

# Diversification is correlated with temperature in white and sulfur butterflies

Ana Paula S. Carvalho<sup>\*1</sup>, Hannah L. Owens<sup>2</sup>, Ryan A. St Laurent<sup>3,1</sup>, Chandra Earl<sup>4</sup>, Kelly M. Dexter<sup>1</sup>, Rebecca L. Messcher<sup>1</sup>, Keith R. Willmott<sup>1</sup>, Kwaku Aduse-Poku<sup>5</sup>, Steve C. Collins<sup>6</sup>, Nicholas T. Homziak<sup>1,7</sup>, Sugihiko Hoshizaki<sup>8</sup>, Yu-Feng Hsu<sup>9</sup>, Athulya G. Kizhakke<sup>10</sup>, Krushnamegh Kunte<sup>10</sup>, Dino J. Martins<sup>11,12</sup>, Nicolás O. Mega<sup>13</sup>, Sadaharu Morinaka<sup>14</sup>, Djunijanti Peggie<sup>15</sup>, Helena P. Romanowski<sup>16</sup>, Szabolcs Sáfián<sup>17,6</sup>, Roger Vila<sup>18</sup>, Houshuai Wang<sup>19</sup>, Michael F. Braby<sup>20,21</sup>, Marianne Espeland<sup>22</sup>, Jesse W. Breinholt<sup>23,1</sup>, Naomi E. Pierce<sup>24</sup>, Akito Y. Kawahara<sup>1,7,25,29</sup>, David J. Lohman<sup>26,27,28,29</sup>

\*Corresponding author: Ana Paula S. Carvalho, 3215 Hull Road, Florida Museum of Natural History, Gainesville FL 32611, United States

Email: [apsdecarvalho@gmail.com](mailto:apsdecarvalho@gmail.com)

1. McGuire Center for Lepidoptera and Biodiversity, Florida Museum of Natural History, Gainesville, FL, United States
2. Center for Macroecology, Evolution, and Climate, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark
3. Department of Entomology, Smithsonian Institution, National Museum of Natural History, Washington, DC, United States
4. Department of Natural Sciences, Bernice Pauahi Bishop Museum, Honolulu, HI, United States
5. Department of Life and Earth Sciences, Perimeter College, Georgia State University, Decatur, GA, United States
6. African Butterfly Research Institute, 256 Dagoretti Rd, PO Box 14308 0800, Karen, Nairobi, Kenya
7. Department of Entomology and Nematology, University of Florida, Gainesville, FL, United States
8. Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Bunkyo-ku, Tokyo 113-8657, Japan
9. Department of Life Science, National Taiwan Normal University, 88, Sec 4, Tingzhou Rd, Taipei, Taiwan 116, R.O.C.
10. National Centre for Biological Sciences, Tata Institute of Fundamental Research, GKVK Campus, Bengaluru 560065, India
11. Turkana Basin Institute, Stony Brook University, Stony Brook, NY, United States
12. Insect Committee of Nature Kenya, The East Africa Natural History Society, PO Box 44486, Nairobi 00100, Kenya
13. Programa de Pós-Graduação em Biologia Animal, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves 9500/43435, Porto Alegre, RS 91501-970, Brazil
14. Saitama Study Center, The Open University of Japan, 682-2, Nishikityo, Omiya-ku, Saitama City 330-0853, Japan
15. Museum Zoologi Bogor, Research Center for Biosystematics and Evolution, Research Organization for Life Sciences and Environment, National Research and Innovation Agency, Jl. Raya Jakarta-Bogor Km. 46, Cibinong - Bogor 16911, Indonesia
16. Laboratório de Ecologia de Insetos, Departamento de Zoologia, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS 91501-970, Brazil
17. Institute of Silviculture and Forest Protection, University of Sopron, Bajcsy-Zsilinszky út 4. H9400 Sopron, Hungary
18. Institut de Biologia Evolutiva (CSIC-UPF), 03008 Barcelona, Spain
19. Department of Entomology, College of Plant Protection, South China Agricultural University, Guangzhou 510642, China

20. Division of Ecology and Evolution, Research School of Biology, RN Robertson Building, 46 Sullivans Creek Road, The Australian National University, Acton, ACT 2601, Australia
21. Australian National Insect Collection, National Research Collections Australia, GPO Box 1700, Canberra, ACT 2601, Australia
22. Leibniz Institute for the Analysis of Biodiversity Change, Museum Koenig, 53113 Bonn, Germany
23. Intermountain Healthcare, Intermountain Precision Genomics, St. George, UT, United States
24. Department of Organismic and Evolutionary Biology, Museum of Comparative Zoology, Harvard University, Cambridge, MA, United States
25. Department of Biology, University of Florida, Gainesville, FL, United States
26. Biology Department, City University of New York, New York, NY, United States
27. PhD Program in Biology, Graduate Center, City University of New York, New York, NY, United States
28. Entomology Section, National Museum of Natural History, Manila, Philippines
29. Equal senior contribution

**Keywords:** Cenozoic, Diversity, Lepidoptera, Pieridae, Speciation

## Abstract

Temperature is thought to be a key variable explaining global patterns of species richness. However, to investigate this relationship carefully, it is necessary to study clades with broad geographic ranges that are comprised of species inhabiting diverse biomes with well-characterized species ranges. In the present study, we investigate the link between temperature and diversification in the butterfly family Pieridae (sulfurs and whites) by combining Next Generation sequences and published molecular data with fine-grained distribution information. After building the most comprehensive phylogeny of the group, with almost 600 species and all higher taxa (subfamilies, tribes and subtribes), we found strong support for the following relationships within the family: Dismorphiinae + (Coliadinae + (Pseudopontiinae + Pierinae)). With a curated distribution dataset of over 800,000 occurrences, we conducted multiple comparative phylogenetic analyses that provided strong evidence that species in environments with more stable daily temperatures or with cooler maximum temperatures in the warm seasons have higher diversification rates. We also found a positive correlation between speciation and extinction with paleotemperature: as global temperature decreased through geological time, so did diversification rates. Although many studies demonstrate higher diversity in the tropics, we have been able to identify specific climate variables associated with changes in diversification, while also inferring the most robust and well sampled phylogenomic framework for Pieridae to date.

## Introduction

Understanding factors that affect variation in species diversity across Earth is a central pursuit of ecology (1). While multiple biotic and abiotic factors drive speciation and extinction (2–5), temperature is thought to be a key variable impacting diversity patterns (6–8). Diversity is expected to be greater in warmer environments due to higher evolutionary rates (the evolutionary rate hypothesis (2)). According to this hypothesis, species in warmer areas tend to have higher mutation rates, faster metabolism, and shorter generation times, all of which are directly associated with increased rates of evolution (2, 4, 9–11). Due to the lack of pronounced temperature variation in species-rich areas (e.g., some tropical regions), authors also hypothesize that temperature stability affects diversification (12). This relationship is based on the premise that climatic stability favors the evolution of specialized adaptations (climatic stability hypothesis

(12)). The inherent temperature stability of tropical climates compared to temperate ones is therefore thought to reduce extinction rates.

To test the link between temperature and diversification, it is necessary to study widely distributed clades of species inhabiting diverse biomes that have sufficiently fine-grained locality data. Additionally, a well-supported, dated phylogenetic tree is necessary to account for relationships among species when evaluating patterns of diversity. Although many studies have assessed the correlation between ambient temperature and diversity, these studies have primarily focused on vertebrates (3, 13–19). Considerably less attention has been directed to other, substantially more diverse groups like insects, which—as ectotherms—have life histories more likely to be intrinsically connected to temperature (20). Butterflies are among the best documented insect groups, with nearly 34.5 million locality records on the Global Biodiversity Information Facility (GBIF, as of September 2022).

Studies of butterfly diversity have largely focused on the coevolution with their host plants (21). Prior investigations have found strong evidence for reciprocal adaptations between Pierinae butterflies and their host plants as predicted by an escape-and-radiate coevolutionary scenario (22, 23). However, an extensive study examining evolutionary patterns of host use across all butterflies found weak evidence for coevolution in many clades (24). Host plants are just one factor that might affect insect diversification rates; less attention has been paid to the role of climate.

Butterfly species richness is generally much higher in equatorial regions than temperate regions (25), and is also correlated with ambient temperature (26). Like other ectotherms, temperature strongly influences insect larval development and diapause (27). Ambient temperature and sunlight have direct effects on behaviors such as basking, and indirect effects on life history traits including phenology and voltinism (e.g., 28, 29). Consequently, extreme or highly variable temperatures can negatively impact fitness (30). Temperature mediated diversification is another important context in which to study diversification dynamics in ectothermic organisms, such as butterflies.

The butterfly family Pieridae comprises 1,159 species in 86 genera distributed across the globe (31). Most of their diversity is concentrated in the tropics (32), but Pieridae is known for having species adapted to extreme cold or dry conditions, such as the Arctic (33) and deserts (34). Pierid monophyly is well established (35, 36), but relationships among subfamilies and tribes remain contentious (see 37–39 for clade or region specific studies) (Fig. 1). A comprehensive phylogeny of Pieridae will not only disentangle poorly understood relationships but also illuminate persistent questions regarding the role of thermal ecology in insect macroevolution. Here we construct a comprehensive species-level dated tree of Pieridae with 593 species from 84 genera and infer the evolution of host plant interactions. We combine our phylogenetic reconstruction with climate data extracted using distribution records to examine whether speciation and extinction are correlated with 1) warmer climates (annual mean temperature), 2) seasonality and daily temperature variability (temperature annual range and annual mean diurnal range), 3) extreme temperatures (maximum temperature of the warmest month and minimum temperature in the coldest month), and/or 4) paleotemperatures.

## Results

### Diversity patterns and temperature

Pieridae lineages have steadily increased since the family's origin around 70 Ma ago, according to the deterministic lineage through time plot (dLTT) (Fig. S1). Fitted pulled speciation rates (PSR) gradually decreased from the stem lineage until 40 Ma, when rates started to increase and had two peaks between 20 Ma and the present (Fig. S2).

The BAMM (Bayesian Analysis of Macroevolutionary Mixtures) estimation of net diversification rates identified five clades with increases in diversification rates: *Colias*; a subclade of *Catasticta* (+ *Archonias*); *Delias*; and clades within *Pieris* and *Mylothris* (Fig. 2, Fig. S3). These patterns were consistent with our RevBayes analysis (Fig. S4). The distinct shift configuration of

BAMM suggests a high ( $>0.75$ ) marginal probability of shifts for *Colias*, *Pareronia*, and a subclade of *Catasticta* (+ *Archonias*) (Fig. S3).

Our phylogenetic generalized least squares (PGLS) analyses found significant relationships between temperature and speciation/extinction for two variables: 1) minimum temperature of the coldest month (positive for speciation, negative for extinction), and 2) temperature annual range (negative for speciation, positive for extinction). We also found a significant negative relationship between speciation and mean diurnal range (Fig. S5). The annual mean temperature and maximum temperature of the warmest month did not show significant relationships with either speciation or extinction.

However, when QuaSSE (Quantitative State Speciation and Extinction) was used to fit different likelihood functions of speciation on the tree for different temperature-related variables, the only significant ( $\Delta AIC > 2$ ) variables were: 1) mean diurnal range, and 2) maximum temperature in the warmest month (Dataset S1). We found a negative sigmoidal relationship with drift (Dataset S2) between speciation and these two variables. This means that speciation rates are higher in habitats with low diurnal temperature ranges and with mild summer temperatures, but there is a steep decline in speciation rates when the diurnal range exceeds  $11^{\circ}\text{C}$  and, separately, with summer temperatures are higher than  $31^{\circ}\text{C}$ .

PGLS and QuaSSE both found a negative relationship between speciation and mean diurnal temperature range (Dataset S1). The PGLS analysis of maximum temperature of the warmest month was not significant, but QuaSSE suggests a negative correlation as the best fit for this variable, indicating that extremely high temperatures negatively affect speciation rates. The direction of correlations between speciation and minimum temperature of the coldest month and temperature annual range (which was one of the variables with the highest  $\lambda$  for phylogenetic signal) differed between the two methods. However, PGLS P-values were significant while the QuaSSE  $\Delta AIC$  was less than 2. These results suggest that milder temperatures in the coldest month and less variable temperature throughout the year are correlated with higher speciation.

The discrepancies between PGLS and QuaSSE are likely an algorithmic artifact. The PGLS used tip rate estimations extracted from BAMM, while QuaSSE evaluated correlations using its own lambda and mu estimates. Additionally, QuaSSE fits the data to five different models while PGLS is limited to a linear regression, which may not be the best model for the data.

The best fit time-dependent and (paleo)temperature-dependent model explaining Pieridae diversification dynamics was a positive exponential relationship for both speciation and extinction rates and (paleo)temperature (BEnv.VarDEnv.Var\_EXPO) (Dataset S3). Global temperature, in general, underwent a negative trend during the ~40 my of Pieridae crown group evolution. Thus, the exponential positive relationships of both speciation and extinction rates with paleoclimate means these rates also decreased (slowed) with time towards the present. The best model fit for the Pierinae tree describes a pattern in which speciation rate was constant whereas extinction rate decreased with paleotemperature (BCSTDEnv.Var\_EXPO), meaning that net diversification remained positive in the group.

## Comparative phylogenetics

An ancestral state reconstruction (ASR) identified Fabales as the ancestral host plant order of Pieridae as well as the subfamilies Dismorphiinae and Coliadiinae (Fig. S6). Well supported (posterior probability [pp]  $> 0.75$ ) host plant switches were found in the ancestors of Pseudopontiinae (to Santalales) and Pierinae (to Brassicales). A host switch to Santalales was also detected in a clade within Pierinae that includes species-rich groups including *Mylothris*, *Delias*, and *Catasticta*.

Pagel's  $\lambda$  scores suggest significant phylogenetic signal from all WorldClim climatic variables tested: mean diurnal range, maximum temperature in the warmest month, minimum temperature in the coldest month, temperature annual range and, especially, for annual mean temperature (Dataset S4). Conversely, the *Blomberg's K* analyses were only significant for

annual mean temperature and highest temperature in the warmest month, both with low P-values ( $<0.06$ ) (Dataset S4).

An ASR of annual mean temperature suggests that the ancestor of Pieridae was likely found in warm climates; pronounced switches to colder, temperate climates occurred multiple times, such as in the genera *Colias* and *Pieris* (Fig. S7). The reconstruction of minimum temperature in the coldest month shows prominent habitat shifts to colder temperatures in *Colias* (Fig. S8). There was a clear shift towards habitats with greater annual temperature ranges (seasonality) in *Aporia*, *Colias*, and *Pieris* (Fig. S9). *Delias*, the most species-rich genus of butterflies, is often associated with warmer habitats, but many species are found in environments with cooler temperatures (Fig. S7) including tropical and subtropical montane habitats. Reconstructions of mean diurnal range (Fig. S10) and maximum temperature in the warmest month (Fig. S11) did not show pronounced variation across the tree.

### Pieridae phylogeny

We inferred the most taxonomically comprehensive phylogeny of Pieridae to date (Fig. 3, Fig. S12). We included genetic data for 593 pierid species and 15 outgroups, which amounted to a six-fold increase in taxon sampling compared to the most recent phylogeny of the family (36). Eighty-four of the 86 genera recognized in Pieridae by Lamas (31) are represented; the two missing genera, *Calopieris* and *Piercolias* are rarely collected, and the validity of the latter has been recently questioned (40). Subfamilial relationships were strongly supported (UFBoot and aLRT  $>95\%$ ) in general, but support for the position of Pseudopontiinae, which we recovered as the sister group to Pierinae, was low (UFBoot = 75.2, SHaLRT = 77). This corroborates the results of Kawahara et al. (24) and Espeland et al. (41), based on 391 loci, but not of Chazot et al. (42), which used 9 loci (Fig. 1). Recently, Zhang et al. (40) used phylogenomic analyses to validate previous studies indicating the non-monophyly of some taxa. They suggested that *Catasticta* should be a junior synonym of *Archonias*, and that *Eurema* is paraphyletic, among other issues. In regards to *Catasticta* and *Archonias*, for example, phylogenetic, morphological, and natural history traits had already suggested that these genera are closely related to each other (38, 43, 44). However, sampling in Zhang et al. (40) was limited in some clades, and our study highlights additional issues. Namely, *Euchloe* is paraphyletic, *Rhabdodryas* is nested within *Phoebis*, *Glennia* is nested in *Ganyra*, and *Eurema* is still polyphyletic even after taxonomic changes by Zhang et al. (40).

Divergence time estimates varied depending on the calibration scheme, with calibrations taken from Kawahara et al. (45) resulting in younger dates (Dataset S5). The crown age of Pieridae was between 69.6 – 64.5 Ma with a stem age between 83.5 – 76.1 Ma. Subfamily crown ages were estimated to be: 35.0 – 32.7 Ma for Dismorphiinae; 49.8 – 44.2 Ma for Coliadinae; and 49.9 – 48.9 Ma for Pierinae (Dataset S5). The five Afrotropical Pseudopontiinae species are on a long branch with extant taxa diversifying only 5.1 – 3.8 Ma (Dataset S5).

## Discussion

### Diversity patterns and temperature

Based on multiple analyses, our study aimed to isolate different aspects of ambient temperature and test how these could individually be correlated with rates of speciation and extinction. Our results support the hypothesis of climatic stability only when considering the range of temperature throughout the day, but not through the year (seasonality). We found evidence that contradicts the evolutionary rate hypothesis since speciation rates are lower for species inhabiting areas with high summer temperatures.

Seasonality influences species' distributions (46). In general, diversity is greater in the tropics, and these areas are often characterized by high temperatures that are generally more constant throughout the year (47). In warm seasons, sunlight and ambient heat are the main factors influencing butterfly activity, and if conditions are not ideal, butterflies will shift the timing and length of their activity (48). The diversification of white and yellow butterflies is higher in these



environments with stable circadian temperatures. However, based on our analyses, when considering seasonality (as a function of temperature annual range), we found results which do not support the hypothesis of climatic stability (12).

Warmer environments are linked to life history traits that accelerate speciation (9–11); however, our results provide evidence that in temperate environments, speciation rates are elevated. Butterflies have an optimal internal temperature range for survival, but can tolerate greater environmental ranges because they have some ability to control their thermal physiology (28, 49, 50). Nonetheless, anything below or above the tolerable range can be lethal (28). The optimal temperature range usually trends towards warmer temperatures, and most species are active during warm seasons. In some cases, the warmest month is the only time of the year when butterflies can reproduce. This is especially true for Arctic species. However, extremely warm temperatures can reduce activity in some species (48). Recognizing such trends is crucial considering global temperature changes due to climate change, particularly in light of our evidence of speciation rate decreasing with increasing temperature.

Franzén et al. (48) found that the survival of the palearctic papilionid *Parnassius apollo* is negatively affected by increased median temperatures and that there is a tendency for stabilizing selection with temperature in the lycaenid *Phengaris arion* (48). In general, if temperatures are extreme, behavior and morphology/wing coloration will adapt to regulate body temperature (51–54). Butterflies use different strategies and mechanisms to maintain body temperature, which can disrupt their regular behavior (50). They may, for example, find a shaded area to hide (28), which then prevents them from feeding and finding mates.

We found that paleotemperatures partially explain current pierid diversification patterns. Contemporary temperatures and species distributions provide only a snapshot of how temperature impacts diversity and have limited ability to explain how diversity was shaped throughout millennia. Condamine et al. (18) found that (paleo)temperature-dependent models explain diversity across a wide breadth of taxonomic groups. We found that global cooling during the Cenozoic correlates with decreases in both speciation and extinction rates across all of Pieridae. This co-variance is likely a result of high rate heterogeneity across the tree as made evident by lineage specific birth-death analyses conducted herein. However, our investigations of diversification dynamics of the most species-rich subfamily, Pierinae, show that net diversification has increased since the Cenozoic, as a result of decreasing extinction rates combined with constant speciation. Future research could further clarify these dynamics by examining smaller clades within Pieridae but with complete or nearly complete species-level sampling.

In many taxa, temperature at least indirectly correlates with diversity (6, 8, 18, 55–57). Condamine et al. (25) found evidence to support the diversification rate hypothesis in swallowtail butterflies (Papilionidae) by comparing tropical and temperate species. However, these authors caution against concluding that one trait could be responsible for explaining diversity because it could be confounded with others. In butterflies, host plant use has been frequently suggested to play an important role in diversification through coevolution (21). Wheat and colleagues found evidence in Pieridae for the signature of escalating arms races via the coupled evolution of plant chemical defenses (glucosinolates) and detoxification mechanisms (nitrite-specifier proteins) (22), as well as other key innovations involving chemical defenses derived from gene and genome duplications (23). A genus-level phylogenetic investigation of all butterflies by Kawahara et al. (24) found limited evidence that host plant switches affect diversification rates. Our exploratory ancestral state reconstruction of plant orders did not uncover any striking host shifts that coincided with diversification rate shifts, but we did not conduct any direct analyses to test for effects of host plants on diversification. Further work will be necessary to evaluate the prevalence of coevolutionary arms races along the plant/ herbivore interface in generating organic diversity.

### Comparative phylogenetics

We used Pagel's  $\lambda$  and Blomberg's  $K$  to evaluate the phylogenetic correlation of each WorldClim variable. The discrepancies between  $\lambda$  and  $K$  highlight the importance of assessing differences between these two methods. Pagel's  $\lambda$  indicated phylogenetic signal consistent with a Brownian motion model for all variables. This was not the case for  $K$  (Dataset S4), which is a ratio

of variation within clades compared to overall tree trait variation. For some of the traits that we analyzed, such as maximum temperature of the warmest month (Fig. S11), variation throughout the tree is low, but the presence of closely related taxa with very different trait values, especially on short branches (e.g., in *Pontia* + *Reliquia*) pulls *K* values down to near zero. Thus, for our dataset, *Blomberg's K* gives the impression that phylogenetic signal is low, despite trait values being largely conserved across the tree (Fig. S11).

The general trend we observe in the Pieridae phylogeny is a uniform distribution of taxa in warm regions with mild winters (Fig. S7, Fig. S8), which is expected considering Pieridae are most diverse in warm climates. However, there are multiple independent origins of cold-adapted species (e.g., *Baltia* spp. and *Reliquia santamarta*) and clades that are sister to taxa living in warm or mild climates (such as the clade that includes *Tatochila*, *Hyposchila*, *Theochila*, *Pierphulia*, *Phulia*, and *Infraphulia*). Some of these shifts seem to be associated with adaptations to habitats at higher elevations, as shown by the convergence of morphology among groups that inhabit similarly cold, high-elevation habitats around the world (35). The monotypic genus *Reliquia*, endemic to a single mountain range in tropical Colombia, was recently found, as well as in our study, to be nested in the mostly Holarctic *Pontia* (including *Baltia*) (40). *Pontia* inhabits colder environs at similarly high elevations in the Northern Hemisphere. This observation partially uncovers the intriguing interplay of Pieridae biogeography and adaptation to high elevations around the world. Interestingly, the high-altitude, narrowly endemic monotypic Riodinidae genera *Corrachia* and *Styx*, occurring only in mountains in Costa Rica and Peru, respectively, are also nested within an Old World clade (58).

### Phylogenetic relationships of Pieridae

This study provides a robust phylogenetic framework of pierid butterflies and a foundation for testing many macroevolutionary questions. We show, with strong support (UFBoot and SH-aLRT = 100), that Dismorphiinae are sister to a clade containing Coliadinae, Pseudopontiinae, and Pierinae, in agreement with Wahlberg et al. (36) and Kawahara et al. (24). All tribal relationships of Pierinae were strongly supported and differ from previous studies (24, 36, 42). All of the Pieridae sequence data used by Kawahara et al. (24) are incorporated into our dataset, but the inferred trees differ in tribal topology (Fig. 3, Fig. S12). This is likely a result of our considerably increased taxon sampling.

We recovered the subfamily Pseudopontiinae on an extremely long branch (as shown previously in Mitter et al. (59)) of approximately 50 my in length (Dataset S5), which likely contributed to the difficulty of placing the subfamily, as well as inferring ancestral states and diversification patterns in this group. This morphologically unique Afrotropical subfamily was thought to only include a single species until 2011 (59) when molecular data demonstrated the existence of four additional cryptic species. Even though we incorporated additional molecular data for several species in the group, support for its position in relation to the other subfamilies is still low (as in Espeland et al. (41) and Chazot et al. (42)). Furthermore, several genera, like *Euchloe* and *Ganyra*, are not monophyletic (Fig. 2), and will require further taxonomic investigations.

The crown age of Pieridae of around 70 Ma overlaps the K-Pg (Cretaceous–Paleogene) boundary, a major extinction event. The lineage giving rise to the family pre-dated this event, and pierids started diversifying during the Paleogene (Fig. S1).

### Conclusions

Our study provides a robust phylogenetic framework for Pieridae and will invigorate future studies on butterfly ecology, evolution, and conservation. We demonstrate that diversity patterns in pierid butterflies are linked to current and paleoclimates, and that rates of both speciation and extinction slowed as paleotemperatures decreased during the Cenozoic.

### Materials and Methods

## Molecular data

### 1. Locus sampling

We obtained sequence data in 3 different ways. These were: 1) the anchored hybrid enrichment (AHE) BUTTERFLY2.0 kit (60), which targets 13 genes, 2) AHE BUTTERFLY1.0 (41), which targets 425 genes, and 3) publicly available sequence data. Details of these approaches are outlined below.

#### 1.a Anchored Hybrid Enrichment

We sequenced DNA from 338 Pieridae species and two outgroups sourced from various alcoholic and dry insect collections (Dataset S6) using methods described in Kawahara et al. (24). The BUTTERFLY2.0 AHE probe set captured up to 13 loci commonly used in Sanger sequencing studies of past decades (60, 61). Raw sequencing reads underwent filtering and quality checks for base call quality (phred > 20) and length (> 30 bases) using Trim Galore! v.0.40 (62); and assembled via a custom Python v.2.7.6 script that performs an iterative baited assembly (IBA) (63) for each locus. IBA uses similarity searching in usearch (64) to assemble loci beginning with a database of reference loci from a closely related species (in this case *Danaus plexippus*) using bridger (65). Second, assembled loci were filtered by blasting the original probe regions of the assembly against the *Danaus plexippus* reference genome (66) followed by mapping single hits to check for orthology. We screened the orthologous loci for sequence contamination and removed any sequences with 99% similarity to species in another family. The resulting sequences were aligned with MAFFT v.7.0.1 (67). Different copies of DNA were collapsed into consensus sequences using FASconCAT-G v.1.0.4 (68).

We augmented the dataset with 103 pierid species sequenced for Kawahara et al. (24) with the BUTTERFLY1.0 (Dataset S6) probe set (41) which have up to 425 AHE loci (including the same 13 loci in BUTTERFLY2.0).

#### 1.b Published sequences

We complemented the AHE molecular dataset with sequences from 134 pierid species in GenBank downloaded using the software GeneDumper v.0.8 (<https://github.com/sunray1/genedumper>) (Dataset S6). We used the 13 loci captured by BUTTERFLY2.0 from *Colias erate* as a reference sequence for BLAST to find other available Pieridae sequences in GenBank. GeneDumper can integrate taxonomic name updates in the GenBank search and thus requires a synonym list to determine name validity. We used the species list of Lamas (31). After the original BLAST, GeneDumper filters and checks the sequence hits. When multiple sequences for the same locus of a species are available, GeneDumper considers the characteristics of the data and selects the longest sequence with the most unambiguous DNA content. The identification of the selected sequences are verified via a self-blast against GenBank's nucleotide collection. A tiling approach is also used, where sequences are tiled along the bait sequence used in the original BLAST and the least number of tiles providing the highest coverage across the baits are chosen. Therefore, one species may be represented by sequences from multiple specimens if there are no sequences available that fully cover the bait sequence (Further details of the method can be found at <https://github.com/sunray1/genedumper>).

The sequences of Warren-Gash et al. (39) were made available on GenBank after we used GeneDumper, so we downloaded those sequences from the database for species not already in our molecular dataset. We also incorporated sequences from other targeted studies (38, 41), and used genomes and transcriptomes assembled to the BUTTERFLY1.0 probe kit in Kawahara et al. (24) (including two as outgroups) (Dataset S6).

There were 15 species for which we created chimeras by adding cytochrome oxidase subunit 1 (COI) from one sample to another with more loci available but no COI (see Dataset S6 for more details).

#### 1.c Combining molecular datasets



All genetic sequences were then combined to create the molecular data matrix. We undertook this process by first combining data from the two AHE datasets. The 13 loci of BUTTERFLY2.0 are included (as loci numbers 1 through 13) in BUTTERFLY1.0 (41) and simplify concatenation and combination of datasets. We then manually added GenBank sequences to this genetic dataset, and subsequently aligned each locus with MAFFT following the gene alignment protocol of Breinholt et al. (63). After alignment, each locus was examined in AliView v.1.27 (69) to check for indels, and determine reading frames. Loci were concatenated with FASconCAT-G v.1.0.4 (68) to create a concatenated dataset of 425 loci.

The final dataset had fifteen outgroups (Dataset S6) with at least two species from each of the six other butterfly families. The taxonomy of butterfly names follows Lamas (31).

## 2. Partitioning, tree inference, and branch support

The best nucleotide partition schemes were identified using ModelFinder as implemented in IQ-TREE v.2.0.3 (70, 71). For this analysis, we allowed partitions to be identified and merged with '-m TESTNEWMERGEONLY' to limit overparameterization. In order to reduce computation times, we used the rcluster method (set to 10, with a maximum of 1000) and limited the model set to GTR. We consolidated the original 425 locus partitions into 69 partitions. Separately, we ran ModelFinder to identify the best models of nucleotide evolution for our partitions using all available models. With these 69 partitions and models, we then performed 100 independent IQ-TREE tree searches with 1000 Ultrafast bootstraps (UFBoot) and 1000 Shimodaira-Hasegawa approximate likelihood ratio test (SH-aLRT) replicates to test for branch support. We implemented the '-bnni' command when running these analyses to alleviate concerns about model violation inherent in the Ultrafast bootstrap method (72). All phylogenetic analyses were run on the University of Florida HiPerGator2 Cluster ([www.rc.ufl.edu/services/hipergator](http://www.rc.ufl.edu/services/hipergator)).

## 3. Divergence time estimation

We used two different secondary calibration schemes obtained from the trees of Espeland et al. (41) and Kawahara et al. (45) to estimate pierid divergence times (Dataset S7). Ages, based on 95% confidence intervals in these studies, were treated as hard minimum/maximum ages and we employed the penalized likelihood tree dating with TreePL (73). TreePL requires an input tree, for which we used the most likely tree (highest likelihood of 100 runs) inferred with IQ-TREE. TreePL outputs divergence times as single dates per node. In order to provide a range of dates for each divergence event, we performed TreePL independently 100 times and summarized results, following the approach of St Laurent et al. (74). This method uses several custom Python scripts (available at <https://github.com/sunray1/treepl>) and creates 100 bootstrap replicates generated in RAXML (75) using the 425-locus data matrix, partitioned according to the same 69 partitions inferred with ModelFinder. These 100 bootstrap alignments were then used to infer branch lengths on the constrained input phylogeny. We generated 100 bootstrap trees, utilizing optimization parameters that were determined with the 'prime' command and random subsample and replicate cross-validation with the 'randommcv' command for identifying the best smoothing parameter. This optimization and smoothing step was performed three times on each tree to identify the best combination of parameters and assess consistency in the smoothing parameter. All analyses used the 'thorough' command in TreePL. These analyses resulted in 100 dated trees varying only in branch lengths, which were summarized as a maximum clade credibility (MCC) tree in TreeAnnotator in the BEAST package (76). The MCC tree was used in all downstream analyses that required a dated phylogeny.

In addition to making use of secondary calibrations in TreePL, we conducted an analysis using fossil calibrations as the minimum ages. To calibrate the root, a median age of Angiosperms (139.4 Ma) from Magallón et al. (77) was used (as done by Espeland et al. (41)). In total, eight fossil calibrations were applied to the following places in the tree (fossil names and their age noted in parentheses): 1) the Parnassiinae stem age (*Thaïtes ruminiana*, 23 Ma); 2) the Hesperidae stem (*Protocoeliades kristenseni*, 54 Ma); 3) an internal Hesperidae crown age (*Pamphilites abdita*, 23 Ma); 4) Nymphidiini (Riodinidae) stem (*Theope* sp., 15 Ma); 5) the crown node of Satyrinae + Heliconiinae (Nymphalidae) (*Vanessa amerindica*, 33.7 Ma); 6) *Aporia*

(Pieridae) crown age (*Aporia* cf. *crataegi*, 2.6 Ma); 7) *Belenois* (Pieridae) crown age (*Belenois crawshayi*, 0.02 Ma); 8) split between Coliadinae and (Pseudopontiinae + Pierinae) (*Vanessa pluto*, 16 Ma). Each of these fossils were selected because they have been vetted as reasonable fossils to include in calibration analyses (78). TreePL was run on this dataset as noted above, with the exception that the root was a hard maximum age and all fossils were hard minimum ages.

For the purpose of our discussion, we focus on estimated dates obtained using calibrations from Kawahara et al. (45), because this is the most comprehensive, fossil-calibrated dated phylogeny of Lepidoptera to date.

### Distribution and bioclimatic data

We used the R library occCite v.0.4.9 (79) to assemble distribution data. It provides a platform to compile georeferenced data from global databases while also generating a list of primary data sources for each record from institutional collections. We extracted occurrence data using the command 'occQuery' and followed the taxonomy of Pieridae according to Lamas (31). This command checks species names against the Global Names Index (gni.globalnames.org), then uses rgbif (80) to send a query to GBIF (81), requesting all records with geographic coordinates. We removed records that did not have decimal places to increase accuracy and removed records with identical coordinates to the first four decimal places because they were too precise to provide meaningful additional information. Original occurrence data sources with digital object identifiers (DOIs) were gathered using 'occCitation' (Dataset S8). We also extracted additional occurrence data from the literature and directly from GBIF for 37 species that were not obtained using the method described above (references also available in Dataset S8).

We mapped all records by species in an atlas using the package ggplot2 v.3.3.5 (82), with a custom R script (<https://github.com/hannahlowens/PieridaeDiversity>). The atlas was inspected for occurrence points that appeared as outliers or errors based on published distribution maps and opinions of Pieridae experts (authors and collaborators - see Acknowledgements) and we subsequently removed these putatively erroneous records. We also combined the records of the cryptic species *Leptidea sinapis*, *L. juvenica* and *L. reali* as the "sinapis complex" due to difficulty in delimiting distribution in these species (83). The final dataset had over 800,000 records for 541 pierid species, representing 91% of species in the phylogeny (Dataset S9).

We extracted bioclimatic data for the geographic locations of these records from the WorldClim dataset (2.5 arc-minute resolution; (84)) using the package raster v.3.4-10 (85). Since climatic stability can refer to different timescales including paleoclimates (86), we henceforth refer to temperature stability in regards to seasonality (annual variation) and daily range. Therefore, considering our hypotheses regarding the roles of extreme conditions and seasonality, we focused on the following variables: annual mean temperature (average of BIO1), mean diurnal temperature range (average of BIO2), maximum temperature in the warmest month (95% quantile of BIO5), minimum temperature in the coldest month (5% quantile BIO6), and temperature annual range (average of BIO7). The averages (BIO1, BIO2, BIO7), or 5% (BIO6) or 95% (BIO5) quantiles were calculated for each species that had at least one record. The data are summarized in Dataset S9.

### Host plant data

Larval host records were compiled from over 200 sources including books, websites, existing databases, and journal articles (SI Appendix), resulting in over 13,150 unique records for 347 of the 593 species in our tree. Data were captured at the taxonomic resolution provided in the original source. Many of our sources document the same butterfly-host interactions, and such duplicate records with different citations were retained, as we regarded multiple records as an indication of reliability.

Butterfly taxonomy was standardized to Lamas (31) and plant taxonomy was standardized to The World Flora (87) using taxotools v.0.0.1 (88). To provide the clearest picture of host use evolution, we elected to analyze patterns of pierid species host use at the level of

plant orders rather than plant families, genera, or species. This seemed justified because many pierid species feed on Brassicaceae and Capparaceae (both Brassicales) or Santalaceae and Loranthaceae (both Santalales), a pattern noted by Ehrlich and Raven (21).

Many extensive databases of butterfly host plant records aim to aggregate all known records without scrutinizing their veracity (89, 90). In an attempt to quality-control records from the literature, we calculated the “ordinal proportion,” which is the proportion of records for a given butterfly species found in a given plant order. We scrutinized records for which the ordinal proportion was less than 20%, usually by reviewing the original references and ascertaining whether the plant order had been observed as a host plant of that butterfly species more than once. Several unusual host association records were discarded during this vetting procedure. Even if these records are accurate, they most likely document uncommon host taxa, which are sometimes observed during population outbreaks but do not represent the most important ecological patterns (91, 92). The dataset used for the analysis is provided in Dataset S10.

### Diversification and correlation analyses

We conducted a BAMM (Bayesian Analysis of Macroevolutionary Mixtures) analysis to infer diversification rates on the pierid phylogeny. The input topology was the TreePL tree, dated with secondary calibrations from Kawahara et al. (45), excluding outgroups. We estimated prior parameters using the R package BAMMtools v.2.1.6 (93) with the command ‘setBAMMpriors’ (lambdaInitPrior ~ 3.33; lambdaShiftPrior ~ 0.01; muInitPrior ~ 3.33). We ran a reversible-jump MCMC for 100 million generations, sampling every 100,000 generations with six different expected shifts (5, 10, 15, 20, 25, and 30). We used a fraction file to account for taxon sampling bias across the tree (Dataset S11). BAMM output files were analyzed in BAMMtools with a 10% burn-in. We plotted shift posterior probabilities for all six analyses and determined that they converge around 20 shifts (Fig. S13). Thus, the result based on 20 shifts was used for diversity analyses described below.

We also conducted a lineage-specific birth-death-shift (LSBDS) analysis in RevBayes (94) with an MCMC chain of 10,000 steps; following the parameters employed in Höhna et al. (95). Rate variables were set to lognormal distributions for both mean speciation and extinction rates, the rate of rate-shift events was set to uniform, and six rate categories were set as a global parameter. LSBDS does not have an option for fine level sampling bias, and therefore a global sampling fraction of 51% of 1,159 described species of Pieridae was applied.

Considering recent criticisms of estimating diversification rates using phylogenies of extant taxa (e.g., 96), we also used the castor v.1.5.7 R package (97), to estimate pulled speciation rates (PSR) and deterministic lineage through time plots (dLTT). To ensure time slices are biased more towards the present where extant time trees are expected to be more reliable, we first used the function ‘fit\_hbd\_psr\_on\_grid’ with arguments ‘Ngrid’ set to 15, and ‘Ntrials’ to 100 in order to get density values for X and Y axes. These density values were used as arguments in the castor function ‘castor:::get\_inhomogeneous\_grid\_1D’ to create an inhomogeneous ‘age\_grid’ which could then be used as an argument in the function ‘fit\_hbd\_psr\_on\_grid’ to calculate pulled speciation rate (PSR) and dLTTs on the inhomogeneous grid.

We performed comparative phylogenetic analyses using Phytools v.0.7-90 (98) in R. To estimate the degree of phylogenetic conservatism present for each of the five WorldClim BIO climatic variables, we calculated both *Pagel's λ* (99) and *Blomberg's K* (100) using ‘phylosig’.

A maximum likelihood-based ancestral state reconstruction (ASR) was conducted in Phytools using the ‘contMap’ function (‘anc.ML’ model) of all temperature-related variables on the dated Pieridae tree without outgroups. We also conducted a host plant ancestral state reconstruction using stochastic character mapping in *Phytools* with the ‘make.simmap’ command. For this ASR we combined five plant orders (Picrostiales, Zygophyllales, Gentianales, Sapindales, Saxifragales) that were recorded as hosts for only one species each into one state (category “OTHER”) to facilitate resolving state history. Without this addition, we were unable to resolve most ancestral feeding associations. We used a symmetrical model (‘SYM’) and 1000 simulations (nsim=1000). The presence of a state was coded as 1, and the absence as 0. When

there were missing data, all tested states were coded with equal prior probability. If a species was known for feeding on more than one order, we assigned equal probability for those states and 0 for all others (Dataset S11).

To assess a possible correlation between temperature and rates of diversification, we calculated phylogenetic generalized least squares (PGLS) in the R package nlme v.3.1-149 (101). Speciation and extinction rates were extracted from the results from the BAMM analysis using the 'getTipRates' command in BAMMtools for all species in the phylogeny. With the command 'gls', we set the correlation function to Brownian and method to 'ML'. We performed ANOVA tests on resultant PGLS models to check for significance and identify the relationship of diversification rates and variable (slopes) as positive or negative.

As another means of testing for the relationship between temperature and speciation rates, we used Quantitative State Speciation and Extinction (QuaSSE) as implemented in *diversitree* v.0.9-15 (102). We used a pruned tree including only species with available trait data (541 species). Our methods largely followed Zhang et al. (103), whereby we tested increasingly complex QuaSSE models to create likelihood functions where speciation is constant, linear, sigmoidal, or hump-shaped, with and without a drift parameter. The Akaike information criterion (AIC) was used to identify the best fit QuaSSE model for each of the BIO variables.

Climatic BIO variables used in the analysis above stem from present day temperature. To evaluate how global temperature changes through geological time affected diversity in the family as well as its most diverse subfamily Pierinae, we fitted four temperature-dependent models to our phylogeny in the R. We also considered the possibility of time influencing diversity by fitting six time-dependent models. To run these analyses we used the pipeline and custom R scripts developed by Condamine et al. (104). Specifically, time-dependent models were derived from Morlon et al. (105) and Stadler (106) and the paleoclimate models from Condamine et al. (107). The original script provided by the authors included linear and exponential models, but due to criticism directed toward linear diversification dependency models by Gamisch (108) (but also see Morlon et al. (109)), we commented out code for linear models. The time-dependent models were fitted with maximum likelihood using the 'fit\_bd' function and the paleoclimate models used the 'fit\_env' function, both functions available in the pipeline of Condamine et al. (104). Sampling fractions of 0.5 for Pieridae and 0.44 for Pierinae were assigned to account for missing data on the phylogenies.

## Acknowledgments

David Plotkin provided suggestions on phylogenetic dating, Caroline Storer helped molecular data assembly, and Martin Wiemers helped curate distribution data. Ana B.B. Morais and Yutaka Inayoshi provided samples. Fabien Condamine made suggestions on the time- and paleotemperature-dependent analyses. Stilianos Louca assisted with the inhomogenous time grid in castor. Liam Revell provided the code for plotting temperature bars in Figure 2. We thank the NCBS Sequencing Facility for probeset sequencing of an Indian sample. We acknowledge Research Computing at the University of Florida for providing computational resources (<https://www.rc.ufl.edu>).

## References

1. K. J. Gaston, Global patterns in biodiversity. *Nature* **405**, 220–227 (2000).
2. K. Rohde, Latitudinal gradients in species diversity: the search for the primary cause. *Oikos* **65**, 514–527 (1992).
3. B. A. Hawkins, E. E. Porter, J. A. Felizola Diniz-Filho, Productivity and history as predictors of the latitudinal diversity gradient of terrestrial birds. *Ecology* **84**, 1608–1623 (2003).

4. D. J. Currie, *et al.*, Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecol. Lett.* **7**, 1121–1134 (2004).
5. D. Schluter, Speciation, ecological opportunity, and latitude. *Am. Nat.* **187**, 1–18 (2016).
6. N. J. Sanders, J.-P. Lessard, M. C. Fitzpatrick, R. R. Dunn, Temperature, but not productivity or geometry, predicts elevational diversity gradients in ants across spatial grains. *Glob. Ecol. Biogeogr.* **16**, 640–649 (2007).
7. Q.-G. Zhang, H.-S. Lu, A. Buckling, Temperature drives diversification in a model adaptive radiation. *Proc. R. Soc. B* **285**, 20181515 (2018).
8. F. O. Cerezer, *et al.*, Latitudinal gradient of termite diversity indicates higher diversification and narrower thermal niches in the tropics. *Glob. Ecol. Biogeogr.* **29**, 1967–1977 (2020).
9. A. P. Allen, J. H. Brown, J. F. Gillooly, Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Science* **297**, 1545–1548 (2002).
10. J. F. Gillooly, A. P. Allen, G. B. West, J. H. Brown, The rate of DNA evolution: Effects of body size and temperature on the molecular clock. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 140–145 (2005).
11. S. Wright, J. Keeling, L. Gillman, The road from Santa Rosalia: a faster tempo of evolution in tropical climates. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 7718–7722 (2006).
12. P. H. Klopfer, Environmental determinants of faunal diversity. *Am. Nat.* **93**, 337–342 (1959).
13. J. G. Owen, Patterns of herpetofaunal species richness: relation to temperature, precipitation, and variance in elevation. *J. Biogeogr.* **16**, 141–150 (1989).
14. J. J. Lennon, J. J. D. Greenwood, J. R. G. Turner, Bird diversity and environmental gradients in Britain: a test of the species–energy hypothesis. *J. Anim. Ecol.* **69**, 581–598 (2000).
15. K. H. Kozak, J. J. Wiens, Accelerated rates of climatic-niche evolution underlie rapid species diversification. *Ecol. Lett.* **13**, 1378–1389 (2010).
16. J. Rolland, F. L. Condamine, F. Jiguet, H. Morlon, Faster speciation and reduced extinction in the Tropics contribute to the mammalian Latitudinal Diversity Gradient. *PLoS Biol.* **12**, e1001775 (2014).
17. C. Gómez-Rodríguez, A. Baselga, J. J. Wiens, Is diversification rate related to climatic niche width? *Glob. Ecol. Biogeogr.* **24**, 383–395 (2015).
18. F. L. Condamine, J. Rolland, H. Morlon, Assessing the causes of diversification slowdowns: temperature-dependent and diversity-dependent models receive equivalent support. *Ecol. Lett.* **22**, 1900–1912 (2019).
19. J. Garcia-Porta, *et al.*, Environmental temperatures shape thermal physiology as well as diversification and genome-wide substitution rates in lizards. *Nat. Commun.* **10**, 4077 (2019).
20. M. H. Sayle, The metabolism of insects. *Q. Rev. Biol.* **3**, 542–553 (1928).
21. P. R. Ehrlich, P. H. Raven, Butterflies and plants: a study of coevolution. *Evolution* **18**, 586–608 (1965).
22. C. W. Wheat, *et al.*, The genetic basis of a plant–insect coevolutionary key innovation. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20427 (2007).
23. P. P. Edger, *et al.*, The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8362 (2015).
24. A. Y. Kawahara, *et al.*, Evolution and diversification dynamics of butterflies. *bioRxiv [Preprint]*, 2022.05.17.491528 (2022) <https://doi.org/10.1101/2022.05.17.491528>.
25. F. L. Condamine, F. A. H. Sperling, N. Wahlberg, J. Y. Rasplus, G. J. Kergoat, What causes latitudinal gradients in species diversity? Evolutionary processes and ecological constraints on swallowtail biodiversity. *Ecol. Lett.* **15**, 267–277 (2012).
26. J. R. G. Turner, C. M. Gatehouse, C. A. Corey, Does solar energy control organic diversity? Butterflies, moths and the British climate. *Oikos* **48**, 195 (1987).



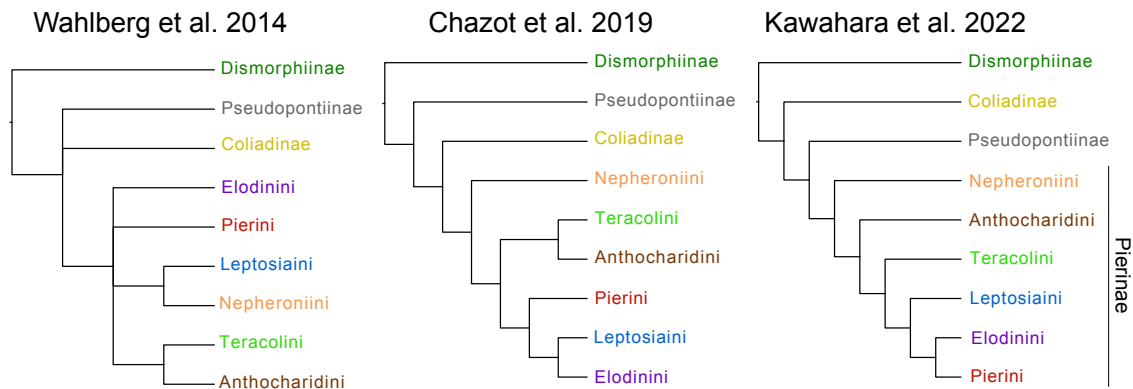
27. N. P. Kristensen, *Handbook of Zoology, Volume IV, Arthropoda: Insecta, Part 36. Lepidoptera, Moths and Butterflies, Volume 2: Morphology, Physiology, and Development*, N. P. Kristensen, Ed. (De Gruyter, 2003).
28. H. K. Clench, Behavioral thermoregulation in butterflies. *Ecology* **47**, 1021–1034 (1966).
29. A. Caporale, H. P. Romanowski, N. O. Mega, Winter is coming: Diapause in the subtropical swallowtail butterfly *Euryades corethrus* (Lepidoptera, Papilionidae) is triggered by the shortening of day length and reinforced by low temperatures. *J. Exp. Zool. A Ecol. Integr. Physiol.* **327**, 182–188 (2017).
30. M. Klockmann, F. Kleinschmidt, K. Fischer, Carried over: heat stress in the egg stage reduces subsequent performance in a butterfly. *PLOS ONE* **12**, e0180968 (2017).
31. G. Lamas, Catalog of the butterflies (Papilionoidea). (2015).
32. P. R. Ackery, R. de Jong, R. I. Vane-Wright, “The butterflies: Hedyloidea, Hesperioidea and Papilionoidea” in *Handbook of Zoology, Volume IV, Arthropoda: Insecta, Part 35. Lepidoptera, Moths and Butterflies, Volume 1: Evolution, Systematics, and Biogeography, 4. Arthropoda: Insecta, Part 35*, N. P. Kristensen, Ed. (Walter de Gruyter, 1998), pp. 263–300.
33. B. C. Schmidt, Diagnosis and distribution of the Kivalliq sulphur (*Colias rankinensis* Verhulst), an enigmatic Arctic butterfly. *J. Lepid. Soc.* **72**, 227–232 (2018).
34. A. El-Gabbas, F. Gilbert, The desert beauty *Calopieris eulimene*: a butterfly new to Egypt (Insecta: Lepidoptera). *Zool. Middle East.* **62**, 279–281 (2016).
35. M. F. Braby, R. Vila, N. E. Pierce, Molecular phylogeny and systematics of the Pieridae (Lepidoptera: Papilionoidea): higher classification and biogeography. *Zool. J. Linn. Soc.* **147**, 239–275 (2006).
36. N. Wahlberg, J. Rota, M. F. Braby, N. E. Pierce, C. W. Wheat, Revised systematics and higher classification of pierid butterflies (Lepidoptera: Pieridae) based on molecular data. *Zool. Scr.* **43**, 641–650 (2014).
37. M. F. Braby, N. E. Pierce, Systematics, biogeography and diversification of the Indo-Australian genus *Delias* Hübner (Lepidoptera: Pieridae): phylogenetic evidence supports an ‘out-of-Australia’ origin. *Syst. Entomol.* **32**, 2–25 (2007).
38. M. F. Braby, N. E. Pierce, R. Vila, Phylogeny and historical biogeography of the subtribe Aporiina (Lepidoptera: Pieridae): implications for the origin of Australian butterflies. *Biol. J. Linn. Soc.* **90**, 413–440 (2007).
39. H. Warren-Gash, K. Aduse-Poku, L. Murillo-Ramos, N. Wahlberg, Systematics and evolution of the African butterfly genus *Mylothris* (Lepidoptera, Pieridae). *Nota Lepidopterol.* **43**, 1–14 (2020).
40. J. Zhang, Q. Cong, J. Shen, P. A. Opler, N. V. Grishin, Genomics-guided refinement of butterfly taxonomy. *Taxon. Rep. Int. Lepid. Surv.* **9**, 1–54 (2021).
41. M. Espeland, *et al.*, A comprehensive and dated phylogenomic analysis of butterflies. *Curr. Biol.* **28**, 770–778.e5 (2018).
42. N. Chazot, *et al.*, Priors and posteriors in Bayesian timing of divergence analyses: the age of butterflies revisited. *Syst. Biol.* **68**, 797–813 (2019).
43. M. F. Braby, K. Nishida, The immature stages, larval food plants and biology of Neotropical mistletoe butterflies (Lepidoptera: Pieridae). II. The *Catasticta* group (Pierini: Aporiina). *J. Nat. Hist.* **44**, 1831–1928 (2010).
44. L. L. Mota, A. K. Silva, A. V. L. Freitas, L. A. Kaminski, Immature stages of *Archonias brassolis tereas* (Godart) (Pieridae: Pierini), with notes on interspecific interactions between mistletoe butterflies. *J. Lepid. Soc.* **70**, 289–294 (2016).
45. A. Y. Kawahara, *et al.*, Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 22657–22663 (2019).
46. K. S. Sheldon, J. J. Tewksbury, The impact of seasonality in temperature on thermal tolerance and elevational range size. *Ecology* **95**, 2134–2143 (2014).
47. A. G. Fischer, Latitudinal variations in organic diversity. *Evolution* **14**, 64–81 (1960).

48. M. Franzén, *et al.*, Differences in phenology, daily timing of activity, and associations of temperature utilization with survival in three threatened butterflies. *Sci. Rep.* **12**, 7534 (2022).
49. A. J. Bladon, *et al.*, How butterflies keep their cool: physical and ecological traits influence thermoregulatory ability and population trends. *J. Anim. Ecol.* **89**, 2440–2450 (2020).
50. C.-C. Tsai, *et al.*, Physical and behavioral adaptations to prevent overheating of the living wings of butterflies. *Nat. Commun.* **11**, 551 (2020).
51. W. B. Watt, Adaptive significance of pigment polymorphisms in *Colias* butterflies, ii. Thermoregulation and photoperiodically controlled melanin variation in *Colias eurytheme*. *Proc. Natl. Acad. Sci. U.S.A.* **63**, 767–774 (1969).
52. A. M. Shapiro, “Seasonal Polyphenism” in *Evolutionary Biology*, M. K. Hecht, W. C. Steere, B. Wallace, Eds. (Springer US, 1976), pp. 259–333.
53. J. G. Kingsolver, D. C. Wiernasz, Seasonal polyphenism in wing-melanin pattern and thermoregulatory adaptation in *Pieris* butterflies. *Am. Nat.* **137**, 816–830 (1991).
54. S. Gautam, K. Kunte, Adaptive plasticity in wing melanisation of a montane butterfly across a Himalayan elevational gradient. *Ecol. Entomol.* **45**, 1272–1283 (2020).
55. D. Jacobsen, R. Schultz, A. Encalada, Structure and diversity of stream invertebrate assemblages: the influence of temperature with altitude and latitude. *Freshw. Biol.* **38**, 247–261 (1997).
56. C. Chu, *et al.*, Direct and indirect effects of climate on richness drive the latitudinal diversity gradient in forest trees. *Ecol. Lett.*, ele.13175 (2018).
57. K. Aduse-Poku, *et al.*, Miocene climate and habitat change drove diversification in *Bicyclus*, Africa’s largest radiation of satyrine butterflies. *Syst. Biol.* **71**, 570–588 (2021).
58. M. Espeland, *et al.*, Ancient Neotropical origin and recent recolonisation: Phylogeny, biogeography and diversification of the Riodinidae (Lepidoptera: Papilionoidea). *Mol. Phylogenet. Evol.* **93**, 296–306 (2015).
59. K. T. Mitter, *et al.*, The butterfly subfamily Pseudopontiinae is not monobasic: marked genetic diversity and morphology reveal three new species of *Pseudopontia* (Lepidoptera: Pieridae). *Syst. Entomol.* **36**, 139–163 (2011).
60. A. Y. Kawahara, *et al.*, Phylogenetics of moth-like butterflies (Papilionoidea: Hedyliidae) based on a new 13-locus target capture probe set. *Mol. Phylogenet. Evol.* **127**, 600–605 (2018).
61. N. Wahlberg, C. W. Wheat, Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of Lepidoptera. *Syst. Biol.* **57**, 231–242 (2008).
62. F. Krueger, Trim Galore Version 0.4.0 (2015).
63. J. W. Breinholt, *et al.*, Resolving relationships among the megadiverse butterflies and moths with a novel pipeline for Anchored Phylogenomics. *Syst. Biol.* **67**, 78–93 (2018).
64. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
65. Z. Chang, *et al.*, Bridger: a new framework for de novo transcriptome assembly using RNA-seq data. *Genome Biol.* **16**, 30 (2015).
66. S. Zhan, C. Merlin, J. L. Boore, S. M. Reppert, The monarch butterfly genome yields insights into long-distance migration. *Cell* **147**, 1171–1185 (2011).
67. K. Katoh, D. M. Standley, MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
68. P. Kück, K. Meusemann, FASconCAT: convenient handling of data matrices. *Mol. Phylogenet. Evol.* **56**, 1115–1118 (2010).
69. A. Larsson, AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278 (2014).
70. L.-T. Nguyen, H. A. Schmidt, A. Von Haeseler, B. Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

71. B. Q. Minh, *et al.*, IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
72. B. Q. Minh, M. A. T. Nguyen, A. von Haeseler, Ultrafast approximation for phylogenetic Bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).
73. S. A. Smith, B. C. O'Meara, TreePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**, 2689–2690 (2012).
74. R. A. St Laurent, A. P. S. Carvalho, C. Earl, A. Y. Kawahara, Food plant shifts drive the diversification of sack-bearer moths. *Am. Nat.* **198**, E170–E184 (2021).
75. A. Stamatakis, RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
76. A. J. Drummond, M. A. Suchard, D. Xie, A. Rambaut, Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
77. S. Magallón, S. Gómez-Acevedo, L. L. Sánchez-Reyes, T. Hernández-Hernández, A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New. Phytol.* **207**, 437–453 (2015).
78. R. de Jong, Fossil butterflies, calibration points and the molecular clock (Lepidoptera: Papilionoidea). *Zootaxa* **4270**, 1–63 (2017).
79. H. L. Owens, *et al.*, occCite: Tools for querying and managing large biodiversity occurrence datasets. *Ecography* **44**, 1228–1235 (2021).
80. S. Chamberlain, *et al.*, rgbif: Interface to the Global Biodiversity Information Facility API. R package v.3.7.3, <https://cran.r-project.org/package=rgbif> (2022).
81. GBIF: The Global Biodiversity Information Facility, What is GBIF? (2022) (May 30, 2022).
82. H. Wickham, ggplot2: Elegant Graphics for Data Analysis, ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org> (2016).
83. V. Dincă, V. A. Lukhtanov, G. Talavera, R. Vila, Unexpected layers of cryptic diversity in wood white *Leptidea* butterflies. *Nat. Commun.* **2**, 324 (2011).
84. S. E. Fick, R. J. Hijmans, WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).
85. R. J. Hijmans, raster: geographic data analysis and modeling. R package version 3.4-10 (2021).
86. A. C. Carnaval, M. J. Hickerson, C. F. B. Haddad, M. T. Rodrigues, C. Moritz, Stability predicts genetic diversity in the Brazilian Atlantic Forest hotspot. *Science* **323**, 785–789 (2009).
87. WFO, World Flora Online, (2022) <http://www.worldfloraonline.org> (August 1, 2020).
88. V. Barve, Taxotools: tools to handle taxonomic lists. R package (2021) <https://doi.org/10.5281/ZENODO.3934939> (June 23, 2021).
89. G. S. Robinson, P. R. Ackery, I. J. Kitching, G. W. Beccaloni, HOSTS - a database of the World's lepidopteran hostplants (2010) (April 16, 2020).
90. J. R. Ferrer-Paris, *et al.*, A compilation of larval host-plant records for six families of butterflies (Lepidoptera: Papilionoidea) from available electronic resources. 6927120 Bytes (2014).
91. R. Eastwood, An aberrant food-plant record for "*Jalmenus evagoras*" (Donovan) (Lepidoptera: Lycaenidae). *Aust. Entomol.* **26**, 13–14 (1999).
92. N. E. Pierce, D. R. Nash, "The Imperial Blue, *Jalmenus evagoras* (Lycaenidae)" in *The Biology of Australian Butterflies (Monographs on Australian Lepidoptera)*, (CSIRO Press, 1999), pp. 277–316.
93. D. L. Rabosky, *et al.*, BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods Ecol. Evol.* **5**, 701–707 (2014).
94. S. Höhna, *et al.*, RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst. Biol.* **65**, 726–736 (2016).
95. S. Höhna, *et al.*, A Bayesian approach for estimating branch-specific speciation and extinction rates. *bioRxiv [Preprint]*, 555805 (2019) <https://doi.org/10.1101/555805>.

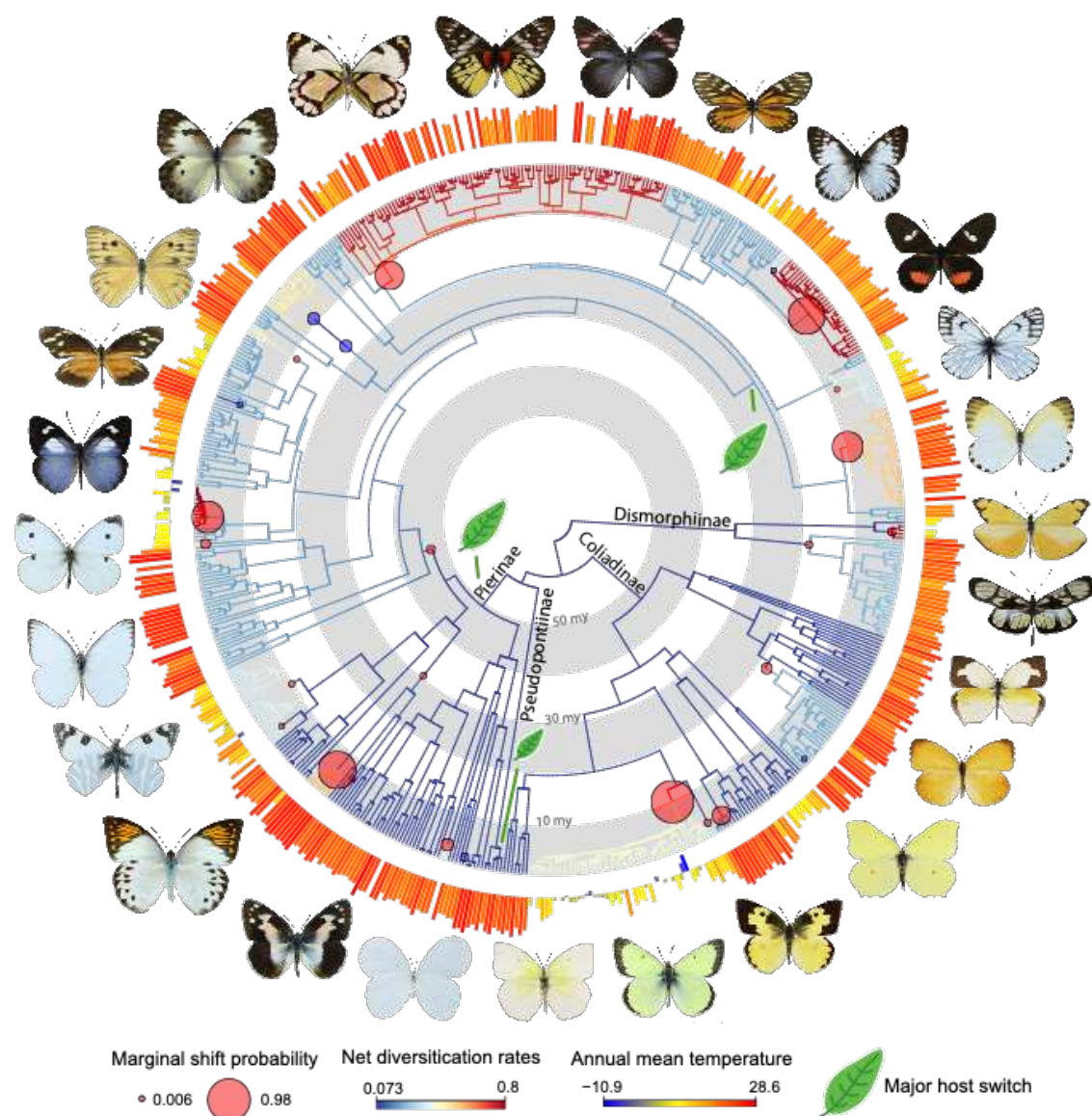
96. S. Louca, M. W. Pennell, Extant timetrees are consistent with a myriad of diversification histories. *Nature* **580**, 502–505 (2020).
97. S. Louca, M. Doebeli, Efficient comparative phylogenetics on large trees. *Bioinformatics* **34**, 1053–1055 (2018).
98. L. J. Revell, phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
99. M. Pagel, Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884 (1999).
100. S. P. Blomberg, T. Garland JR., A. R. Ives, Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57**, 717–745 (2003).
101. J. Pinheiro, D. Bates, R Core Team, nlme: Linear and Nonlinear Mixed Effects Models (2022).
102. R. G. FitzJohn, Diversitree: comparative phylogenetic analyses of diversification in R. *Methods Ecol. Evol.* **3**, 1084–1092 (2012).
103. X. Zhang, *et al.*, Macroevolutionary pattern of *Saussurea* (Asteraceae) provides insights into the drivers of radiating diversification. *Proc. R. Soc. B* **288**, 20211575 (2021).
104. F. Condamine, A. Antonelli, L. P. Lagomarsino, C. Hoorn, L. H. Liow, “Teasing Apart Mountain Uplift, Climate Change and Biotic Drivers of Species Diversification” in *Mountains, Climate and Biodiversity*, (John Wiley & Sons Ltd., 2018), pp. 257–272.
105. H. Morlon, T. L. Parsons, J. B. Plotkin, Reconciling molecular phylogenies with the fossil record. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 16327 (2011).
106. T. Stadler, Mammalian phylogeny reveals recent diversification rate shifts. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 6187 (2011).
107. F. L. Condamine, J. Rolland, H. Morlon, Macroevolutionary perspectives to environmental change. *Ecol. Lett.* **16**, 72–85 (2013).
108. A. Gamisch, Technical comment on Condamine *et al.* (2019): a cautionary note for users of linear diversification dependencies. *Ecol. Lett.* **23**, 1169–1171 (2020).
109. H. Morlon, J. Rolland, F. L. Condamine, Response to technical comment ‘A cautionary note for users of linear diversification dependencies.’ *Ecol. Lett.* **23**, 1172–1174 (2020).

## Figures

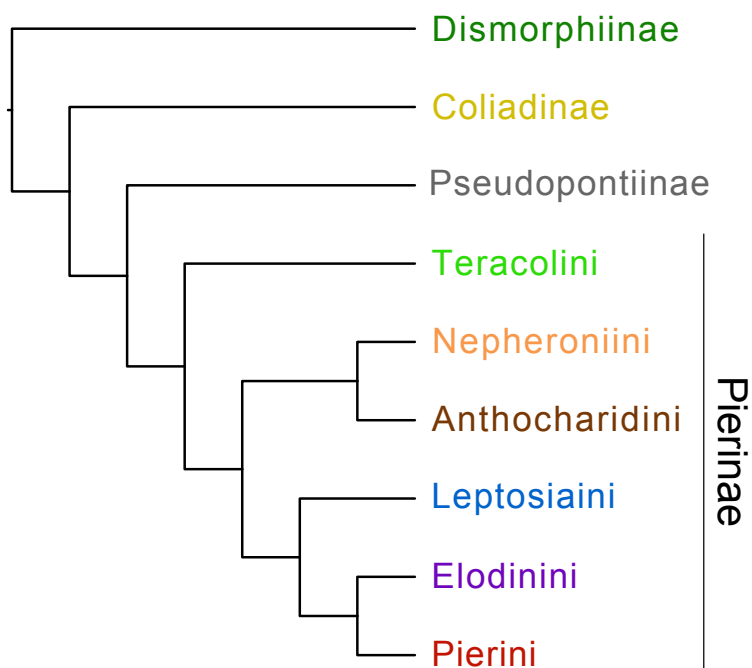


**Fig. 1.** Summary of the branching patterns of Pieridae among recent molecular phylogenetic studies.





**Fig 2.** Evolutionary relationships and diversification patterns of Pieridae. Time-calibrated tree of 593 species. Branches with significant net diversification rate shifts as estimated by BAMM are indicated with red and blue circles. Annual mean temperature (BIO1, WorldClim) for each species is indicated with colored bars in the outer ring. A green leaf indicates a major shift in host plant use along the branch as reconstructed by SIMMAP. Butterfly images are not to scale.



**Fig. 3.** Simplified topology showing the evolutionary relationships of Pieridae subfamilies and tribes as inferred in the present study.