

Title: *Host use diversification during range shifts shapes global variation in Lepidopteran dietary breadth*

Lesley T. Lancaster^{1*}

Nature Ecology and Evolution, 2020

¹University of Aberdeen, School of Biological Sciences, Aberdeen, United Kingdom

*Corresponding author information:

University of Aberdeen

School of Biological Sciences

Tillydrone Ave.

Aberdeen, UK AB24 2TZ

lesleylancaster@abdn.ac.uk

Abstract: Niche breadths tend to be greater at higher latitudes. This pattern is frequently assumed to emerge from the cumulative effects of multiple, independent local adaptation events along latitudinal environmental gradients, although evidence that generalization is more beneficial at higher latitude locations remains equivocal. Here I propose an alternative hypothesis: that latitudinal variation in niche breadths emerges as a non-adaptive consequence of range shift dynamics. Based on analysis of a global dataset comprising more than 6934 globally distributed dietary records from 4410 Lepidopteran species, this hypothesis receives robust support: Population-level dietary niche breadths are better explained by the relative position of the population within its geographic range and the species' poleward range extent than by the latitude of diet observation. Broader diets are observed closer to poleward range limits and in species which have attained higher latitudes. Moreover, latitudinal increase in diet breadth is more prominent within and among species undergoing rapid, contemporary range shifts than for species with more stable ranges. Together these results suggest that latitudinal patterns in niche breadth represent a transient and emergent property of recent geographic range dynamics and

need not require underlying gradients in selective agents or fitness trade-offs. The results have wide-ranging implications for global ecology and for anticipating changes in host use during ongoing distributional shifts of pests and disease vectors.

Main Text:

Populations characterized by broad ecological niches often have greater representation in high-latitude ecosystems than in the tropics¹⁻³. However, because the drivers of latitudinal variation in niche breadth are poorly understood, recognition of this widespread macroecological pattern has limited utility for understanding the processes driving resource specialization, generalization, and niche shifts. Typically, latitudinal clines in niche breadth are thought to reflect an emergent property of multiple, independent local adaptation events along latitudinal ecological gradients. For instance, higher environmental variability and lower species richness of temperate ecosystems in comparison to the tropics have been proposed as ecological explanations for broader niches at higher latitudes^{1,3-6}. In contrast, the role of biogeographic process in shaping global variation in ecological niche breadths has received little prior investigation⁷⁻⁹, despite the fact that causal associations between colonisation events and niche shifts have been well documented¹⁰⁻¹². If biogeographic processes are important drivers of latitudinal trends in niche breadth, distinguishing their effects from those of local adaptation is important for understanding regional variation in ecological interactions, and for predicting novel impacts of range shifting species on natural and agricultural ecosystems and human health^{13,14}.

Here, I investigate potential biogeographic drivers of the global latitudinal cline in dietary niche breadth among insect herbivores^{1,5,6}. I use a new, synthetic global dataset on Lepidopteran dietary niche breadths (Fig. 1) to test whether the cline might reflect widespread, independent responses to Holocene range expansion events^{7,15,16} rather than local adaptation along latitudinal gradients. The dataset comprises 6934 local (site-specific) records of dietary niche breadth, across 4410 Lepidopteran species (Table S1), plus the relative position of the site within each species' geographic range^{17,18}. In addition, n=52 species (Table S2) with known rates of contemporary range expansion^{19,20}, and for which dietary breadth records are available from at

least two latitudes, were used to examine whether the lability of a species' range margin influences its (intraspecific) latitudinal patterns in dietary breadth.

Latitudinal clines in dietary niche breadth may result from changes in levels of host discrimination in marginal populations and during poleward range shifts, if both of the following are true: 1) More recently-established populations are found towards the poleward portion of species' geographic ranges, such that population age tends to be inversely related to latitude, and 2) Colonising and marginal populations exhibit, on average, broader ecological niches than more established populations of the same species, irrespective of underlying ecological conditions. These two combined processes can generate latitudinal variation in niche breadths, without depending on an assumption of a link between latitudinal clines in niche breadth and underlying latitudinal gradients in local ecological conditions— an assumption which has, in any case, received mixed support⁴.

Evidence for the first process, that younger populations tend to lie closer to the poles, is well established, with numerous studies accumulating to indicate widespread poleward movement of species under both postglacial and anthropogenic climate change in both hemispheres during the Holocene and Anthropocene^{15,21–24}. Although high latitude Pleistocene glacial refugia have been identified, which suggest complex patterns of postglacial movement that were not always in a poleward direction²⁵, many studies have shown that the current warming period has characteristically led to genetic and biogeographic signal of more recently expanded populations in poleward than equatorial portions of species' ranges^{15,23,24,26}. Contemporary warming processes have subsequently dramatically increased rates of poleward movement^{22,27}.

Support for the second hypothesis, that younger populations tend to have broader niches, comes from global and intraspecific latitudinal variation in thermal niche breadths^{7,11} and from previous studies in which within-species range extensions were associated with dietary niche breadth expansions^{28–32}. In further support of this hypothesis within Lepidoptera, Singer and Parmesan³³ recently analyzed 30 years of detailed data on dietary niche breadth across Californian populations of Edith's checkerspot butterfly (*Euphydras editha*), and found evidence

of non-adaptive increases in dietary breadth in recently colonized populations, and temporal declines in dietary breadth with increased population age.

Poleward range expansions are occurring generally; if geographic range expansions have caused diet expansion, a multi-population and multi-species analysis should reveal that, after controlling for (including or rejecting) any absolute effects of sampling latitude, there should be significant impacts on diet breadth of the proximity of the study population to the species' poleward range margin. Further, intraspecific latitudinal clines in niche breadth should be most apparent in those species currently undergoing rapid poleward range shifts or which have been able to expand to higher latitudes. Under this process, the net poleward expansion of many species in the Anthropocene and Holocene may therefore in part shape latitudinal variation in latitudinal niche breadth.

Results:

Dietary niche breadth observations in the compiled dataset were best approximated by a truncated, discrete log-normal distribution (Extended Data 1), and the most common strategy was to use a single host in any one location (with $n = 3100$ observations of single-host populations, Extended Data 2). Additional descriptive relationships are presented in the Supplementary Information.

Latitudinal and biogeographical drivers of dietary niche breadth variation among populations:

Local estimates of dietary breadth, estimated as $\log(\text{number of host species})$ within each Lepidopteran population, exhibited a significant, positive latitudinal trend across the 6934 observations (Table 1B), providing a new, independent source of support for the classical latitude-niche breadth hypothesis², a pattern which has also previously been found in multiple, other global datasets^{1,5,6}. However, and in accord with the predictions outlined above, model comparisons revealed that species' range dynamics better predict latitudinal variation in dietary breadth than does sampling latitude itself (Table 1A-C). Populations situated closer to the species' poleward range limit exhibited broader diets than populations nearer to their equatorial

range limit (Table 1A, Fig. 2A), and greater likelihood of local diet generalisation was also found for populations within species that had, as a whole, attained higher latitudes. The effect of latitude itself drops out of the best fit model, suggesting that the variation in diet breadth commonly attributed to latitude is in fact better attributable to placement *within* and *of* the geographic range. This result suggests that the biogeographic processes which give rise to the relative and absolute locations of species' poleward range margins are more important drivers of latitudinal variation in dietary breadth than are latitudinal gradients in the local ecological conditions, such as competitive regimes or resource availability, at sites where dietary breadth was recorded.

Reported results reflect dietary breadth assessed to the plant species level. However, results were qualitatively similar in models in which host use breadths were assessed either to the plant genus or plant family level, suggesting that patterns of host specialization and generalization across the species' geographic range are independent of taxonomic scope of preference functions. Moreover, the phylogenetic niche breadth of Lepidopteran populations, estimated as the summed phylogenetic distance among host plants used by a particular population, is also better predicted by position within the geographic range and maximum range extent than by latitude of observation *per se* (Table S3).

Latitudinal and biogeographical drivers of dietary niche breadth turnover within species:

For species with more than one, geographically-distinct estimate of dietary niche breadth, spatial variation in local host use as a proportion of the species-level dietary niche breadth (i.e., as a proportion of the total number of hosts used by the species across all sampling locations) was also best explained by a model including the relative position of a population within the species' range: populations used a greater proportion of their species-level dietary niche when situated closer to their poleward range margin, in comparison to populations located closer to the equatorial range margin (Table 1D, Fig. 2B). The overall amount of intraspecific geographic variation (i.e., host beta-diversity) in dietary niche breadth also increased with number of sampling locations and the species' geographic range size (Table 1D; the proportional dietary niche breadth within each location is lower, indicating greater host beta-diversity across

observations, for species with more records over a wider area), reflecting the fact that geographic variation in host use within Lepidopterans is common. There was an additional effect of range extent in the best model, such that species which had attained higher latitudes at their poleward range margins were generally less geographically differentiated (lower host beta-diversity) in diet across their ranges than species confined to lower latitudes, a result which is consistent with the proposed mechanism if having attained more polar latitudes at the species level is often generally caused by a series of recent colonization events at the population level, with little time for divergent specialisation to have occurred across the new range (see below).

In contrast, while proportional niche breadth does also increase with latitude (Table 1E), the effect of latitude on intraspecific dietary niche turnover (host beta-diversity) has little explanatory power in comparison to features of the species' geographic range (Table 1D-F), and latitude drops out of the best-fit model in accordance with the hypothesis that range dynamics drive latitudinal variation. While not discounting the importance of local processes, the combined results of these analyses suggest that global variation in dietary niche breadth, assessed either as absolute breadth (previous results section) or as relative changes in host use within species within a species (this section), is better attributed to widespread biogeographic processes and historical colonization events than to patterns of local adaptation along latitudinal ecological gradients.

Latitudinal variation in niche breadth and contemporary range shifts:

The biogeographic legacy of range shifts on niche breadth was also reflected in contemporary range shifts. Species exhibiting more rapid contemporary expansions were more likely to exhibit positive correlations between local dietary niche breadth and latitude (binomial GLMM; interaction effect of latitude \times range shift rate on proportional use of the species' full host range = 0.07 ± 0.03 SE, $z = 2.35$, $P = 0.02$; $n = 70$; Fig. 2C). Estimated range shifting rates were uncorrelated with the latitudes at which host use was observed for each species (see Methods). Thus species which presently exhibit very dynamic ranges are also those for which post-glacial range processes likely have had a greater overall effect on clinal variation in generalism across the species' range^{15,16}, with more recent or rapid colonisation events

generating wider dietary niche breadths towards the poleward range margin than for species exhibiting older or slower colonization events (Fig. 2C).

Discussion:

Implications for range size – niche breadth correlations:

Dietary generalism is often thought to facilitate geographic range shifts, because broader dietary tolerances are presumed to pre-adapt lineages to tolerate novel resources in the new part of the range^{10,34}. However, the evidence for this is mixed. While some studies have found that generalists are more likely to undergo contemporary range shifts³⁵, others have found that dietary specialisation predisposes species to range shifts^{36,37}. Other studies have reported more complex relationships³⁸, or no association at all²⁰. The present data indicate that dietary niche breadth in any particular locale does not covary with species' range shifting rates (Supplementary results, left side of figure 2C) or geographic range size (Table 1A,D; see also³⁴), and therefore suggest that dietary generalism, at least at the population level, is a poor predictor of a species' future capacity for poleward range shifts. In contrast, the present data indicates that host species variety tends to increase towards the poleward range margin irrespective of absolute latitude, and that reduced or diversified host discrimination towards the poleward expansion front is most readily observed in the most recently- or rapidly-expanding lineages (Fig. 2B,C). Together these results strongly suggest that the global trend of increased local dietary niche breadth towards higher latitudes is likely a consequence rather than a cause of geographic range shifts.

Underlying mechanisms:

Evolutionary and ecological transitions between generalization and specialization are commonly labile and bidirectional in insect herbivores³⁹. However, the genetic and physiological mechanisms by which insects switch, expand, or contract their host ranges are poorly known^{40,41}. Singer and Parmesan³³ provide a complementary study to this one, which can provide additional clues about how range shifts result in increased niche breadth. Their analysis of over 30 years of data on preference functions in California populations of Edith's checkerspot butterfly

(*Euphydryas editha*) found diet breadth declined within populations over time but increased after colonization events. Those increases resulted primarily from diversification of host preferences among individuals. Moreover, dietary niche breadth was inversely correlated with site-level genetic diversity. These data suggested that local increases in niche breadth were a consequence of colonization, but that the broad diets of young populations did not reflect admixture of multiple host races. Based on these data and previous work, Singer and Parmesan³³ suggest that transient increases in host breadth following colonization are a consequence of geographic variation within the ancestral host (plant) species: even when butterflies colonizing a new habitat patch use the same host species that they fed on in the source patch, adaptation to the novel host *population* may drive non-adaptive inclusion of a variety of other host *species* into the diet⁴². After repeated colonization events, this dynamic process may ultimately lead to iterative, positive feedback between geographic range expansion and increased dietary breadth⁴³. The data presented here also support this hypothesis, because the reported patterns are independent of the taxonomic level at which host plants are considered: Lepidopterans experience equivalent increases in host plant species, genus or family-level dietary breadth towards their poleward range limits, which provides additional support for the hypothesis that Lepidopteran host preference functions operate primarily at the plant population rather than species (or higher taxonomic) level^{42,44,45}.

Conclusions:

Despite that fact that herbivore host shifts have previously been linked to range dynamics¹⁰, the preponderance of tropical specialists has typically been attributed to underlying site-specific environmental variables which influence trade-offs between generalization and specialization differently within tropical vs. temperate communities⁴⁻⁶. In contrast, while acknowledging the importance of understanding such community dynamics at local scales, the current results raise the possibility that the *global* trend towards increasing niche breadths with latitude¹ may simply result from neutral or nearly-neutral processes⁴⁶ including: easier loss than gain of hosts under stable or ancestral conditions¹⁰, the necessity of accepting novel host genotypes under novel conditions⁴⁷ such as during range shifts, and the hard boundary of minimum requirement (Extended Data 2). Thus while local adaptation cannot be ruled out as an

explanatory factor driving transient increases in niche breadth during range shifts, it is also possible that gains in host plant variety during colonisation, and the subsequent loss of hosts following establishment, may also reflect non-adaptive processes³³. The results presented here clearly indicate that local adaptation is neither necessary nor sufficient to explain latitudinal variation in niche breadth, and that, after considering biogeographic processes that can drive shifts in niche breadth, invocation of latitudinal environmental gradients is likely not required to explain this global pattern.

Materials and Methods

Dataset construction:

To examine how range dynamics drive latitudinal variation in dietary niche breadth, I used a publicly-available dataset of Lepidopteran food plants, which includes 180,000 global Lepidopteran – host plant records for 22,000 species, based on compiled data from 1600 published sources (the HOSTS database¹⁷). I cross-referenced this database with occurrence records in the Global Biodiversity Information Facility (GBIF¹⁸), to obtain information on species' range sizes and the relative range positions of sampling locales. I then combined these data with published information about contemporary Lepidopteran range shift rates^{19,20}. In the HOSTS database, each observed Lepidopteran – host plant species combination is represented once per sampling location, irrespective of underlying frequency of observation in that location, where sampling locations represent political regions largely corresponding to countries (Fig. 1, Extended Data 4; all map figures were created in the *rworldmap* package for R^{48,49}). If a Lepidopteran species was observed in more than one location, it appears multiple times in the dataset, once for each location in which its dietary niche was estimated. Table S1 lists the taxonomic distribution, among superfamilies, of Lepidopterans included in the final dataset.

Local dietary niche breadth was estimated as the number of host species that the Lepidopteran species was observed to use in each location. For Lepidopteran species with host use observed in multiple locations, the species' total niche breadth was estimated as the total number of hosts used over all locales. The species' total niche breadth was used only as a

denominator in analyses of interspecific variation in local niche breadth, and never used as a standalone variable, as it is not ecologically relevant for understanding geographic variation in niche breadths. Moreover, estimates of species-level niche breadth are strongly biased by geographic scope and intensity of sampling effort (see *Spatial Trends in the Data* within Supplementary Information).

The latitudinal geographic range extent for each Lepidopteran species in the dataset was obtained using occurrence records from GBIF¹⁸. Species with fewer than 3 observation records were excluded from the analysis initially, but sensitivity analysis for this cutoff was conducted (see *Effects of spatial resolution and accuracy* in Supplementary Information, and Table S5). Species observations were also excluded if the latitudinal range, calculated using GBIF records, did not include the latitude of the site at which their dietary niche was assessed. Reducing the HOSTS dataset to fit these criteria resulted in 6934 species observations representing 4410 unique species, 1239 species which occurred more than once in the dataset, and 148 countries or other political regions (Fig. 1, Extended Data 4).

The distributional form of dietary niche breadth (number of host species used) over the 6934 observations was estimated using the *powerLaw* package⁵⁰ for R v.3.3.2⁵¹. I tested the data for fit to four candidate distributional forms: discrete power law, discrete log-normal, discrete exponential, and discrete Poisson distributions. Distributions were truncated at the minimum point where the distribution no longer fit the data (x_{\min}), using a standard Kolmogorov-Smirnov approach⁵². The significance of each fit, predicated on the estimated x_{\min} , was estimated using bootstrapping⁵⁰, where H_0 = that the data is generated from the fitted distribution. The number of host species used by each Lepidopteran species in each location was best explained by a discrete, truncated log-normal distribution (Extended Data 1, Extended Data 2) with $\mu = 0.21$, and $\sigma = 1.16$.

Latitudinal and biogeographical drivers:

To test the stated hypotheses, I analysed the data in three, complementary ways. First, I estimated drivers of among-location variation in niche breadth in the *lme4* and *lmerTest*

packages for R^{53,54}. For this, data on local dietary niche breadth (number of host species) were first log-transformed, and then analysed using LMM with a Gaussian-distributed error function. For species with estimates of dietary niche breadth available in more than one location, a separate analysis was run to identify the drivers of intraspecific geographic variation in dietary niche breadth (proportional host use). For the latter test, a binomial GLMM with a logit link structure was applied, where the binomial response variable included the number of host species used by a Lepidopteran species in its particular location, and the total number of host species used by the species over all locations ($y.1$ = the number of host species in the location, $y.2$ = the extent of the total species' host range which was not used in that location).

Explanatory variables used in the above-described regression analyses included: 1) the latitude at which the species' dietary niche was recorded (the HOSTS database uses political boundaries to establish locales for observation, so the midpoint of each country/state was used), 2) the area over which dietary niche breadth was sampled (i.e., the area of the sampling location, in km²), 3) The maximum absolute latitude and latitudinal range extent for each species, 3) the proportional distance between the location where dietary niche breadth was estimated (HOSTS database location) and the absolute maximum latitude of the species' range, within the same hemisphere (resulting in a relative measure of proximity of the population to the species' poleward range margin). Sensitivity of the analysis to differences in spatial resolution of the HOSTS and GBIF sources of spatial data is presented in the Supplementary Information, under *Effects of spatial resolution and accuracy*, Table S5. All continuous variables were transformed to standard deviation units, in order to facilitate comparison of effect sizes. Random effects were also included in all models to account for taxonomic effects: species (because some species were observed in more than one location) nested within family (to control for taxonomic constraints on dietary niche breadth). An additional random effect was fit to account for sampling locale (lat/long; this variable records the midpoint of administrative regions in the HOSTS database and controls for effects of shared or similar recording efforts or sampling programmes within locations; note that effects of sampling or recording protocol could not be modelled directly, as the HOSTS database has already synthesized the primary information regarding Lepidopteran – host relationships within each location). In the analysis of intraspecific variation in (proportional) niche breadth, I further included a fixed-effect term for 4) the number of locations in the dataset

in which the species' dietary niche breadth was estimated (a measure of sampling effort which can affect the value of the denominator of the response variable; see *Spatial Trends in the Data* within Supplementary Information). After checking for correlations among predictor variables, all appropriate combinations of fixed, main effects were tested, and the best-fit model was selected as the one with the lowest value of AICc, calculated using the AICcmodavg package for R⁵⁵. For all reported models (Table 1) correlations among predictor variables are each < 0.55, and variance inflation factors, calculated using the mer-utils package for R⁵⁶, are each < 2. Mean proximities of generalists and specialists to their species' poleward range margins are depicted in Fig. 1A using the ggplot2 package for R⁵⁷. To visualise effects of geographic range position on host range (Fig. 1B), residuals are depicted of a model accounting for effects of geographic range size and number of locations in which the species was observed (Table 1D effects). Extended Data 3 shows the full range of data underpinning the means and standard errors depicted in Figs. 1A & 1B.

Patterns were confirmed to be independent of the taxonomic scope at which Lepidopteran preference functions are assessed, by both (a) testing for drivers of genus- and family-level niche breadth (i.e., spatial turnover in total numbers of host plant genera or families used by each Lepidopteran population, following models as described above), and (b) by analyzing how phylogenetic niche breadth changes with range position and latitude (again following models as described above). For the latter approach, a phylogeny was generated for all n = 8320 plant species in the HOSTS database using the the V.PhyloMaker package for R⁵⁸, based on a combination of the Zanne et al. (2014)⁵⁹ and Smith and Brown (2018)⁶⁰ plant phylogenies. Then I sequentially pruned the tree to reflect the particular set of host species used by each unique Lepidopteran population in the final dataset, and the total phylogenetic distance among host plants used by each Lepidopteran population was calculated from the pruned tree. Due to some hosts not matching to the phylogeny, this reduced the dataset from n=6934 Lepidopteran populations analysed, to n=5317. Resulting phylogenetic niche breadth values were log transformed for further analysis, with models and explanatory variables as described above, and the results are presented in Table S3.

Contemporary range shifts:

To understand the effects of contemporary range dynamics on intraspecific latitudinal variation in niche breadth, I collated a list of $n=69$ Lepidopteran species for which published accounts of poleward range boundary dynamics, specifically in the form of temporal changes in the position of the poleward range margin, were available. Two high-quality studies in particular were used, Pöyry et al.²⁰, which reports changes in poleward range position (in km) of butterflies in Finland between the periods 1992-1996 and 2000-2004, and Mason et al.¹⁹ from which we extracted the reported changes in poleward range positions for butterflies in Great Britain between the periods 1986-1995 and 2001-2010. These time periods are highly similar, facilitating combined use of the data. I then extracted all records from the HOSTS dataset for each of these species. Species from Mason et al. and Pöyry et al. which *also* appeared in the HOSTS dataset at least twice were selected for final analysis ($n = 52$ species, 18 of which occurred in both the Pöyry et al. and the Mason et al. datasets, resulting in 70 species range shift observations overall; Table S2). The relative host range of a population within each locale was estimated as a binomial response variable as in the models described above ($y.1 =$ number of host species in the current locale, $y.2 =$ total number of unique host species used by the Lepidopteran species elsewhere, but not in the current locale). This response variable was analysed in a GLMM with fixed effects for latitude, poleward range shift distance (km), the interaction, and random effects for the locale where dietary breadth was sampled, family, and species nested within family. The interacting predictor variables were statistically independent of each other (correlation of fixed effects in the main model = -0.01; Effect of range shift distance on latitude of host plant use observation = -0.04 ± 0.06 SE, $t = -0.62$, $P = 0.53$ in a linear mixed model including species as a random effect, $n = 70$). For species found in both the Mason et al. and the Pöyry et al. datasets, both range shift estimates were included as predictors in the analysis as repeated measures, with random effects of species and sampling locale included to control for pseudoreplication across datasets. Averaging the range shift estimates between the two datasets resulted in qualitatively similar results. The predicted effect of range shift distance on the relationship between host use and latitude was generated using the sjPlot⁶¹ package for R, as the marginal terms of the interaction, after averaging the other terms on the model (Fig. 2C).

Acknowledgments: Thank you to Mike C. Singer for helpful discussions and comments. Thank you to Ian Kitching for permissions to use the Lepidopteran host plant database curated by the Natural History Museum London. Thanks also to the authors of^{19,20} for making their range shift data publicly available, and to all contributors to and curators of the open access databases used in this study.

Data availability: All data used in this study is freely publicly available from the cited sources^{17,19,20}. Reuse of the HOSTS dataset is conditional upon permission from the curators.

Competing interest statement: The author declares no competing interests.

References and Notes:

1. Forister, M. L. *et al.* The global distribution of diet breadth in insect herbivores. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 442–7 (2015).
2. MacArthur, R. H. *Geographical ecology : patterns in the distribution of species.* (Harper and Row, 1972).
3. Addo-Bediako, A., Chown, S. L. & Gaston, K. J. Thermal tolerance, climatic variability and latitude. *Proc. Biol. Sci.* **267**, 739–45 (2000).
4. Vázquez, D. P. & Stevens, R. D. The latitudinal gradient in niche breadth: concepts and evidence. *Am. Nat.* **164**, E1-19 (2004).
5. Novotny, V. *et al.* Why Are There So Many Species of Herbivorous Insects in Tropical Rainforests? *Science (80-.)*. **313**, (2006).
6. Dyer, L. A. *et al.* Host specificity of Lepidoptera in tropical and temperate forests. *Nature* **448**, 696–699 (2007).
7. Lancaster, L. T. Widespread range expansions shape latitudinal variation in insect thermal limits. *Nat. Clim. Chang.* **6**, 618–621 (2016).
8. Slove, J., Janz, N., Strimmer, K., Midford, P. & Leibowits, T. The Relationship between Diet Breadth and Geographic Range Size in the Butterfly Subfamily Nymphalinae – A Study of Global Scale. *PLoS One* **6**, e16057 (2011).
9. Nylin, S. *et al.* Embracing Colonizations: A New Paradigm for Species Association Dynamics. *Trends Ecol. Evol.* **33**, 4–14 (2018).
10. Janz, N. & Nylin, S. The Oscillation Hypothesis of Host-Plant Range and Speciation. in *Specialization, Speciation, and Radiation* (ed. Tilmon, K.) 203–215 (University of California Press, 2008).

11. Lancaster, L. T., Dudaniec, R. Y., Hansson, B. & Svensson, E. I. Latitudinal shift in thermal niche breadth results from thermal release during a climate-mediated range expansion. *J. Biogeogr.* **42**, 1953–1963 (2015).
12. Losos, Jackman, Larson, Queiroz & Rodriguez-Schettino. Contingency and determinism in replicated adaptive radiations of island lizards. *Science* **279**, 2115–8 (1998).
13. Agosta, S. J., Janz, N. & Brooks, D. R. How specialists can be generalists: resolving the ‘parasite paradox’; and implications for emerging infectious disease. *Zool.* **27**, 151–162 (2010).
14. Lancaster, L. T., Morrison, G. & Fitt, R. N. Life history trade-offs, the intensity of competition, and coexistence in novel and evolving communities under climate change. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, (2017).
15. Hewitt, G. The genetic legacy of the Quaternary ice ages. *Nat. www.nature.com* **405**, (2000).
16. Davis, M. B., Shaw, R. G. & Etterson, J. R. Evolutionary responses to changing climate. *Ecology* **86**, 1704–1714 (2005).
17. Robinson, G. S., Ackery, P. R., Kitching, I. J., Beccaloni, G. W. & Hernández, L. M. HOSTS - A Database of the World’s Lepidopteran Hostplants. Natural History Museum, London. <http://www.nhm.ac.uk/hosts>. *Natural History Museum, London.* (2010). Available at: <http://www.nhm.ac.uk/hosts>. (Accessed: 15th June 2016)
18. GBIF.org. The Global Biodiversity Information Facility. 622,629 georeferenced records Available at: www.gbif.org. (Accessed: 7th August 2016)
19. Mason, S. C. *et al.* Geographical range margins of many taxonomic groups continue to shift polewards. *Biol. J. Linn. Soc.* **115**, 586–597 (2015).
20. Pöyry, J., Luoto, M., Heikkinen, R. K., Kuussaari, M. & Saarinen, K. Species traits explain recent range shifts of Finnish butterflies. *Glob. Chang. Biol.* **15**, 732–743 (2009).
21. Schmitt, T. Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Front. Zool.* **4**, 1–13 (2007).
22. Parmesan, C. *et al.* Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature* **399**, 579–583 (1999).
23. Kerdelhué, C. *et al.* Quaternary history and contemporary patterns in a currently expanding species. *BMC Evol. Biol.* **9**, 220 (2009).
24. de Jong, M. A., Wahlberg, N., van Eijk, M., Brakefield, P. M. & Zwaan, B. J. Mitochondrial DNA Signature for Range-Wide Populations of *Bicyclus anynana* Suggests a Rapid Expansion from Recent Refugia. *PLoS One* **6**, e21385 (2011).
25. Eidesen, P. B. *et al.* Genetic roadmap of the Arctic: Plant dispersal highways, traffic barriers and capitals of diversity. *New Phytol.* **200**, 898–910 (2013).
26. Todisco, V. *et al.* Mitochondrial phylogeography of the Holarctic *Parnassius phoebus* complex supports a recent refugial model for alpine butterflies. *J. Biogeogr.* **39**, 1058–1072 (2012).
27. Chen, I.-C., Hill, J. K., Ohlemüller, R., Roy, D. B. & Thomas, C. D. Rapid range shifts of species associated with high levels of climate warming. *Science* **333**, 1024–1026 (2011).

28. Pateman, R. M., Hill, J. K., Roy, D. B., Fox, R. & Thomas, C. D. Temperature-dependent alterations in host use drive rapid range expansion in a butterfly. *Science* **336**, 1028–30 (2012).
29. Braschler, B. & Hill, J. K. Role of larval host plants in the climate-driven range expansion of the butterfly *Polygonia c-album*. *J. Anim. Ecol.* **76**, 415–423 (2007).
30. Suehiro, W. *et al.* Radiocarbon analysis reveals expanded diet breadth associates with the invasion of a predatory ant. *Sci. Rep.* **7**, 15016 (2017).
31. Shively, R., Barboza, P., Doak, P. & Jung, T. S. Increased diet breadth of little brown bats (*Myotis lucifugus*) at their northern range limit: a multimethod approach. *Can. J. Zool.* **96**, 31–38 (2018).
32. Eloy de Amorim, M. *et al.* Lizards on newly created islands independently and rapidly adapt in morphology and diet. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 8812–8816 (2017).
33. Singer, M. C. & Parmesan, C. Colonizations cause host shifts, diversification of preferences and expansion of butterfly diet breadth. *bioRxiv Prepr.* (2020). doi:<https://doi.org/10.1101/2020.03.31.017830>
34. Slatyer, R. A., Hirst, M. & Sexton, J. P. Niche breadth predicts geographical range size: a general ecological pattern. *Ecol. Lett.* **16**, 1104–1114 (2013).
35. Mattila, N., Kaitala, V., Komonen, A., Päivinen, J. & Kotiaho, J. S. Ecological correlates of distribution change and range shift in butterflies. *Insect Conserv. Divers.* **4**, 239–246 (2011).
36. Fordyce, J. A. & Nice, C. C. Contemporary patterns in a historical context: Phylogenetic history of the pipevine swallowtail, *Battus philenor* (Papilionidae). *Evolution (N. Y.)*. **57**, 1089 (2003).
37. Bridle, J. R., Buckley, J., Bodsworth, E. J. & Thomas, C. D. Evolution on the move: specialization on widespread resources associated with rapid range expansion in response to climate change. *Proc. Biol. Sci.* **281**, 20131800 (2014).
38. Dapporto, L. & Dennis, R. L. H. The generalist–specialist continuum: Testing predictions for distribution and trends in British butterflies. *Biol. Conserv.* **157**, 229–236 (2013).
39. Janz, N., Nyblom, K. & Nylin, S. Evolutionary dynamics of host-plant specialization: A case study of the tribe Nymphalini. *Evolution (N. Y.)*. **55**, 783–796 (2001).
40. de la Paz Celorio-Mancera, M. *et al.* Mechanisms of macroevolution: polyphagous plasticity in butterfly larvae revealed by RNA-Seq. *Mol. Ecol.* **22**, 4884–4895 (2013).
41. Snell-Rood, E. C., Troth, A. & Moczek, A. P. DNA Methylation as a Mechanism of Nutritional Plasticity: Limited Support From Horned Beetles. *J. Exp. Zool. Part B Mol. Dev. Evol.* **320**, 22–34 (2013).
42. Janz, N. & Nylin, S. The role of female search behaviour in determining host plant range in plant feeding insects: a test of the information processing hypothesis. *Proc. R. Soc. B Biol. Sci.* **264**, 701–707 (1997).
43. Jahner, J. P., Bonilla, M. M., Badik, K. J., Shapiro, A. M. & Forister, M. L. Use of exotic hosts by Lepidoptera: Widespread species colonize more novel hosts. *Evolution (N. Y.)*. **65**, 2719–2724 (2011).

44. Singer, M. C., Stefanescu, C. & Pen, I. When random sampling does not work: standard design falsely indicates maladaptive host preferences in a butterfly. *Ecol. Lett.* **5**, 1–6 (2002).
45. Singer, M. C. & Lee, J. R. Discrimination within and between host species by a butterfly: implications for design of preference experiments. *Ecol. Lett.* **3**, 101–105 (2000).
46. Forister, M. L. & Jenkins, S. H. A Neutral Model for the Evolution of Diet Breadth. *Am. Nat.* **190**, E40–E54 (2017).
47. Singer, M. C., Wee, B., Hawkins, S. & Butcher, M. Rapid Natural and Anthropogenic Diet Evolution: Three Examples From Checkerspot Butterflies. in *Specialization, speciation, and radiation: the evolutionary biology of herbivorous insects* 311–324 (Univeresiy of California Press, 2008).
48. South, A. rworldmap: A new R package for mapping global data. *R J.* **3**, 35–43 (2011).
49. Natural Earth. Free vector and raster map data @ [naturalearthdata.com](http://www.naturalearthdata.com). Available at: <http://www.naturalearthdata.com/downloads/110m-cultural-vectors/110m-admin-0-countries/>.
50. Gillespie, C. S. Fitting heavy tailed distributions: the powerLaw package. *J. Stat. Softw.* **64**, 1–16 (2015).
51. R Core development Team. R: A language and environment for statistical computing. (2012).
52. Clauset, A., Shalizi, C. R. & Newman, M. E. J. Power-Law Distributions in Empirical Data. *SIAM Rev.* **51**, 661–703 (2009).
53. Bates, D., Maechler, M., Bolker, B. & Walker, S. lme4: Linear mixed-effects models using Eigen and S4. (2014).
54. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest: Tests for random and fixed effects for linear mixed effects models (lmer objects of lme4package). (2014).
55. Mazerolle, M. J. AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c). (2015).
56. Frank, A. F. mer-utils.R. <https://github.com/aufrank/R-hacks/blob/master/mer-utils.R>. (2014).
57. Wickham, H. *Ggplot2 : elegant graphics for data analysis*. (Springer-Verlag, 2009).
58. Jin, Y. & Qian, H. V. PhyloMaker: an R package that can generate very large phylogenies for vascular plants. *Ecography (Cop.)*. 1353–1359 (2019). doi:10.1111/ecog.04434
59. Zanne, A. E. *et al.* Three keys to the radiation of angiosperms into freezing environments. *Nature* **506**, 89–92 (2014).
60. Smith, S. A. & Brown, J. W. Constructing a broadly inclusive seed plant phylogeny. *Am. J. Bot.* **105**, 302–314 (2018).
61. Lüdecke, D. sjPlot: Data Visualization for Statistics in Social Science, R package version 2.4.0. in *R package version 2.4.0* (2017).

Fig. 1. Distribution of Lepidopteran dietary niche breadths in the dataset. Shading (n = 20 shades) represents the average host range size of Lepidopteran species within each location, where average host ranges were calculated from all n = 6934 observations of 4410 species for each of n = 148 locations. Locations lacking colour are not found within the dataset.

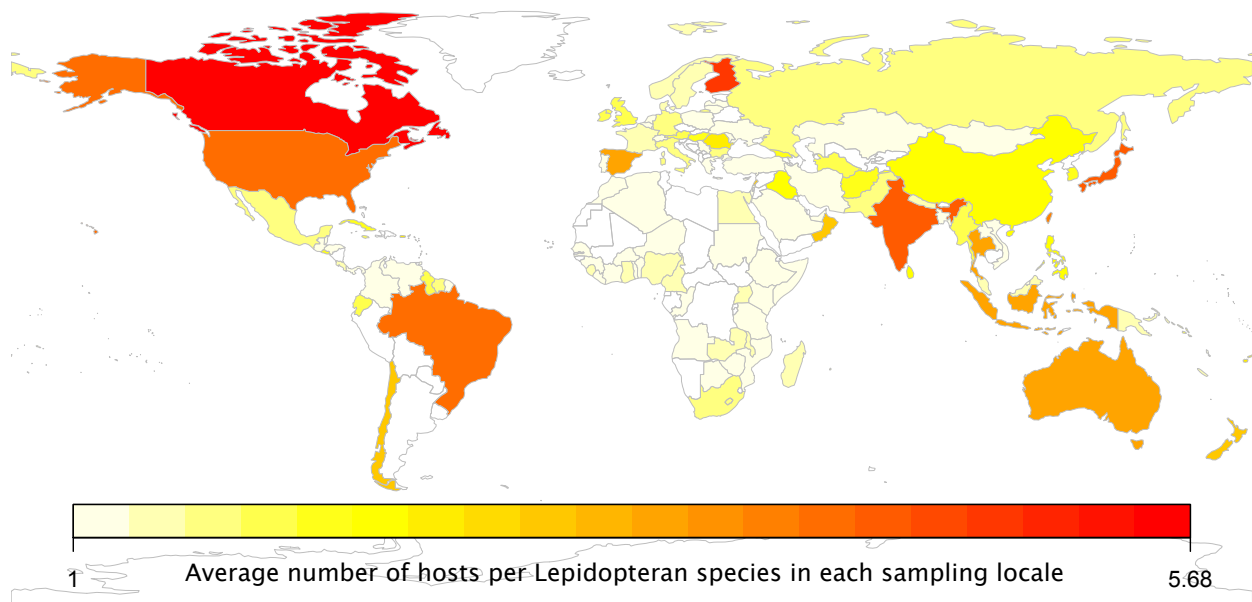


Fig. 2. Effect of relative range position, range dynamics, and latitude on dietary niche breadth in Lepidopterans. (A) Mean (\pm S.E.) relative position of Lepidopteran populations within their species' range as a function of their feeding strategy (specialist = only one host recorded for the population, $n=3100$ populations; generalist = more than 1 host recorded for the population, $n=3834$ populations). Equivalent boxplots are presented in Extended Data 3. (B) For the $n=1239$ species with host range observations in multiple locations (totalling $n=3769$ observations), depicted are the mean (\pm S.E.) residual proportional host use at different proximities to the species' poleward range margin, after correcting for other factors in the model (residuals calculated from the model in Table 1D, omitting the effect of proximity), as a function of proximity to the poleward range margin. Corresponding data points are depicted in Extended Data 3. (C) Latitudinal variation in dietary niche breadth is most apparent for species undergoing fastest rates of contemporary range shifts (solid line), while non-range-shifting species fail to show latitudinal variation in niche breadth (broken line); see text for details of the full models from which prediction curves were extracted. Lines represent the shape of the relationship between latitude and host use, estimated from the interaction surface at maximum and minimum values of range shift rates across species. For all panels, see Methods for plotting details.

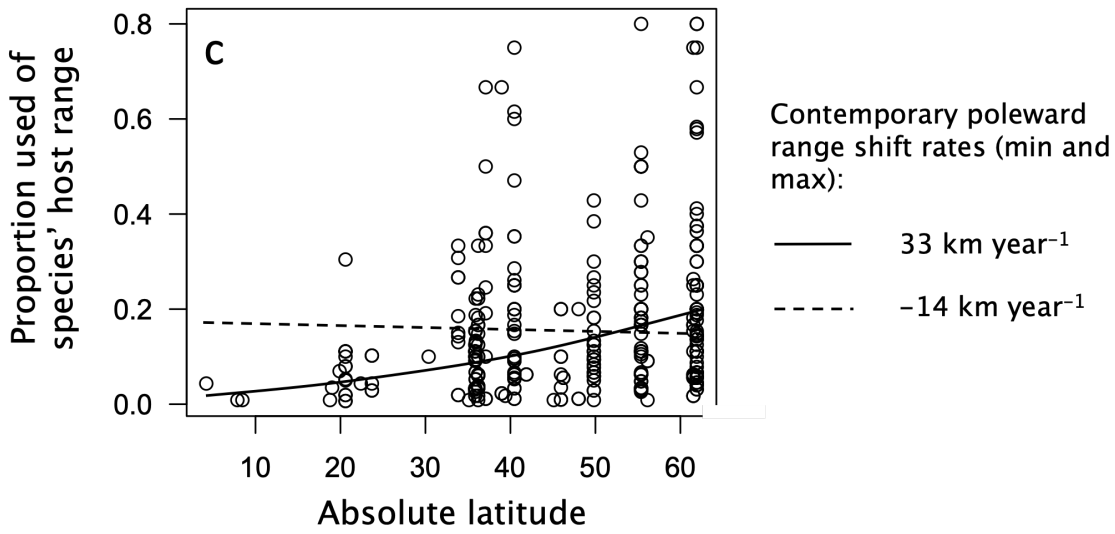
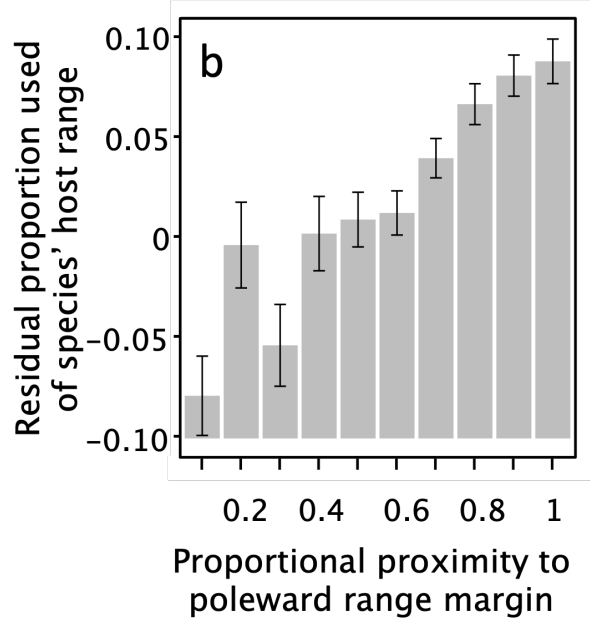
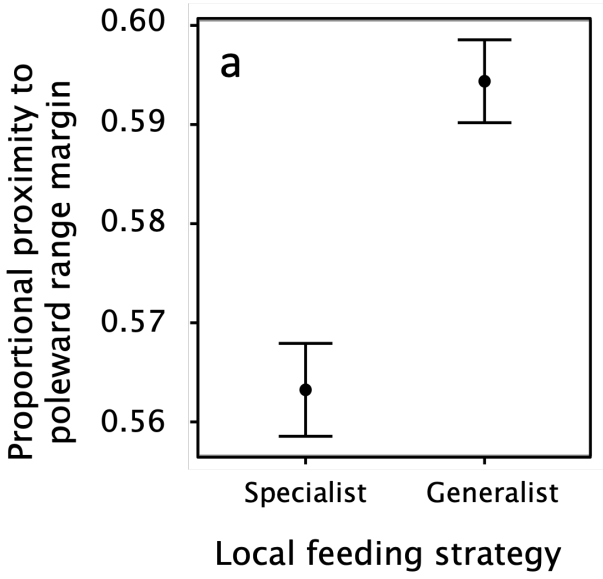


Table 1. Fixed effects in comparisons between models explaining Lepidopteran dietary niche breadth as a function of range position and latitude. Explanatory variables were z-transformed to facilitate comparison. I: Linear mixed model, n = 6934 observations. II: Binomial, general linear mixed model, n = 3769 observations.

I. Global drivers of dietary niche breadth (log[number of host species]):

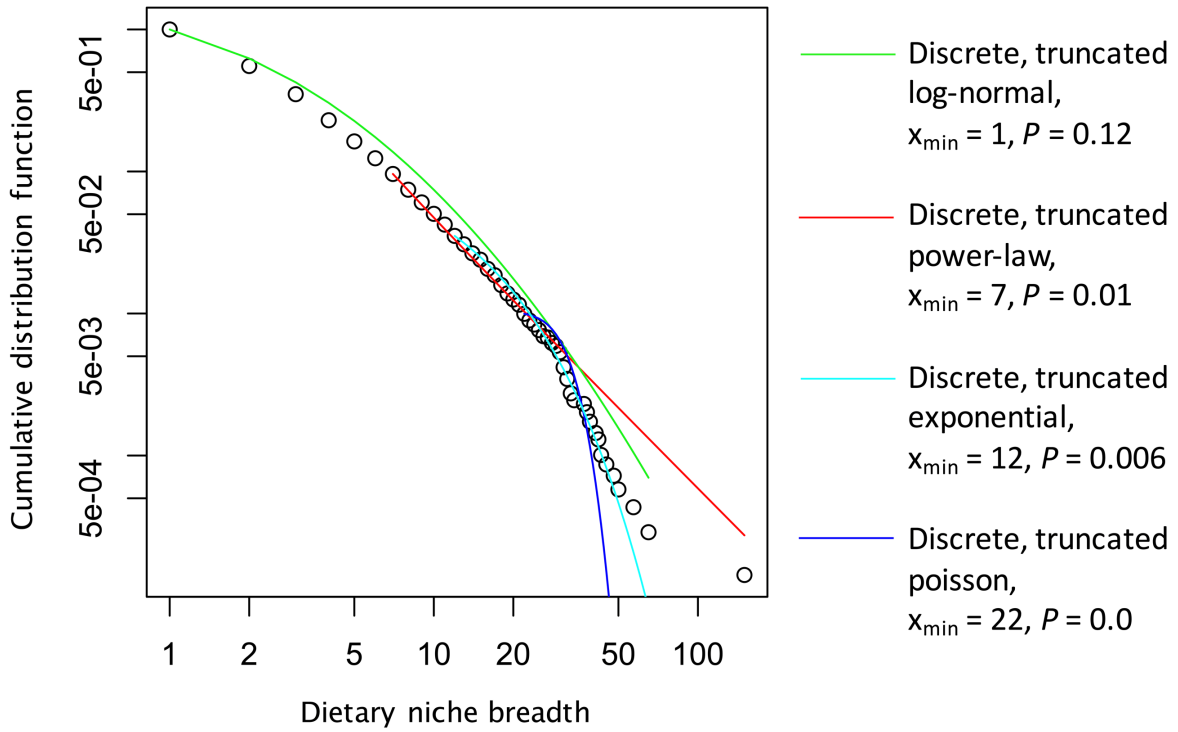
	<u>effect</u>	<u>est</u>	<u>s.e.</u>	<u>z</u>	<u>P</u>	<u>ΔAICc</u>
A) Best fit model:	Proportional proximity to the species' poleward range margin	0.04	0.008	5.03	<<0.0001	0
	Maximum poleward reach of species' range	0.13	0.01	8.97	<<0.0001	
B) Latitude model:	Latitude	0.11	0.03	3.84	0.0002	56.7
C) Full model:	Proportional proximity to the species' poleward range margin	0.03	0.01	2.79	0.005	17.9
	Maximum poleward reach of species' range	0.10	0.02	4.25	<<0.0001	
	Log(latitudinal range extent in degrees)	0.02	0.01	1.90	0.06	
	Latitude	0.02	0.04	0.67	0.50	
	Sampling area	0.04	0.03	1.36	0.18	

II. Drivers of intraspecific spatial patterns (population level niche breadth as a proportion of species-level niche breadth):

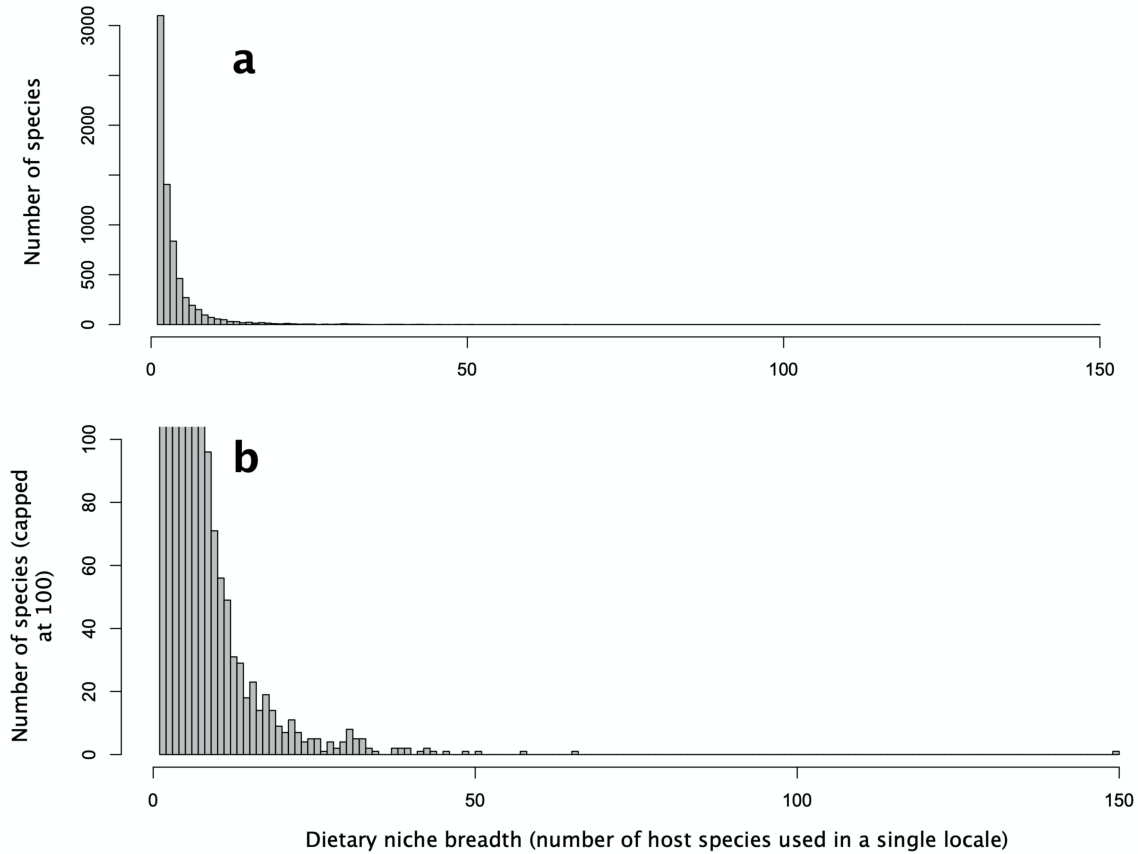
D) Best fit model:	Proportional proximity to the species' poleward range margin	0.20	0.03	8.61	<<0.0001	0
--------------------	--	------	------	------	----------	----------

	Maximum poleward reach of species' range	0.22	0.04	5.15	<<0.0001	
	Log(latitudinal range extent in degrees)	-0.20	0.03	-6.09	<<0.0001	
	Number of locations where dietary niche was observed	-0.60	0.04	-16.99	<<0.0001	
E) Latitude model:	Latitude	0.34	0.09	3.93	<<0.0001	57.9
	Number of locations where dietary niche was observed	-0.61	0.03	-19.78	<<0.0001	
F) Full model:	Proportional proximity to the species' poleward range margin	0.20	0.09	13.20	<<0.0001	2.0
	Maximum poleward reach of species' range	0.20	0.05	4.26	<<0.0001	
	Log(latitudinal range extent in degrees)	-0.19	0.03	-5.49	<<0.0001	
	Sampling area	0.06	0.08	0.72	0.47	
	Latitude	0.10	0.10	1.08	0.28	
	Number of locations where dietary niche was observed	-0.60	0.04	-16.94	<<0.0001	

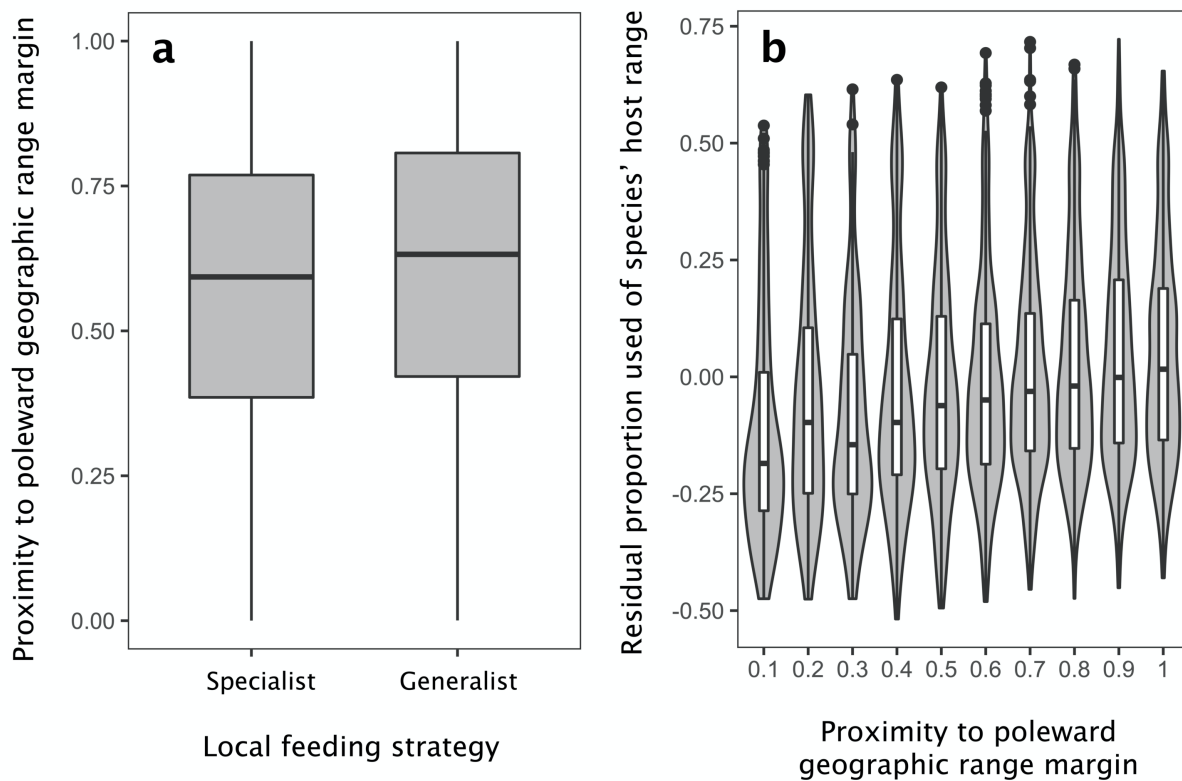
Extended Data Fig. 1. Alternative distributional forms fitted to dietary niche breadths in the dataset. The number of host plants used per Lepidopteran population ($n = 6934$ observations) was best fit by a discrete, truncated lognormal distribution. P -values reflect the null hypothesis that data were sampled from a particular distribution.



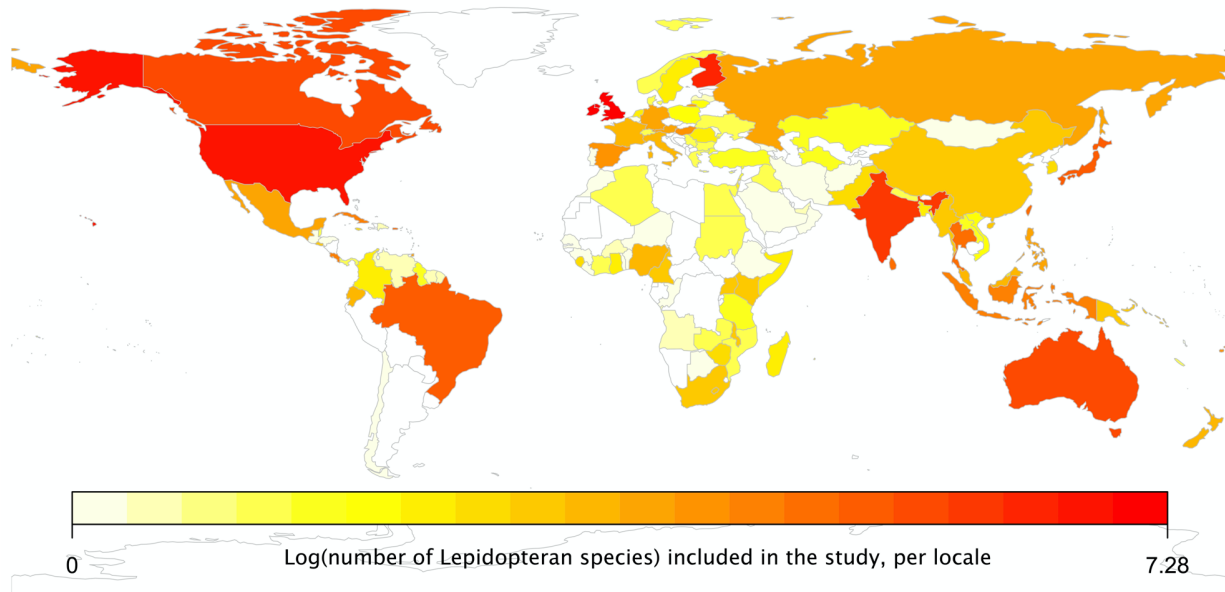
Extended Data Fig 2. Distribution of Lepidopteran dietary niche in the dataset. Histogram of number of host species used (i.e., dietary breadth) per Lepidopteran populations in the dataset (n = 6934 observations). A) The full distribution, B) Inset of the distribution (limited to n=100 records per category), to show the distribution of the tail.



Extended Data Fig 3. Data range corresponding to main text Fig. 2: Relationships between geographic range position and diet breadth. A) Boxplots comparing geographic range positions of generalist and specialist populations of Lepidopterans (specialist = only one host recorded for the population, n= 3100 populations; generalist = more than 1 host recorded for the population, n=3834 populations). Midline = median value; upper and lower limits of box = 3rd and 1st quartile; whiskers are 1.5x interquartile range. B) For the n = 1239 species with host range observations in multiple locations (totalling n = 3769 observations), boxplots represent the median \pm quartiles of residual proportional host use of populations at different distances from the poleward geographic range margins, whiskers are 1.5x the interquartile range, and the full data range is depicted in grey. Residual host range calculated from the model presented in Table 1D, see also Figure 2 legend.



Extended Data Fig. 4. Geographic locations of dietary niche data used in this study. Each location (n=148) appearing in the final dataset is colour-coded by the log-transformed number of unique Lepidopteran species-host plant associations in that location (n = 20 categories). Locations lacking colour are not present in the dataset.



Host use diversification during range shifts shapes global variation in Lepidopteran dietary breadth

Lesley T. Lancaster

Nature Ecology and Evolution, 2020

Supplementary Information:

Spatial trends in the data

Effects of spatial resolution and accuracy

Tables S1-S5

Spatial trends in the data

The number of dietary niche records reported per location in the present dataset did not correlate with absolute latitude, globally (Kendall's $\tau = 0.02$, $z = 0.40$, $P = 0.69$, $n = 148$ locations) or within individual biogeographical regions (Table S4). The number of records reported per location also did not correlate with geographic area of the location (Kendall's $\tau = 0.09$, $z = 1.53$, $P = 0.13$, $n = 148$). These results indicate that, in the present dataset, sampling effort for host plant associations did not suffer from latitudinal or sampling-area biases (Extended Data 4); such positive correlations where present are important sources of potential bias in some previous studies (as discussed in refs.^{1,2}). On average there were 47 ± 153 S.D. dietary records per sampling location. The area of each sampling location also did not correlate with its latitude (Kendall's $\tau = -0.03$, $z = -0.49$, $P = 0.62$, $n = 148$), which could further generate spurious results if such an association were present, because host use becomes more geographically differentiated over larger areas (see main text results). The lack of association between sampling area and latitude in turn implies that the spatial resolution of observations is not latitudinally-dependent (i.e., the distance between successive latitudes of sampling locations in the dataset did not correlate with absolute latitude; Kendall's $\tau = 0.09$, $z = 1.71$, $P = 0.09$, $n = 6934$, $n=146$). Moreover, the latitudinal location of each niche breadth observation was uncorrelated with the maximum latitude achieved by the species, after accounting for the maximum boundary effect (i.e., that a species cannot be observed in the wild beyond its

range limit) (Kendall's $\tau = 0.001$, $z = 0.14$, $P = 0.89$, $n = 6934$), suggesting that dietary observations were random with respect to range position. Global occurrence (GBIF) records³ for Lepidopteran species in the dataset are concentrated in the north-temperate latitudes of the northern hemisphere, suggesting that species' geographic range boundaries and extents are potentially better characterised for species in these locations; however I assessed the effect of range limit accuracy on the results as described below. As expected if species use different hosts across sampling locations, species characterised for their dietary niche breadth in greater numbers of locations were also observed to use greater numbers of unique host species across all locations (among-species correlation for total species-level dietary niche breadth vs. total number of locations in which niche breadth was estimated, Kendall's $\tau = 0.51$, $z = 74.53$, $P \ll 0.001$, $n = 4410$). This latter result strongly cautions against the use of data in which sampling area or intensity is positively correlated with the explanatory variable of interest, because moderate spatial turnover of multiple, specialist host races can create a spurious impression of dietary generalism.

Effects of spatial resolution and accuracy

The dataset includes two distinct sources of location data: range limit location derived from GBIF data and locations of dietary breadth observation⁴, taken from administrative area midpoints, and these sources of data differ in spatial resolution. Range limit locations for all species in the dataset occur on average every $0.045 (\pm 0.26 \text{ S.D.})^\circ$ Latitude, whereas locations of dietary breadth observation occur on average every $0.81 (\pm 1.27 \text{ S.D.})^\circ$ Latitude. I assessed the robustness of the analyses to changes or differences in spatial resolution of these variables by coarsening the resolution of all range limit/extent and location variables to the nearest 1° Latitude, which results in equivalent resolution across data sources, and assessing whether the reported relationships still hold. This was confirmed (Table S4, final column).

I further addressed the robustness of the results to the resolution of range limit locations, as determined by the number of available observation records required to infer the location of a range limit and thus include a species in the analysis. Requiring only a few records to estimate range limits introduces inaccuracy in the estimation of range limit positions, particularly for tropical and S. hemisphere species, for which fewer location records are available overall. However, requiring more location records potentially introduces a bias in the dataset towards primarily N. temperate species. To address this trade-off, sensitivity analyses were run on subsets of the data where >20 , >50 , >100 or >200 observations were required to delineate range limit locations and include the species. The results suggest that the model results are robust to the cutoff used (Table S5). These additional analyses confirm that the findings are robust to variation in the resolution and precision of spatial locations and distances.

Table S1: Taxonomic distribution of dietary niche breadth records in the dataset.

<u>Lepidopteran</u> <u>Superfamily</u>	<u>Number of</u> <u>species</u>	<u>Number of diet</u> <u>observations</u>
Alcitoidea	1	2
Bombycoidea	218	394
Carposinoidea	1	1
Choreutoidea	5	6
Cossoidea	9	13
Drepanoidea	25	35
Epermenioidea	7	7
Eriocranioidea	5	6
Gelechioidea	415	465
Geometroidea	551	797
Gracillarioidea	104	146
Hepialoidea	4	5
Hyblaeoidea	3	12
Incurvarioidea	21	21
Lasiocampoidea	31	39
Mimallonoidea	1	1
Nepticuloidea	45	82
Noctuoidea	1148	1804
Papilionoidea	968	1814
Pterophoroidea	38	56
Pyraloidea	261	517
Schreckensteinoidea	1	3
Sesioidea	26	39
Thyridoidea	3	6
Tineoidea	21	28
Tischerioidea	7	10
Tortricoidea	366	467
Yponomeutoidea	84	109
Zygaenoidea	41	49
Total:	4410	6934

Table S2: Species used to estimate the effects of contemporary range shifts on latitudinal variation in dietary niche breadth. Range shift rates for each species are available in the original reports from Pöyry et al. (2009)⁵ and Mason et al. (2015)⁶.

Species

<i>Anthocharis cardamines</i>	<i>Ochlodes venata</i>
<i>Apatura iris</i>	<i>Oeneis jutta</i>
<i>Aphantopus hyperantus</i>	<i>Papilio machaon</i>
<i>Aporia crataegi</i>	<i>Pararge aegeria</i>
<i>Araschnia levana</i>	<i>Parnassius apollo</i>
<i>Argynnis adippe</i>	<i>Parnassius mnemosyne</i>
<i>Argynnis paphia</i>	<i>Plebejus argus</i>
<i>Aricia artaxerxes</i>	<i>Polygonia c-album</i>
<i>Brenthis ino</i>	<i>Pyrgus malvae</i>
<i>Callophrys rubi</i>	<i>Pyronia tithonus</i>
<i>Carterocephalus palaemon</i>	<i>Quercusia quercus</i>
<i>Celastrina argiolus</i>	<i>Satyrium pruni</i>
<i>Coenonympha glycerion</i>	<i>Satyrium w-album</i>
<i>Coenonympha pamphilus</i>	<i>Thecla betulae</i>
<i>Coenonympha tullia</i>	<i>Thymelicus lineola</i>
<i>Euphydryas aurinia</i>	<i>Thymelicus sylvestris</i>
<i>Euphydryas maturna</i>	<i>Vanessa cardui</i>
<i>Fabriciana niobe</i>	
<i>Glaucopsyche alexis</i>	
<i>Gonepteryx rhamni</i>	
<i>Hesperia comma</i>	
<i>Hipparchia semele</i>	
<i>Inachis io</i>	
<i>Issoria lathonia</i>	
<i>Lasiommata maera</i>	
<i>Lasiommata megera</i>	
<i>Lasiommata petropolitana</i>	
<i>Leptidea sinapis</i>	
<i>Limenitis populi</i>	
<i>Lycaena phlaeas</i>	
<i>Maniola jurtina</i>	
<i>Melanargia galathea</i>	
<i>Melitaea cinxia</i>	
<i>Nymphalis antiopa</i>	
<i>Nymphalis polychloros</i>	

Table S3: Fixed effects in comparisons between models explaining Lepidopteran phylogenetic dietary niche breadth as a function of range position and latitude (full model description provided in the Methods section). Explanatory variables were z-transformed to facilitate comparison. Linear mixed model, n = 5317 observations.

Global drivers of dietary niche breadth (log[total phylogenetic distance among host species]):

	<u>effect</u>	<u>est</u>	<u>s.e.</u>	<u>t</u>	<u>P</u>	<u>ΔAICc</u>
A) Best fit model:	Proportional proximity to the species' poleward range margin	0.17	0.04	3.98	<<0.0001	0
	Maximum poleward reach of species' range	0.50	0.07	6.76	<<0.0001	
B) Latitude model:	Latitude	0.32	0.14	2.35	0.02	37
C) Full model:	Proportional proximity to the species' poleward range margin	0.19	0.05	3.49	0.0005	22
	Maximum poleward reach of species' range	0.56	0.12	4.74	<0.0001	
	Log(latitudinal range extent in degrees)	-0.03	0.06	-0.49	0.63	
	Latitude	-0.18	0.19	-1.03	0.30	
	Sampling area	0.20	0.15	1.40	0.17	

Table S4: Correlations between the number of Lepidopteran – host plant records per location and the absolute latitude of each location, separately within each biogeographic region:

Biogeographic region:	Number of locations:	Kendall's Tau:	<i>P</i>
Nearctic	3	0.33	1.00
Neotropical	24	0.14	0.36
Palearctic	50	0.15	0.12
Afrotropical	31	-0.10	0.43
Indomalayan	29	0.01	0.91
Australasian	10	0.30	0.24

Table S5: Sensitivity of model comparisons to changes in spatial resolution and accuracy. AICc values are compared among competing models (see main text Table 1), under different levels of range limit accuracy (columns 1-5; requiring a greater number of observation records to determine range limits increases the accuracy of range limit locations, but biases the data towards species / regions with more observations available) and resolution of spatial data (column 6, which coarsens the resolution of both HOSTS-derived spatial data and GBIF-derived spatial data to make the spatial resolutions equivalent across these datasets).

Range limit accuracy:	GBIF records >3	GBIF records >20	GBIF records >50	GBIF records >100	GBIF records >200	GBIF records >3
Spatial resolution:	original resolution	original resolution	original resolution	original resolution	original resolution	1 degree resolution

Table 1(I) model comparison (Δ AICc):

A) Best fit model	0	0	0	0	0	0
B) Latitude only model	56.7	13.7	3	4.4	4	72.6
C) Full model	17.9	20.1	19.3	18.4	12.2	16.1
<i>n species included:</i>	4410	3159	2444	1896	1156	4410

Table 1(II) model comparison (Δ AICc):

A) Best fit model	0	0	0	0	0	0
B) Latitude only model	57.9	97.6	94.27	83.3	22.6	48.8
C) Full model	2	3.6	3.8	3.9	0.4	0.3
<i>n species included:</i>	1239	1021	863	727	490	1239

Supplementary References:

1. Ollerton, J. & Cranmer, L. Latitudinal trends in plant-pollinator interactions: are tropical plants more specialised? *Oikos* **98**, 340–350 (2002).
2. Slatyer, R. A., Hirst, M. & Sexton, J. P. Niche breadth predicts geographical range size: a general ecological pattern. *Ecol. Lett.* **16**, 1104–1114 (2013).
3. GBIF.org. The Global Biodiversity Information Facility. 622,629 georeferenced records Available at: www.gbif.org. (Accessed: 7th August 2016)
4. Robinson, G. S., Ackery, P. R., Kitching, I. J., Beccaloni, G. W. & Hernández, L. M. HOSTS - A Database of the World's Lepidopteran Hostplants. *Natural History Museum, London*. (2010). Available at: <http://www.nhm.ac.uk/hosts>. (Accessed: 15th June 2016)
5. Pöyry, J., Luoto, M., Heikkinen, R. K., Kuussaari, M. & Saarinen, K. Species traits explain recent range shifts of Finnish butterflies. *Glob. Chang. Biol.* **15**, 732–743 (2009).
6. Mason, S. C. *et al.* Geographical range margins of many taxonomic groups continue to shift polewards. *Biol. J. Linn. Soc.* **115**, 586–597 (2015).