkus-Rendón et al. 2005, Privet et al. 2020) or ants (Kaspari et al. 2000a, b).

In a meta-analysis of species-level LGTRs by Hillebrand (2004), the first major conclusion is that negative LGTRs are 'a ubiquitous phenomenon' (Hillebrand 2004). Other results were that the sample area mattered that local-scale studies had less steep LGTRs than regional-scale studies. In the meta-analysis 222 of 581 (38%) of the gradients examined used local-scale samples. Of these 222 gradients, only 23 (4%) examine terrestrial invertebrates (and eleven were from three families of Hymenoptera). These 23 are from 16 different studies (and thus likely 16 different sampling methods, grains and extents). Of these 23, only seven show the significant negative LGTR slopes expected while none show a significant positive LGTR slope (e.g. Western Hemisphere and Euro-African bees in Michener 1979).

Here we develop two novel methodological pipelines to non-destructively quantify continental-scale spatial patterns of local-scale taxonomic richness from archived samples of terrestrial invertebrate communities. We quantify ground-active invertebrate assemblages collected in pitfall trap arrays across the continental US, Alaska and Puerto Rico by the National Ecological Observatory Network (NEON). We first employ metagenomics methods from environmental DNA techniques (Barnes and Turner 2016, Taberlet et al. 2018) to develop a novel pipeline for non-destructively sampling vouchered invertebrate assemblages collected in pitfall traps by filtering the storage ethanol. Pitfall traps measure the activity density of ground-active animals, and higher abundance can lead to higher richness (Srivastava and Lawton 1998). Therefore, we develop a second pipeline that uses high-resolution imaging of whole pitfall contents to count the number of individuals in each sample. Together, these pipelines provide a means for non-destructive quantification and identification of numerous taxa and vouchered collections without the need for broad taxonomic expertise or complicated specimen preparation (e.g. pinning insects). Using these pipelines, we estimate taxonomic richness and abundance from three major phyla commonly found in pitfall traps: Arthropoda, Annelida and Mollusca. We generate and compare models across taxa and taxonomic levels to test the importance of climate, sampling quality and purely spatial parameters for predicting species richness. Thus, we are leveraging NEON's spatially, temporally and methodologically standardized samples to test for generality in the LGTR for ground-active invertebrates.

Methods

NEON sites and invertebrate by-catch

NEON performs standardized sampling at their 47 terrestrial sites and is planned to continue this sampling for 30 years (Hoekman et al. 2017). By necessity, NEON focuses on a few focal taxa, including mosquitos (Diptera: Culicidae), ticks (Acari: Ixodida) and ground beetles (Coleoptera: Carabidae) using three different collection methods: carbon dioxide traps to collect mosquitos, drag-cloths to collect ticks (Thorpe et al. 2016), and pitfall trap arrays to collect ground beetles. NEON field staff removes, identifies and counts adult ground beetles from the pitfall trap samples and archives the pitfall residuum as 'Invertebrate By-Catch.' This bycatch is then archived in 95% ethanol in (typically) 50 ml centrifuge tubes.

We examined this invertebrate bycatch from 56 NEON trap arrays from 26 sites (Fig. 1, also see the Supporting information for sample information). Each array consisted of four 11 cm diameter cups with dilute propylene glycol placed on the perimeter of a 20 × 20 m square. Where a site-by-habitat type was represented more than once, we averaged taxonomic richness (read number, etc.) for that site-by-habitat type, leaving 49 site-by-habitat combinations. We sampled three 2-week pitfall trap events per array, selecting the first, last and a trap event in the middle of the season. All arrays we use here were collected during the growing season of 2016 and we processed them in the summer of 2019.

Environmental DNA filtration and extraction

To quantify the taxonomic richness in the NEON samples, we used methods modified from current environmental DNA techniques (Taberlet et al. 2018), which allowed us to non-destructively sample DNA from the storage ethanol. All of the storage ethanol was vacuum filtered using a 50 ml vacuum filtration system with a 0.45 µm pore size. After filtering, we used sterilized scissors to cut the filter from the tube which was then stored at −20°C. To extract DNA, each filter was torn into smaller pieces and then dried in a SpeedVac (ThermoFisher) at low temperature for 20 min. We added 380 µl ATL buffer and 20 µl Proteinase K (DNeasy Kit, Qiagen Corp.) before vortexing for 20 s and then incubating at 55°C overnight. DNA was extracted from the filters using the Qiagen QIAshredder and Qiagen DNeasy Blood and Tissue extraction kits following the protocol outlined in the Supporting information.

Metagenomic barcoding of invertebrate communities

Following DNA extraction, we utilized a two-step polymerase chain reaction (PCR) protocol to amplify three

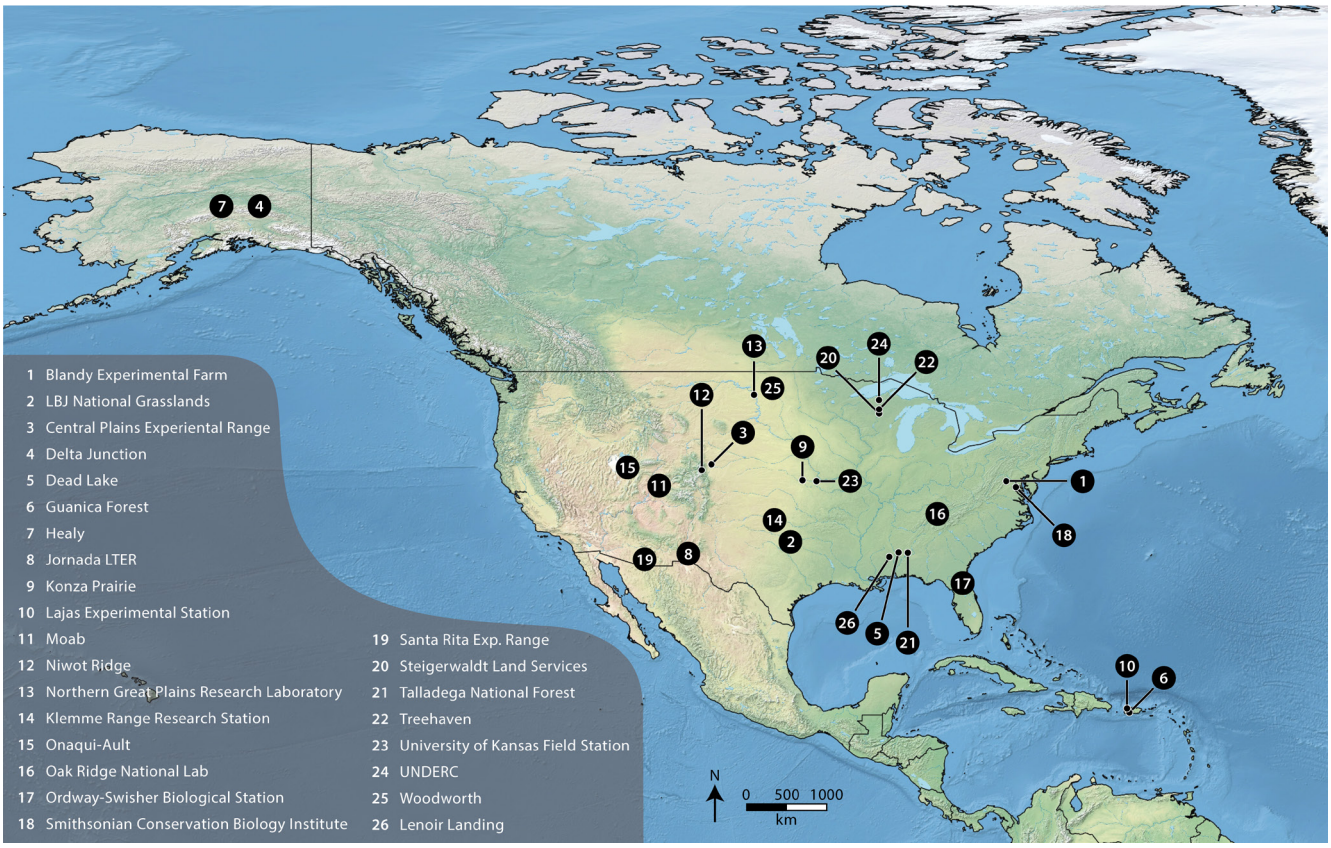

Figure 1. NEON sites used in this study. See the Supporting information for complete listing

fragments of the mitochondrial cytochrome oxidase I gene (COI): *F230R*, *157* and *Lep* (see the Supporting information, Rubbmark et al. 2018, Hajibabaei et al. 2019). Two-step PCR eliminates separate annealing and extension steps; instead, annealing and extension are combined in a single step at a relatively low temperature (48°C), but for a relatively long time (5 min). Initial experiments showed that two-step PCR outperformed traditional three-step PCR. PCR amplification was confirmed by gel electrophoresis and PCR products were cleaned with KAPA Pure Beads (Roche Sequencing Solutions). A second PCR was performed using the cleaned PCR products to attach unique Nextera paired-end indices (IDT for Illumina DNA/RNA UD Indexes, Illumina Inc.) and i5 and i7 adaptor sequences to the amplified DNA so the prepared product was able to bind to the surface of the Illumina MiSeq flow cell (Klymus et al. 2017). Next, PCR products were cleaned with KAPA Pure Beads (Roche Sequencing Solutions), and then quantified using a Qubit fluorometer (Invitrogen). After quantification, all samples were normalized to 6 nM of DNA before pooling samples into a sterile, 1.5 ml microcentrifuge tube. If the DNA quantity of a sample was above 6 nM, 5 ml of the sample was added to a calculated amount of sterile, laboratory grade water to dilute the sample to 6 nM. After dilution, 4 ml of the diluted PCR product was added to the final pool. If the DNA quantity of a sample was below 6 nM, no water was added to the sample and 2 ml of the PCR product was added to the final pool. Sequencing was performed at the University of Oklahoma Consolidated Core Lab on an Illumina MiSeq, producing 2 × 500 bp reads for F230R primers, 2 × 300 bp reads for 157 and 1 × 600 bp reads for Lep.

Taxonomic assignment

We used the command '-fastq_mergepairs' in USEARCH (Edgar 2010) to assemble paired-end reads to create consensus sequences for the 2 paired primers (157 and LCO). We size-filtered all primers to the expected sequence length and quality filtered ('-fastq_filter', '-fastq_maxee') removing any sequence with more than one error. We removed singletons ('-sortbysize') before clustering sequences to OTUs ('-cluster_otus') and then assigning all reads (including singletons) to OTUs ('-otutab'). We then used BLASTn (Altschul et al. 1990) to assign taxonomy to each OTU, selecting the best 10 matches between the consensus OTU sequence and NCBI databases. To set a limit for inclusion, we used linear regression of e-score on percent similarity for each primer to calculate the average e-score at 97% similarity between the OTU consensus sequence and the BLAST search. Where the best match was not determined to species in the NCBI data (e.g. order, family, genus) we used the taxonomy from matches with lower scores, provided they did not contradict any matches with better scores and met the minimum critical quality score based on 97% similarity. We refer to reads that meet these quality and taxonomic criteria as High-Quality reads ('HQreads').

Using the results from BLASTn search, we assigned taxonomy to each consensus sequence, using Integrated Taxonomic Information System (ITIS) as a standard baseline (retrieved between 10 January 2020 and 30 September 2020 (<www.itis.gov>)). From this data we calculated array taxonomic richness at the level of species, genus, family and order. Where sequence identity was not resolved to species, we counted it as a species only if there were no other species from the same taxon present in that sample. For example, if *Pogonomyrmex barbatus* and *Pogonomyrmex* sp. were found in the same sample, the two labels were counted as one species. Across the dataset about 3% of the unique event-by-taxon occurrences had to be merged to a single species due to taxonomic incompleteness, but more than half of these were from two genera with relatively low local abundance and diversity (i.e. the cricket *Gryllus* sp., the roach *Parcoblatta* sp.).

Abundance and biomass calculations through image analysis

To calculate abundance and estimate biomass, we used high resolution imaging. Complete imaging methods are in Weiser et al. (2021). We digitally imaged trap contents at a resolution of 729 pixels per mm^2 (Blaire et al. 2020). We identified each individual specimen to order when possible and did not count disarticulated appendages. We calculated trap array abundance by summing the number of individuals for the three trap events at each array. As a proxy for biomass, we calculated the 'total image area' by summing the number of pixels from each individual image. As another proxy for total biomass, we estimated the total 'biovolume' in ml using the gradations on the outside of the centrifuge tubes (summing if a trap event had multiple tubes). Where a site had two arrays from the same habitat type, we averaged taxonomic richness.

Expert identified NEON Sentinel Taxa

We downloaded the expert identified taxa tables ('expertTaxonomistIDProcessed', <www.neonscience.org/>) which lists individuals identified to species using traditional methods by taxonomic expert for ground beetles, mosquitos and ticks. From these tables, we counted the total number of species collected in 2016 pitfall traps giving a site species richness for that year (Carabidae: 36 sites, Culicidae: 31 sites and Ixodida: 33 sites). As this combines 10 pitfall arrays per site, the spatial scale for the Sentinel Taxa is more regional than the local data from individual trap arrays.

Taxonomic richness and biogeographic analyses

All statistics were performed in R (<www.r-project.org>). We used three sets of generalized linear models (GLMs) with Quasi-Poisson distributions in R with analysis of deviance to calculate p-values using the R package car (Fox and Weisberg 2011). For the invertebrate bycatch data (i.e. trap array taxon richness) the predictor variables included the

purely geographical variables latitude and longitude. While obviously strongly correlated with latitude, we included the quadratic term for latitude to allow for model curvature across latitude and test for a unimodal 'peak' in taxonomic richness within the endpoints. As the quadratic term for latitude is calculated from latitude, we re-ran the models without linear latitude to estimate variance inflation due to this collinearity and calculated the change in mean variance inflation factor using the R package car (Fox and Weisberg 2011). We also used climatic variables (mean annual temperature 'MAT', mean annual precipitation 'MAP'), and sampling variables (abundance counts from images 'N' and number of HQreads in the three focal phyla from the sequencing). Mean annual temperature and mean annual precipitation for each trap array were calculated for the NEON sampling season means taken from the Daymet V4 database (Thornton et al. 2021). Biogeographic analyses presented here were limited to the three major invertebrate phyla found in the pitfalls: Arthropoda, Annelida and Mollusca.

In the first series of GLMs ('cross-taxon models') we modeled invertebrate by-catch richness of all common taxa (10 groups of arthropods a well as mollusks and annelids) as a function of the predictor variables. In the second series of GLMs ('taxon-scale models'), we modeled against the total taxonomic richness of the three focal phyla combined at four taxonomic scales, the richness of: species, genera, families and orders. In the third series of GLMs ('NEON sentinel taxa') we look at the species richness of the three sentinel (Arthropoda) taxa collected and identified by NEON: ground beetles, mosquitos and ticks.

Each NEON site is sampled for different season lengths, with warmer southern sites sampled across a longer season than colder northern sites. A longer season, combined with temporal turnover in active fauna, may lead to higher taxon richness with longer season lengths. To test for this, we calculated the linear slope of the species accumulation curve for each array, both as an ordinal (i.e. samples 1, 2 and 3) and the date of collection. We used linear regression to test if the slopes of these species accumulation curves varied as a function of latitude. To test for latitudinal bias in sequencing we used a series of linear regressions to examine the relationships between latitude and total read number, HQreads, %HQreads (i.e. HQreads/total reads × 100). To test for biases due to abundance and biomass we used linear regressions of %HQreads on total image area, biovolume and abundance.

As the island fauna of Puerto Rico is known to be depauperate for some taxa (Torres 1984) and the sites in Alaska are more than 22° of latitude north of the northernmost sites in the continental US, we investigated the impact of their inclusion in the dataset on the overall LGTRs. We repeated analyses above on datasets excluding both Alaska and Puerto Rico arrays. Using all the sites, we tested for differences due to habitat using an ANCOVA of species richness on latitude between 'open' (e.g. grasslands, agricultural) or 'closed' (forests).

Results

Sequencing results

The sequencing efforts returned 21.2 million reads, of which 19.6 million could be linked to individual sampling events (i.e. 7.6% of the reads did not have adapter sequences). The three primer sets returned different numbers of total reads, with primer set 157 returning the most reads (14.4 million) and LCO the least (83 754). The Lep primers returned 7.9 million reads, 4.4 million forward and 3.5 million reverse. Summed read number per site (three combined samples across all primers) averaged 355 945 reads with the notable outliers of a grassland in Texas (CLBJ_042, 2256 reads) and tropical forest in Puerto Rico (GUAN_008, 49 112 reads). All other samples returned more than 100 000 reads.

While the BLASTn search returned taxa from 39 phylum-level groups (Supporting information), all further results here focus on three focal phyla: Arthropoda, Mollusca and Annelida (summary statistics in Table 1). These three groups returned 12.2 million total reads that had high quality matches (i.e. ≥ 97% similarity with reference sequences) in the BLASTn search. Arthropoda dominated (10.5 million reads), followed by Annelida (1.1 million reads), and Mollusca (593 166 reads). The 49 site-by-habitat combinations averaged 211 937 HQreads.

While there was no significant relationship between latitude and total read number (best fit linear regression total read number = 216 515 + 3676 × latitude, $r^2 = 0.03$, $p = 0.19$, $n = 49$). High-quality read number increased weakly as a function of latitude (best fit linear regression HQreads = −49 920 + 6904 × latitude, $r^2 = 0.18$, $p < 0.0001$, $n = 49$). The percentage of HQreads from total reads ranged from 11% to 98% and averaged 54%. This percentage was weakly and positively correlated with latitude (%HQreads = 19.9 + 0.8 × latitude, $r^2 = 0.12$, $p = 0.01$, $n = 49$), total image area (%HQreads = 41.0 + 0.001 × image area, $r^2 = 0.15$, $p = 0.005$, $n = 49$) and biovolume (%HQreads = 41.2 + 1.7 × biovolume, $r^2 = 0.08$, $p = 0.037$, $n = 49$) but not abundance ($p = 0.46$). Biovolume was not predictive of the total number of reads nor HQreads (linear model $p = 0.64$ and 0.32, respectively).

We originally assumed that sites with longer trap seasons would have more species due to temporal within-season beta diversity. Our analysis of species accumulation mirrors the overall positive richness gradient. The slope of the species accumulation linear model is significantly and positively correlated with latitude (Slope = 0.02 × Latitude − 0.4, $r^2 = 0.51$, $p = 0.03$, $n = 49$) meaning that the northern sites with shorter seasons accumulate species at higher rates than southern sites with longer seasons.

Latitudinal patterns of taxonomic richness

Taxon-scale results

In the taxon-scale models, with the three focal taxa were combined, all four taxonomic scales (i.e. order, family, genus and species) have positive and significant relationships between

Table 1. Project-wide abundance and taxon richness of the taxa examined here. All statistics tallied from the metagenomics pipeline except number of individuals, which was taken from the imaging pipeline. See Methods text for details.

Taxon	No. of families	No. of genera	No. of species* (seq.)	No. of individuals (images)	Total HQ reads (seq.)	Taxon by site occs. (seq.)
Annelida, Arthropoda and Mollusca	296	850	1424	50 001	12 178 842	5389
Arthropoda	281	828	1391	48 338	10 484 268	5137
Annelida	4	10	18	154	1 101 408	147
Mollusca	11	12	15	1509	593 166	105
Araneae	23	113	201	2755	1 748 831	776
Blattodea	6	6	15	256	1 207 803	136
Coleoptera	35	138	223	4772	1 017 967	637
Collembola	12	19	41	10	718 567	340
Diptera	47	166	318	5425	1 511 075	1273
Hymenoptera	10	40	73	17 919	143 295	178
Isopoda	4	5	6	928	121 153	41
Lepidoptera	31	163	208	218	405 559	455
Myriapoda	6	5	11	2151	350 827	72
Orthoptera	7	32	54	1146	2 766 372	481

taxonomic richness and latitude (Table 2, Fig. 2). The coefficient estimates for the effect of latitude were similar across taxonomic scales, ranging from 0.09 to 0.10. Only the order-scale model had a significant quadratic term, but all four were negative. The latitudinal gradient in species richness did not differ between open and closed sites (ANCOVA $p = 0.88$). The GLMs for the four taxonomic scales varied in only minor details (Table 2) and explained between 58% to 64% of the variance in taxonomic richness. Longitude was significant for species, genera and order levels, indicating higher taxonomic richness in eastern sites.

Cross-taxon results

The GLMs for the 12 common taxa we examined (9 arthropod orders, the subphylum Myriapoda and the phyla Mollusca and Annelida) explained between 15% (Hymenoptera) and 73% (Annelida) of the variance in species richness (Table 3, 4). None of the 12 taxa examined exhibit a significant, negative species richness gradient. Of these 12 taxa examined by metagenomic barcoding, 7 had a significant positive relationship between species richness and latitude (Table 2, Fig. 3) in the GLMs. The remaining 5 taxa show no significant variation across latitude. Only Hymenoptera had a negative coefficient for latitude, but it was not statistically significant, and the overall model fit was poor ($r^2 = 0.15$). Half of the taxa, including Annelida, Blattodea, Isopoda, Lepidoptera, Myriapoda and Orthoptera, had significant negative quadratic terms indicating a unimodal relationship between richness and latitude, but three of these were driven by the inclusion of data from Alaska and Puerto Rico (Supporting information). Longitude was significant and positive for Collembola and Diptera, indicating, on average, higher species richness for these two orders in eastern sites. MAT was significant and positive for three of these groups (Blattodea, Isopoda and Myriapoda) and MAP was significant and positive for the Mollusca.

NEON Sentinel Taxa

Two of the three expert-identified NEON Sentinel Taxa have a positive and significant relationship between latitude and species richness at the more regional scale of ten trap arrays from entire NEON sites (Table 4, Fig. 4). The Carabidae, collected in the same pitfall traps as the bycatch analyzed above, and Ixodida collected by cloth-dragging, both show this negative gradient in species richness as well as a significant and negative quadratic term for latitude, and the model indicates peak species richness around 40–45°N. Longitude was significantly and positively correlated with species richness in the Carabidae, and MAP was significant and positive for the Culicidae.

Alaska and Puerto Rico

The sites in Alaska and Puerto Rico affected the model results but did not reverse the overall pattern. When both regions were removed, latitude was a significant and positive contributor to taxon richness at the levels of families and orders (Supporting information) and not significant at the level of

Table 2. Results of generalized linear taxon-scale models at four taxonomic scales for taxonomic richness of Arthropoda, Annelida and Mollusca combined. 'Long.' is longitude, 'MAT' is mean annual temperature, 'MAP' mean annual precipitation,' HQReads' is the number of high quality reads, 'n' is the number of individuals, 'Q-PDP' is the Quasi-Poisson dispersion parameter estimated by the model. Any non-significant value less than 0.0001 is listed as zero. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	Intercept	Latitude	Latitude^2	Long.	MAT	MAP	HQR	n	Adj. R^2	Q-PDP
Species	2.13	0.09*	−0.0005	0.01*	−0.01	0.0001	0	0	0.64	4.38
Genera	1.92	0.09*	−0.0005	0.01*	−0.01	0.0001	0	0	0.65	3.90
Families	1.35	0.10**	−0.0007	0.01	−0.01	0.0003	0	0	0.68	2.31
Orders	0.08	0.10***	−0.0006*	0.01*	0.02	0.0001	0	0	0.58	0.49

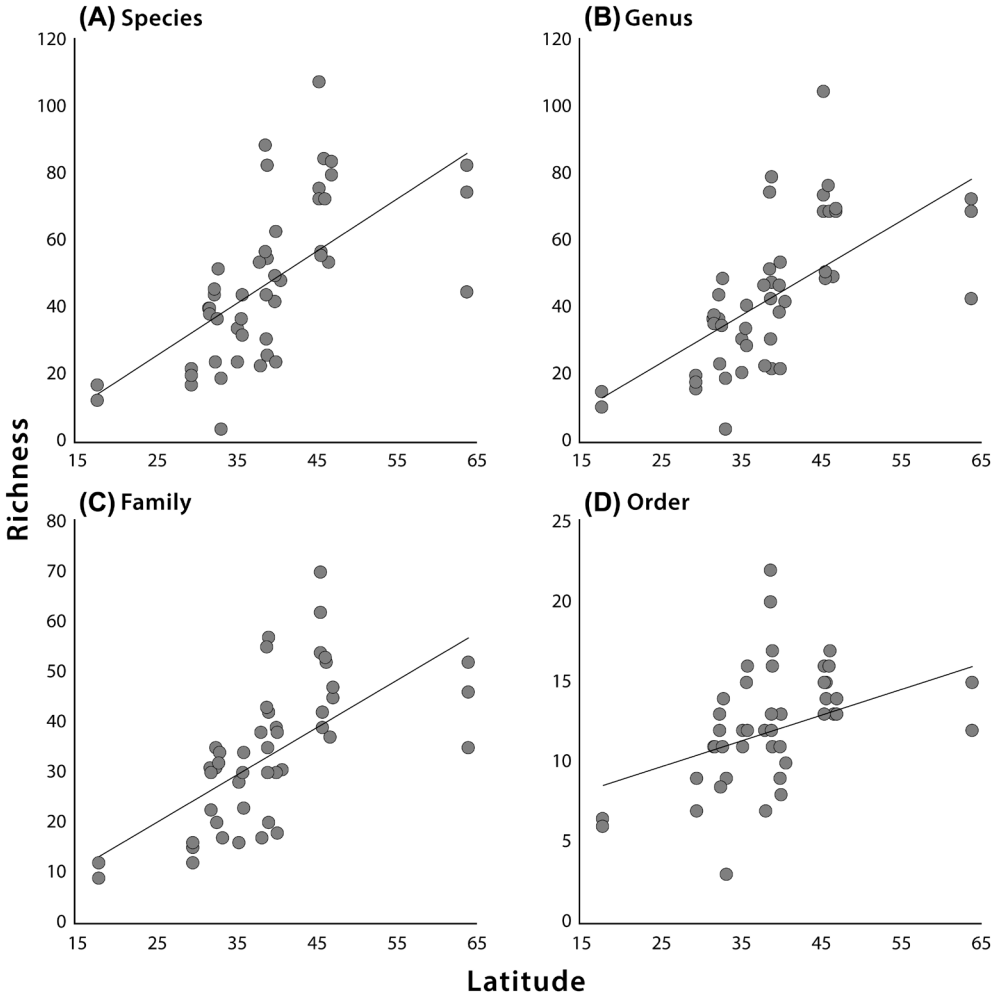

Figure 2. Latitudinal patterns of local taxonomic richness (taxon-scale models) at the scales of species (a), genera (b), families (c) and orders (d) for all focal taxa (Arthropoda, Annelida and Mollusca combined). All four taxonomic scales show a significant positive relationship between latitude and taxon diversity. The line represents the predictions from the generalized linear model (see Table 2 for model details).

species and genera. In the cross-taxon comparisons, the details of the individual models changed (e.g. the positive coefficient for Collembola became significant, the positive coefficient for Coleoptera lost statistical significance, Supporting information) but no taxon's relationship with latitude became significant and negative.

Collinearity

Since we calculated the quadradic term for latitude directly from linear latitude, we calculated the variance inflation factor (VIF) for each model including and excluding the linear latitude. We also ran the GLMs excluding latitude. Across the four taxonomic scales, the mean VIF decreased 75–94% with latitude removed (Supporting information). With latitude excluded, MAT was significant and negative for all but order richness (Supporting information), indicating higher taxonomic richness, on average, at colder temperatures. Removing latitude decreased the mean VIF for the 12 cross-taxon groups from 71 to 97% (Supporting information), and only two (Coleoptera and Diptera) of the twelve taxa showed the significant negative relationship with MAT (Supporting information). The mean VIF for NEON sentinel taxa showed similar decreases with latitude removed (85–90, Supporting information) Both the Mollusca (Supporting information) and Culicidae (Supporting information) retained a significant positive relationship with MAP.

Discussion

Using three types of taxonomic richness data, we show that no invertebrate taxon collected by NEON has a statistically significant and negative latitudinal gradient in taxon richness. Instead, we found that seven of the twelve taxa investigated across three phyla have a significant and positive LGTR. Using the combined results across three phyla we show that this significant positive gradient in community taxon richness is observed in the numbers of species, genera, families and orders. At the more regional scale of entire NEON sites, and with species richness detected in a more traditional manner,

Table 3. Results of cross-taxon generalized linear models of NEON pitfall species richness detected by metagenomics for each of the nine most abundant arthropod orders, the subphylum Myriapoda and the phyla Annelida and Mollusca. Non-significant coefficients between −0.0001 and 0.0001 are listed as 0. Variables are described in Table 2. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Taxon	Intercept	Latitude	Latitude^2	Long.	MAT	MAP	HQReads	n	Adj. R^2	Q-PDP
Annelida	−27.36	1.07***	−0.01**	0.02	0.17	0.001	0	0.03	0.73	0.95
Araneae	−0.24	0.1	−0.0007	0.008	−0.007	0.0002	0.000004**	0.004*	0.50	1.49
Blattodea	−0.17**	0.34**	−0.002*	−0.01	0.24***	0.0004	0	0.02	0.19	0.74
Coleoptera	−2.85	0.25*	−0.002	0.02	−0.003	0.0002	0	0.0003	0.49	2.12
Collembola	−1.04	0.15	−0.0006	0.04**	0.02	0.0001	0	0.0001	0.55	1.07
Diptera	1.52	0.08	−0.0005	0.01*	−0.01	−0.00004	0	0.0001	0.53	1.57
Hymenoptera	0.32	−0.06	0.0002	−0.02	−0.04	0.0004	0	0	0.15	0.98
Isopoda	−8.41**	3.16***	−0.03***	−0.07	0.34*	0.002	0.00006*	0.002	0.42	0.58
Lepidoptera	−1.37	0.15*	−0.001*	0.007	0.005	0	0	0.01**	0.32	1.87
Mollusca	−6.88	0.14	−0.0003	0.003	0.03	0.002**	0	0.003	0.37	0.81
Myriapoda	−44.87	1.32**	−0.01**	−0.03	0.34*	0.03	0.00003*	0.002	0.53	0.92
Orthoptera	−2.56	0.20**	−0.002**	0.006	0.03	0.0002	0	0.004	0.21	1.41

Table 4. Results of generalized linear models of regional (i.e. all traps at a given NEON site from 2016) species richness from expert-identifications of the three NEON Sentinel Taxa. Variables described in * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	Intercept	Latitude	Latitude^2	Long.	MAT	MAP	Adj. R^2	Q-PDP
Carabidae	3.548	0.32***	−0.003**	0.016**	0.05	0.0002	0.48	4.52
Culicidae	1.02	0.01	0.0002	0.001	0.03	0.0008*	0.36	2.59
Ticks	0.19*	0.68***	−0.006**	0.02	0.23**	0.0008	0.44	0.69

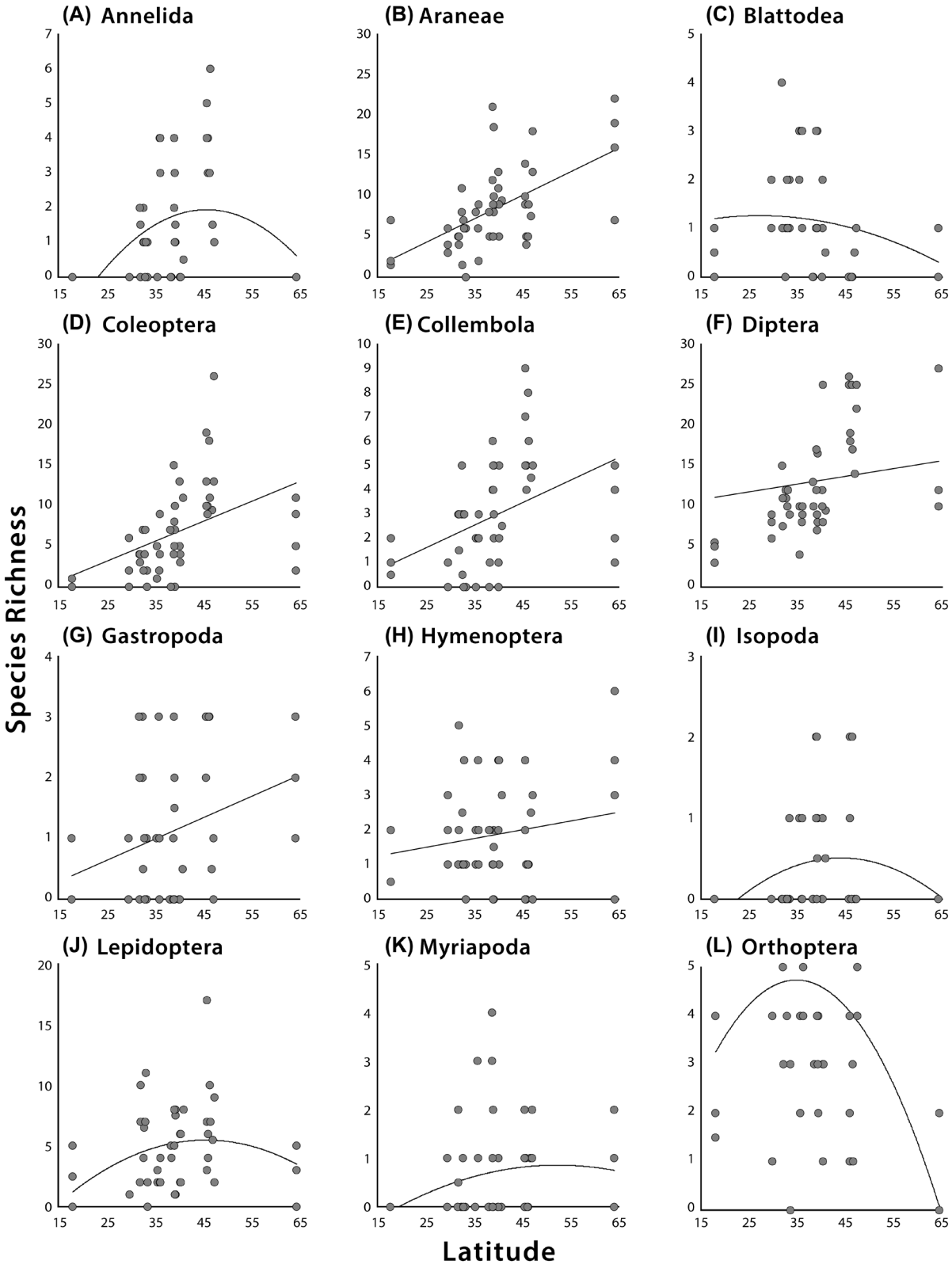

Figure 3. Latitudinal patterns in local species richness for the nine most common orders of arthropods, the subphylum Myriapoda and the phyla Annelida and Mollusca (cross-taxon models). Eight of these twelve taxa show a significant positive relationship between latitude and species diversity. None show a significant negative relationship. The line represents the predictions from the generalized linear models (see Table 3 for model details).

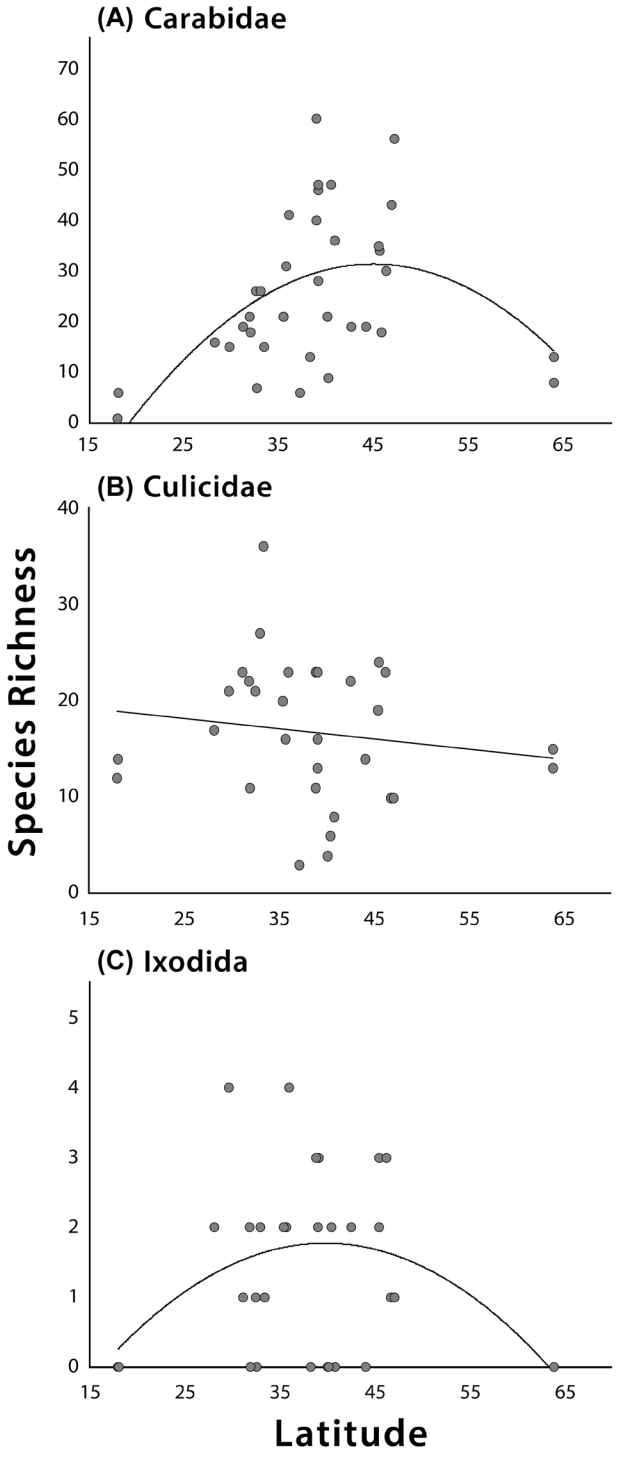

Figure 4. Latitudinal patterns of regional species richness for three NEON Sentinel Taxa identified by taxonomic experts. Ground beetles (Coleoptera: Carabidae) and ticks (Acari: Ixodida) have a significant positive relationship between latitude and species richness. Mosquito (Diptera: Culicidae) richness did not vary significantly with latitude.

two of the three NEON Sentinel Taxa we examined show a significant and positive gradient in species richness.

Positive latitudinal gradients in taxonomic richness indicate north temperate peaks for these data. It is likely that many of these taxa are more diverse in the mainland tropics than anywhere along the gradient we examine. One can assume that at some point in northern Canada and Alaska, all these taxa will drop to zero species (e.g. no terrestrial annelids in permafrost), thus the possibility of a multimodal pattern of richness for those taxa that are also tropical or a unimodal north temperate peak in richness.

The four taxon-scale models were remarkably consistent. Adding a novel higher taxon to a local assemblage automatically adds lower taxa (e.g. adding a novel family adds a novel genus and species), but adding a novel lower taxon does not necessarily add higher taxa (e.g. adding a species from a genus already represented). This consistency across taxonomic ranks, given the hierarchical nature of taxonomy and phylogeny, emphasized the importance of the gain and loss of orders and families to the generation of geographic gradients of diversity.

The cross-taxon models varied. The finding of significant positive LGTR at local scales for some groups is supported by some larger scale studies (e.g. annelids in Phillips et al. 2020). The failure to find a significant gradient for ants (Formicidae) in these data is also supported by the failure to find a significant gradient in local ant species richness for the area studied. Local species richness of ants peaks in Florida, Arizona and North Dakota (Kaspari et al. 2000a, b) while regional species richness for ants shows a clear negative LGTR (Guénard et al. 2012, Jenkins et al. 2013, Economo et al. 2018). It is perhaps notable that the two taxa that show both a positive relationship with precipitation, but not latitude, are dependent on water for reproduction (Culicidae) or moisture to avoid desiccation (Mollusca).

Historically, reverse LGTRs have been notable exceptions to a biogeographical rule (e.g. Hillebrand 2004; parasitic wasps in Janzen 1981, Kouki et al. 1994). There is, however, a growing body of work showing that the exceptions may not be as rare as once thought (e.g. pitcher-plant food webs in Buckley et al. 2003; freshwater arthropods in Morinière et al. 2016; bees in Orr et al. 2020; earthworms in Phillips et al. 2020; snakes in the tribe Lampropeltini in Pyron and Burbrink 2009; ants in Vasconcelos et al. 2018; some plant taxa Weiser et al. 2018). Our results add to this growing body of evidence of positive and unimodal gradients of taxonomic richness outside of the tropics and greatly increase the number of species richness gradients measured at local scales for terrestrial invertebrates.

Failure to find any significant negative gradients in taxonomic richness does not refute the expectation of a negative LGTR but indicates that the underlying drivers of the LGTR do not act continuously across latitudes nor across taxa and taxonomic scales. The tropical conservatism hypothesis (Wiens and Donoghue 2004) argues that relatively few lineages that diversified in the aseasonal tropics have conserved traits that limit their ability to move into and diversify in more seasonal and extreme extra-tropical regions. Pyron and

Burbrink (2009) generalized this hypothesis to apply to conservation of traits for temperate lineages. This biogeographical conservatism hypothesis argues that taxa that diversify in extra-tropical regions would not be expected to have tropical peaks in taxonomic richness. Consistent with a 'temperate' conservatism hypothesis, we show several invertebrate taxa that have a positive and/or unimodal patterns of taxonomic richness peaking temperate North America.

Many, if not most, studies of LGTR are at scales much larger than local species assemblages (Hillebrand 2004). It is possible that there is some other covariate with latitude that limits local taxon richness but not regional (e.g. predation gradients, habitat heterogeneity). It is also likely that some of the patterns and differences in patterns amongst taxa are due to differences in habitats (Phillips et al. 2020). That said, when we look at the more regional scale (10 trap array seasons from entire NEON sites), none of the NEON Sentinel Taxa showed a significant negative LGTR.

Here we use a novel combination of methods on standardized samples to, relatively rapidly, examine community level taxon richness for multiple taxa at and taxonomic scales. While we examine only 12 taxa and three phyla in detail here, the metagenomics effort returned 281 families of arthropods alone. The combination of the spatial extent and standardized methods of NEON sampling and the taxonomic breadth of our methods allows for truly comparative biogeography.

Acknowledgements – The National Ecological Observatory Network is a program sponsored by the National Science Foundation and operated under cooperative agreement by Battelle. This material uses samples collected as part of the NEON Program. All samples will be archived at the NEON Biorepository at the Biodiversity Knowledge Integration Center at Arizona State University. We thank K. Levan and K. Thibault at NEON for their advice and access to the samples used here.
Funding – This work was funded by the National Science Foundation Grant MSB-FRA no. 1702426 grant to MEK (PI), CDS, KEM, MDW and MJM (co-PIs) with additional support to KEM from NSERC RGPIN-2019-04239.
Conflict of interest – All authors have no conflict of interest to report.

Author contributions

Michael D. Weiser: Conceptualization (equal); Data curation (equal); Formal analysis (lead); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Writing – original draft (lead); Writing – review and editing (equal). **Cameron D. Siler**: Conceptualization (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Supervision (equal); Visualization (equal); Writing – review and editing (equal). **Sierra N. Smith**: Investigation (equal); Methodology (equal); Validation (equal). **Katie E. Marshall**: Conceptualization (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Writing – review and editing (equal). **Jessica F. McLaughlin**: Methodology (equal); Validation (equal). **Matthew J. Miller**: Conceptualization (equal); Funding acquisition (equal); Methodology (equal); Resources (equal). **Michael Kaspari**: Conceptualization (equal); Funding acquisition (equal); Investigation (equal); Project administration (equal); Resources (equal); Supervision (equal); Writing – review and editing (equal).

Data availability statement

All data is available directly from the corresponding author. Original images (.CR2 files) and downstream processed images (.tif files) will be archived with the NEON Biorepository/Biodiversity Knowledge Integration Center at Arizona State University. Complete raw sequence reads are available from to the NCBI Sequence Read Archive under the accession PRJNA750744 (<www.ncbi.nlm.nih.gov/bioproject/PRJNA750744>).

Supporting information

The Supporting information associated with this article is available with the online version.

References

Altschul, S. F. et al. 1990. Basic local alignment search tool. – J. Mol. Biol. 215: 403–410.
Barnes, M. A. and Turner, C. R. 2016. The ecology of environmental DNA and implications for conservation genetics. – Conserv. Genet. 17: 17.
Blaire, J. et al. 2020. Robust and simplified machine learning identification of pitfall trap-collected ground beetles at the continental scale. – Ecol. Evol. 10: 1–11.
Buckley, H. L. et al. 2003. Reverse latitudinal trends in species richness of pitcher-plant food webs. – Ecol. Lett. 6: 825–829.
Dixon, A. F. et al. 1987. Why are there so few species of aphids, especially in the tropics. – Am. Nat. 129: 580–592.
Economo, E. P. et al. 2018. Macroecology and macroevolution of the latitudinal diversity gradient in ants. – Nat. Commun. 9: 1778.
Edgar, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. – Bioinformatics 29: 2.
Fox, J. and Weisberg, S. 2011. An R companion to applied regression. – Sage Press.
Guénard, B. et al. 2012. Global models of ant diversity suggest regions where new discoveries are most likely are under disproportionate deforestation threat. – Proc. Natl Acad. Sci. USA 109: 7368–7373.
Hajibabaei, M. et al. 2019. COI metabarcoding primer choice affects richness and recovery of indicator taxa in freshwater systems. – PLoS One 14: e0220953.
Hillebrand, H. 2004. On the generality of the latitudinal diversity gradient. – Am. Nat. 163: 192–211.
Hoekman, D. et al. 2017. Design for ground beetle abundance and diversity sampling within the National Ecological Observatory Network. – Ecosphere 8: 17.
Janzen, D. H. 1981. The peak in North American ichneumonid species richness lies between 38° and 42°N. – Ecology 62: 532–557.
Jenkins, C. N. et al. 2013. Conservation implications of divergent global patterns of ant and vertebrate diversity. – Divers. Distrib. 19: 1084–1092.

Kaspari, M. et al. 2000a. Three energy variables predict ant abundance at a geographic scale. – Proc. R. Soc. B 267: 485–489.
Kaspari, M. et al. 2000b. Energy, density and constraints to species richness: studies of ant assemblages along a productivity gradient. – Am. Nat. 155: 280–293.
Kinlock, N. et al. 2017. Explaining global variation in the latitudinal diversity gradient: meta-analysis confirms known patterns and uncovers new ones. – Global Ecol. Biogeogr. 27: 125–141.
Klymus, K. E. et al. 2017. Metabarcoding of environmental DNA samples to explore the use of uranium mine containment ponds as a water source for wildlife. – Diversity 9: 54.
Kouki, J. et al. 1994. Reversed latitudinal gradient in species richness of sawflies (Hymenoptera, Symphyta). – Ann. Zool. Fenn. 31: 83–88.
Marshall, K. E. and Baltzer, J. L. 2015. Decreased competitive interactions drive a reverse species richness latitudinal gradient in subarctic forests. – Ecology 96: 461–470.
Michener, C. D. 1979. Biogeography of the bees. – Ann. Miss. Bot. Gard. 66: 277–347.
Morinière, J. et al. 2016. Phylogenetic niche conservatism explains an inverse latitudinal diversity gradient in freshwater arthropods. – Sci. Rep. 6: 26340.
Orr, M. C. et al. 2020. Global patterns and drivers of bee distributions. – Curr. Biol. 31: 1–8.
Owen, D. F. and Owen, J. 1974. Species diversity in temperate and tropical Ichneumonidae. – Nature 249: 583–584.
Phillips, H. R. P. et al. 2020. Global distribution of earthworm diversity. – Science 366: 480–485.
Pinkus-Rendón, M. A. et al. 2005. Spider diversity in a tropical habitat gradient in Chiapas, Mexico. – Divers. Distrib. 12: 61–69.
Privet, K. et al. 2020. Relative efficiency of pitfall trapping vs. nocturnal hand collecting in assessing soil-dwelling spider diversity along a structural gradient of Neotropical habitats. – Diversity 12: 1–12.
Pyron, R. A. and Burbrink, F. T. 2009. Can the tropical conservatism hypothesis explain temperate species richness patterns? An inverse latitudinal gradient in the New World snake tribe Lampropeltini. – Global Ecol. Biogeogr. 18: 406–415.
Rohde, K. 1992. Latitudinal gradients in species diversity: the search for the primary cause. – Oikos 65: 514–527.
Rubbmark, O. R. et al. 2018. A broadly applicable COI primer pair and an efficient single-tube amplicon library preparation protocol for metabarcoding. – Ecol. Evol. 8: 12335–12350.
Srivastava, D. S. and Lawton, J. H. 1998. Why more productive sites have more species: an experimental test of theory using tree-hole communities. – Am. Nat. 152: 510–529.
Taberlet, P. et al. (eds) 2018. Environmental DNA for biodiversity research and monitoring. – Oxford Univ. Press.
Thornton, P. E. et al. 2021. Gridded daily weather for North America with comprehensive uncertainty quantification. – Sci. Data 8: 190.
Thorpe, A. S. et al. 2016. Introduction to the sampling designs of the National Ecological Observatory Network Terrestrial Observation System. – Ecosphere 7: e01627.
Torres, J. A. 1984. Diversity and distribution of ant communities in Puerto Rico. – Biotropica 16: 296–303.
Vasconcelos, H. L. et al. 2018. Neotropical savanna ants show a reversed latitudinal gradient of species richness, with climatic drivers reflecting the forest origin of the fauna. – J. Biogeogr. 45: 248–258.
Weiser, M. D. et al. 2018. Taxonomic decomposition of the latitudinal gradient in species diversity of North American floras. – J. Biogeogr. 45: 418–428.
Weiser, M. D. et al. 2021. Batch extraction of morphological an color metrics from invertebrate samples. – <https://dx.doi.org/10.17504/protocols.io.byt4pwqw>.
Wiens, J. J. and Donoghue, M. J. 2004. Historical biogeography, ecology and species richness. – Trends Ecol. Evol. 19: 639–644.
Willig, M. R. and Presley, S. J. 2018. What are latitudinal gradients of biodiversity? – In: Dellasala, D. A. and Goldstein, M. I. (eds), Encyclopedia of the Anthropocene. Elsevier.
Willig, M. R. et al. 2003. Latitudinal gradients of biodiversity: pattern, process, scale and synthesis. – Annu. Rev. Ecol. Evol. Syst. 34: 273–309.