INSECT COMMUNITIES: DIVERSITY PATTERNS AND THEIR DRIVING FORCES

EDITED BY: Ai-bing Zhang, Xin Zhou, Shu-Jun Wei, Jian Zhang, Yasuoki Takami and Gengping Zhu

PUBLISHED IN: Frontiers in Ecology and Evolution

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-83250-235-8
DOI 10.3389/978-2-83250-235-8

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

INSECT COMMUNITIES: DIVERSITY PATTERNS AND THEIR DRIVING FORCES

Topic Editors:
Ai-Bing Zhang, Capital Normal University, China
Xin Zhou, China Agricultural University, China
Shu-Jun Wei, Beijing Academy of Agricultural and Forestry Sciences, China
Jian Zhang, East China Normal University, China
Yasuoki Takami, Kobe University, Japan
Gengping Zhu, Tianjin Normal University, China

Citation: Zhang, A.-B., Zhou, X., Wei, S.-J., Zhang, J., Takami, Y., Zhu, G., eds. (2023). Insect Communities: Diversity Patterns and their Driving Forces. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-235-8

Table of Contents

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 07 July 2021
doi: 10.3389/fevo.2021.702178

Population Genetics of the Black Citrus Aphid *Aphis aurantii* (Hemiptera, Aphididae) in China

*Hong-Ling Liu[1], Zhi-Teng Chen[2], Chao Liu[1], Xing-Long Wu[1], Ke-Jun Xiao[1] and De-Qiang Pu[1]**

[1] Key Laboratory of Integrated Pest Management of Southwest Crops, Institute of Plant Protection, Sichuan Academy of Agricultural Sciences, Chengdu, China, [2] School of Grain Science and Technology, Jiangsu University of Science and Technology, Zhenjiang, China

The black citrus aphid, *Aphis aurantii* Boyer de Fonscolombe, 1841, is one of the most destructive pests in commercial tea plantations and gardens in China. In this study, we investigated the population genetic structure of *A. aurantii* based on the concatenated sequences of two mitochondrial genes, cytochrome c oxidase I (*cox1*) and cytochrome b (*cytb*). A total of 166 haplotypes were identified from 177 individuals collected at 11 locations in China. The whole Chinese *A. aurantii* population showed a low nucleotide diversity (0.00968) and a high population diversity (haplotype diversity; 0.9991). The haplotypes of the 11 local populations were widely distributed in the neighbor-joining phylogenetic tree and haplotype network diagram, whereas no apparent lineages were detected. Gene flow analysis showed gene exchanges among local populations. The pairwise F_{st} values revealed a certain amount of genetic difference among local populations. Analysis of molecular variance (AMOVA) reflected genetic differences both within and among populations. The isolation by distance (IBD) analysis revealed a high positive correlation between the geographic distance and genetic distance of the different populations. Neutral test and mismatch distribution suggested that *A. aurantii* may have experienced recent population expansion events.

Keywords: genetic structure, mitochondrial gene, aphids, *Aphis aurantii*, pest

OPEN ACCESS

Edited by:
Ai-bing Zhang,
Capital Normal University, China

Reviewed by:
Jing-Tao Sun,
Nanjing Agricultural University, China
Cheng-Min Shi,
Chinese Academy of Sciences, China

****Correspondence:***
De-Qiang Pu
pdqpudeqiang@163.com

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *29 April 2021*
Accepted: *07 June 2021*
Published: *07 July 2021*

Citation:
Liu H-L, Chen Z-T, Liu C, Wu X-L, Xiao K-J and Pu D-Q (2021) Population Genetics of the Black Citrus Aphid Aphis aurantii (Hemiptera, Aphididae) in China. Front. Ecol. Evol. 9:702178. doi: 10.3389/fevo.2021.702178

INTRODUCTION

The tea aphid (*Aphis aurantii* Boyer de Fonscolombe, 1841), also known as the "black citrus aphid," is widely distributed in tropical and subtropical regions, including the Mediterranean region, England, Africa, India, Southeast Asia, Australia, South America, Central America, and North America (Tahori and Hazan, 1970; Carver, 1978). *A. aurantii* is an extremely polyphagous species with a very wide host range, inhabiting plants from over 190 genera of 80 families. Many of these hosts are economically important plants, such as citrus, coffee, tea, cacao, avocado, loquat, litchi, mango, fig, *Camellia* spp., *Cinchona* spp., *Annona* spp., *Macadamia* spp., *Piper* spp., and *Artocarpus* spp. (Carver, 1978). *A. aurantii* shows a preference for members of the families Rutaceae, Rosaceae, Apocynaceae, and Rubiaceae. In China, *A. aurantii* frequently occurs throughout the tea regions and causes direct feeding damage to fresh leaves and tender shoots, seriously reducing the output and lowering the quality of commercial teas. The excreted honeydew of *A. aurantii* can also incur mildew on the leaves, which reduces photosynthesis of the tea leaves.

In Chinese tea gardens, the main strategy for controlling these aphids is using chemical pesticides. However, this method has resulted in three main issues: residue, resistance, and

resurgence. Very few genetic studies have been conducted for *A. aurantii*. The partial mitochondrial cytochrome c oxidase I (*cox1*) genes of some *A. aurantii* specimens were sequenced from Guangxi Province in southern China (Wang and Qiao, 2009), and apparent intraspecific genetic diversity was found within *A. aurantii*. The mitochondrial genome of *A. aurantii* from Guangxi Province was sequenced (Wang et al., 2019) and the phylogeny of Aphidoidea was rebuilt, revealing the close relationship between *A. aurantii* and *Aphis craccivora* Koch, 1854. A recent study sequenced the mitochondrial genome of *A. aurantii* from Sichuan Province and constructed the phylogeny of Aphididae (Pu et al., 2020), supporting the close relationship of *A. aurantii* and other aphids in *Aphis*. The transcriptome of *A. aurantii* has also been sequenced and analyzed (Hong et al., 2020), providing a basic transcriptomic dataset to identify simple sequence repeats (SSRs). However, these molecular studies did not involve population genetic analysis for the Chinese *A. aurantii*. To better understand the population genetics of this important pest and provide new insights for controlling this pest, we sampled 11 sites (**Figure 1**) and examined the population genetic structure of Chinese *A. aurantii* based on the concatenated sequences of two mitochondrial genes: *cox1* and *cytb*.

MATERIALS AND METHODS

Sampling and DNA Extraction

A total of 220 individuals of *A. aurantii* were sampled from 11 tea plantations in China in 2020 (**Figure 1** and **Table 1**). The populations from the 11 locations were, respectively, named as P1, P2, and so on. All individuals were preserved in 99.9% ethanol until DNA extraction. Genomic DNA was extracted from each individual using the E.Z.N.A.® Tissue DNA Kit (Omega, Norcross, GA, United States) following the manufacturer's protocol and preserved at −20°C.

PCR Reaction and Sequencing

The *cox1* gene was amplified with the universal primers cox1F 5′-ATTCAACCAATCATAAAGATATTGG-3′ and cox1R 5′-TAAACTTCTGGATGTCCAAAAAATCA-3′ (Hebert et al., 2004). The *cytb* gene was amplified with primers cytbF 5′-GATGATGAAATTTTGGATC-3′ and cytbR 5′-CTAATGCAATAACTCCTCC-3′ (Harry et al., 1998). The conditions for the PCR reactions were as follows: initial denaturation at 94°C for 5 min, followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 54°C for 45 s, and elongation at 72°C for 60 s, and a final elongation at 72°C for 10 min. The PCR products were separated by electrophoresis in 1.0% agarose gels and purified with the Axygen DNA Gel Extraction Kit (Axygen Biotechnology, Hangzhou, China). The purified PCR fragments were sequenced by an ABI 3730 automated sequencer (Biozeron Co., Ltd., Shanghai, China).

Analyses of Genetic Variation

The gene length, conserved sites, variable sites, parsimony informative sites, singleton sites, and the average A + T contents were, respectively, calculated for the *cox1* and *cytb* genes and their concatenated sequences by Mega v6.0 (Tamura et al., 2013). Genetic diversity parameters, including the haplotype number (H), number of polymorphic sites (S), haplotype diversity (H_d), and nucleotide diversity (Pi), were estimated using DnaSP software (Librado and Rozas, 2009). Each haplotype was named as H1, H2, and so on. Molecular variance analysis (AMOVA), pairwise F_{st} values, neutrality test, and mismatch distribution analysis were performed with the Arlequin v3.5 software (Excoffier and Lischer, 2010). Gene flow analysis was also conducted with the Arlequin v3.5 software to generate the N_m values and examine the gene exchanges between different populations. To test the correlation between geographic distance and genetic difference, the IBD (isolation by distance) analysis based on the Mantel test was performed using IBD v1.53 (Bohonak, 2002). Principal coordinates analysis (PCoA) was conducted from the distance matrix of the 11 populations and eight provincial groups.

Phylogenetic Analyses

Phylogenetic reconstruction using the concatenated sequences was performed with the neighbor-joining (NJ) method hosted by Mega v6.0 (Tamura et al., 2013). In the NJ analysis, the bootstrap method was used with 1,000 bootstrap replications. The Kimura two-parameter model was selected as the substitution model. Both transitions and transversions were included, and the rates among sites were regarded as uniform. Gaps/missing data were pairwise deleted. All computed trees were adjusted and visualized in FigTree v1.4.2. The relationships among the different haplotypes were inferred using a median-joining method in PopART software (Leigh and Bryant, 2015).

RESULTS

DNA Sequencing

The 220 sampled individuals of *A. aurantii* yielded 196 high-quality *cox1* sequences and 201 *cytb* sequences (GenBank accession numbers MW265715–MW265910 and MW289592–MW289792). Both genes were simultaneously obtained for 177 individuals. The remaining double-peak sequences were excluded from the analyses. For the *cox1* genes, the consensus length was 702 bp for all populations (**Table 2**). The average A + T content for the *cox1* genes was 75.5%. The *cox1* sequences of the 11 populations included 623 conserved sites, 76 variable sites, 54 parsimony informative sites, and 22 singleton sites. For the *cytb* genes, the consensus length was 829 bp for all populations (**Table 2**). The average A + T content for the *cox1* genes was 76.9%. The *cytb* sequences of the 11 populations included 666 conserved sites, 150 variable sites, 114 parsimony informative sites, and 33 singleton sites. For the concatenated sequences of the *cox1* and *cytb* genes of the 177 individuals, the consensus length was 1,507 bp and the average A + T content was 76.2% (**Table 2**). There were 1,287 conserved sites, 214 variable sites, 160 parsimony informative sites, and 54 singleton sites in the concatenated sequences. These concatenated sequences were used for subsequent genetic and phylogenetic analyses.

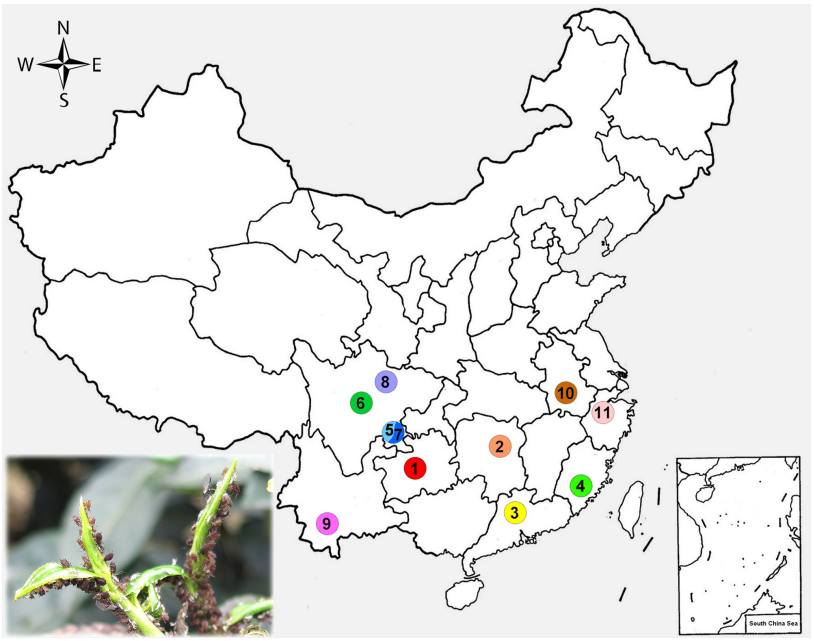


FIGURE 1 | Collection localities of the samples used in this study. (1) Guiyang, Guizhou Province; (2) Changsha, Hunan Province; (3) Qingyuan, Guangdong Province; (4) Quanzhou, Fujian Province; (5) Dadukou, Sichuan Province; (6) Ya'an, Sichuan Province; (7) Huguo, Sichuan Province; (8) Mianyang, Sichuan Province; (9) Pu'er, Yunnan Province; (10) Chizhou, Anhui Province; (11) Hangzhou, Zhejiang Province.

Genetic Variation and Diversity

A total of 166 haplotypes were identified for the 177 samples of *A. aurantii*, which had both *cox1* and *cytb* sequences (**Supplementary Material**). Most of the haplotypes were unique and corresponded to their own population; only three haplotypes (H28, H34, and H63) were shared by different populations. The largest haplotype was H28, which was shared by one individual of P2 from Hunan, two individuals of P7 from Sichuan, and one individual of P10 from Anhui. The entire dataset comprised 92 polymorphic sites, with a high H_d of 0.9991 (±0.0008) and a low Pi of 0.0097 (±0.0005) (**Table 3**). For each population, the number of polymorphic sites (*S*) ranged from 37 (P7) to 88 (P11). The H_d was 1.000 for all populations. The Pi values varied from 0.0112 (±0.0008, P2) to 0.0211 (±0.0014, P9).

The AMOVA showed more variation of genetic differentiation within populations than among populations of *A. aurantii* (**Table 4**). Over 87.49% variation of genetic differentiation occurred within populations, and only 12.51% of the genetic variation was found among populations. The pairwise F_{st} values were highly variable, ranging from 0.0229 (between populations P3 and P4) to 0.2984 (between populations P2 and P11). High genetic differentiations were present between populations P2 and P11 ($F_{st} > 0.25, p < 0.05$), whereas no significant differentiations were present between other pairwise populations (**Table 5**). Most of the N_m values from the gene flow analysis were more than 1, except for the negative values between P11 and populations P2, P3, P4, P7, and P8 (**Table 5**). The highest N_m value was found between P3 and P4.

The IBD analysis revealed a highly significant positive correlation ($r = 0.4190, p < 0.01$) between the geographic distance and genetic distance of the *A. aurantii* populations.

TABLE 1 | Sample information in this study.

Code	Sampling location	Sample size
P1	Guiyang City, Guizhou Province	20
P2	Changsha City, Hunan Province	20
P3	Qingyuan City, Guangdong Province	20
P4	Shennei Village, Anxi County, Quanzhou City, Fujian Province	20
P5	Limu Village, Dadukou Town, Naxi District, Luzhou City, Sichuan Province	20
P6	Dangcunba, Mingshan County, Ya'an City, Sichuan Province	20
P7	Huguo Town, Naxi District, Luzhou City, Sichuan Province	20
P8	Beichuan, Mianyang City, Sichuan Province	20
P9	Heimao, Pu'er City, Yunnan Province	20
P10	Lixin, Chizhou City, Anhui Province	20
P11	Xihu District, Hangzhou City, Zhejiang Province	20

TABLE 2 | Content of gene sequences in *Aphis aurantii*.

Gene	Length (bp)	C	V	Pi	S	A + T (%)
cox1	702	623	76	54	22	75.5
cytb	829	666	150	114	33	76.9
cox1 + cytb	1,507	1,287	214	160	54	76.2

Length, consensus length; C, conserved sites; V, variable sites; Pi, parsimony informative sites; S, singleton sites; A + T, average A + T content.

TABLE 3 | Parameters of genetic diversity in populations of *Aphis aurantii*.

Population	*N*	*H*	*S*	$H_d \pm SD$	Pi ± SD
P1	20	20	50	1.000 ± 0.016	0.0119 ± 0.0007
P2	20	20	50	1.000 ± 0.016	0.0112 ± 0.0008
P3	14	14	47	1.000 ± 0.027	0.0146 ± 0.0008
P4	18	18	57	1.000 ± 0.019	0.0148 ± 0.0008
P5	18	18	71	1.000 ± 0.019	0.0150 ± 0.0014
P6	11	11	55	1.000 ± 0.039	0.0151 ± 0.0014
P7	17	17	37	1.000 ± 0.020	0.0114 ± 0.0006
P8	15	15	61	1.000 ± 0.024	0.0167 ± 0.0011
P9	17	17	87	1.000 ± 0.020	0.0211 ± 0.0014
P10	14	14	62	1.000 ± 0.027	0.0172 ± 0.0013
P11	13	13	88	1.000 ± 0.030	0.0185 ± 0.0031
Total	177	166	92	0.9991 ± 0.0008	0.0097 ± 0.0005

N, sample size; H, number of haplotypes; S, number of polymorphic sites; H_d, haplotype diversity; Pi, nucleotide diversity.

The first two coordinates of the PCoA explained 78.78% of the total variation among the 11 populations (**Figure 2**). The first and second axes explained 52.8 and 25.98% variation, respectively. Most genotypes on the PCoA graph were separately clustered, except for populations 1 and 10, which were partially overlapped. The PCoA at the population level also revealed that P2, P4, P7, and P11 were clearly distinct from the other populations. In the PCoA at the province level, most genotypes from Sichuan were closely located with those from Guizhou, Yunnan, and Anhui; the genotypes from Hunan, Fujian, and Zhejiang were clearly distinct from those of the other provinces.

Neutrality Test

The neutrality test was conducted based on the concatenated *cox1* and *cytb* sequences (**Table 6**). Tajima's *D* values were positive for most populations except for P11, and all these values were not statistically significant (**Table 6**). Fu's F_s values were negative for most populations except for P7, and most of the values were not statistically significant, except for that of P1. For the entire dataset of *A. aurantii*, Tajima's *D* value was positive and not significant, while Fu's F_s value was negative and significant (**Table 6**).

Mismatch Distribution Analysis

The mismatch distribution analyses generated single bell curves for the combined dataset of all *A. aurantii* populations (**Figure 3**). In the two curves calculated by different models, the observed data generally corresponded with the simulated mismatch distribution under the spatial expansion model rather than the sudden expansion model.

For the entire dataset of *A. aurantii* under the sudden expansion model, the goodness-of-fit test computed the sum of squared deviation (SSD) as 0.0040, which was not statistically significant (p = 0.2300); the Harpending's raggedness index (HRI) was 0.0003, which was also not statistically significant (p = 1.0000). For each population, the SSD and HRI values were all not significant (**Table 7**).

For the entire dataset of *A. aurantii* under the spatial expansion model, the goodness-of-fit test computed the SSD as 0.0003, which was not significant (p = 0.8500), and the HRI was 0.0003, which was also not significant (p = 0.9700). For each population, the SSD and HRI values were all not significant (**Table 7**).

Phylogenetic Relationships and Genetic Structure

The NJ tree was constructed using the concatenated *cox1* and *cytb* sequences (**Figure 4**). However, the bootstrap values of most clades were lower than 50%. Individuals of the 11 populations

TABLE 4 | Analysis of molecular variance (AMOVA) for the different populations of *Aphis aurantii*.

Source of variation	Degree of freedom	Sum of squares	Variance components	Percentage of variation	Fixation index
Among populations	10	686.328	2.9796 Va	12.51	F_{st} = 0.1251, p = 0
Within populations	166	3,458.067	20.8317 Vb	87.49	
Total	176	4,144.395	23.8113		

TABLE 5 | Matrix of pairwise F_{st} (below diagonal) and N_m values (above diagonal) among the 11 populations of *Aphis aurantii* based on concatenated sequences.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
P1		2.3241	3.4482	1.9233	6.1174	5.1525	2.5260	3.0217	2.6399	5.1757	1.6130
P2	0.0971**		1.2652	1.0175	1.3433	2.3647	5.2413	1.9300	1.0091	1.4498	0.5878
P3	0.0676**	0.1650**		10.656	2.2023	1.9604	1.1273	2.3558	1.6836	2.1495	0.9280
P4	0.1150**	0.1972**	0.0229		1.7177	2.0142	1.0309	1.9694	1.3188	1.4684	0.7754
P5	0.0393**	0.1569**	0.1020**	0.1271**		5.0046	1.4395	2.2480	5.7561	2.7975	1.5344
P6	0.0463**	0.0956**	0.1131**	0.1104**	0.0476*		3.0960	4.0066	1.7458	5.7984	1.0576
P7	0.0901**	0.0455**	0.1815**	0.1952**	0.1480**	0.0747**		1.9170	1.1984	2.0060	0.9364
P8	0.0764**	0.1147**	0.0959**	0.1126**	0.1001**	0.0587**	0.1154**		1.7894	2.8766	0.8955
P9	0.0865**	0.1986**	0.1293**	0.1594**	0.0416*	0.1253**	0.1726**	0.1226**		1.6402	1.8019
P10	0.0461**	0.1471**	0.1042**	0.1455**	0.0820**	0.0413*	0.1108**	0.0800**	0.1323**		1.2834
P11	0.1342**	0.2984**	0.2122**	0.2438**	0.1401**	0.1912**	0.2107**	0.2182**	0.1218**	0.1630**	

**Significant: 0.01 < p < 0.05; **Highly significant: p < 0.01.*

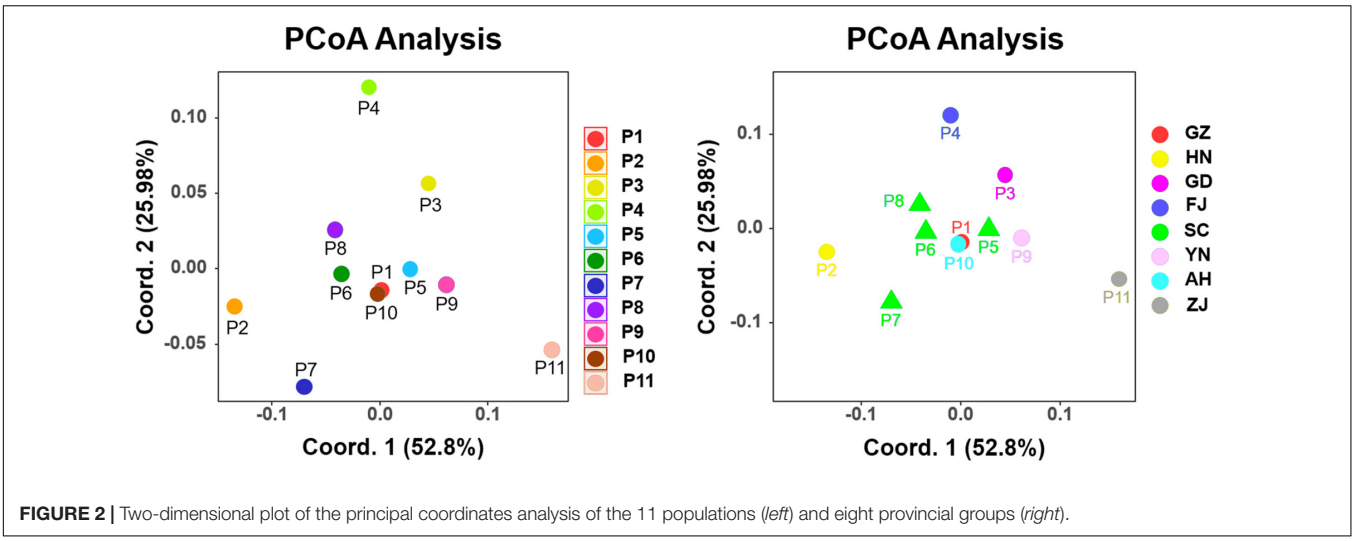


FIGURE 2 | Two-dimensional plot of the principal coordinates analysis of the 11 populations (*left*) and eight provincial groups (*right*).

were scattered in the tree and did not cluster into apparent lineages. Each geographic population cannot be divided in the phylogenetic analysis, indicating more variation of genetic differentiation within populations than among populations of *A. aurantii*. The only exception is P11, most samples of which grouped together in the NJ tree.

The median-joining network was constructed for populations from eight provinces (**Figure 5**). The 166 haplotypes were widely distributed among the sampling locations and were connected without apparently diverged lineages. Most haplotypes were unshared by different geographic groups (each province), except for haplotypes H28, H34, and H63. H28 was the mostly shared haplotype by three geographic groups from Hunan, Sichuan, and Anhui. H34 was shared by two geographic groups from Sichuan and Anhui. H63 was shared by two geographic groups from Hunan and Sichuan.

TABLE 6 | Neutrality test and the corresponding *p*-values for the populations of *Aphis aurantii* based on the concatenated *cox1* and *cytb* sequences.

Population	Tajima's *D* test	Significance (*p*)	Fu's F_s test	Significance (*p*)
P1	0.5631	0.7440	−4.0072	0.0260
P2	0.9156	0.8510	−2.1829	0.1590
P3	1.8761	0.9930	−2.2767	0.0750
P4	1.2028	0.9220	−1.7765	0.1630
P5	0.2844	0.6570	−2.6735	0.0680
P6	1.0998	0.9030	−0.7456	0.2100
P7	1.2288	0.9380	0.0190	0.4550
P8	1.3051	0.9320	−2.2267	0.0900
P9	0.9117	0.8680	−2.3448	0.0920
P10	1.4214	0.9560	−1.6907	0.1210
P11	−0.4017	0.3720	−1.1723	0.1580
Whole	0.1121	0.6610	−23.9127	0.0030

DISCUSSION

Genetic Variations

The purpose of the current study was to determine the genetic diversity and interrelationship of the Chinese populations of *A. aurantii* using mitochondrial gene markers. Assessment of the genetic diversity of the species can reveal its adaptability to environmental change, which is important for pest control or biological resource protection (Schmitt and Hewitt, 2004). The *A. aurantii* populations studied herein were characterized by a high haplotype diversity and a low nucleotide diversity. This suggests that *A. aurantii* might have undergone expansion from a low effective population size, or ancient population bottleneck or founder effect (Hedgecock et al., 1989; Grant and Bowen, 1998). The star-like structures of the haplotype network diagram also provided evidence for the recent expansion of *A. aurantii* from a small number of ancestors (Slatkin and Hudson, 1991; Grant and Bowen, 1998).

Different locations in China are represented by different local tea products. Such long-term monocultures of plant hosts could lead to the low nucleotide diversities of each tea aphid population. The ability of *A. aurantii* to adapt to chemical insecticides is poor, so it is susceptible to regional extinction events. The periodically excessive use of chemical insecticides in tea gardens might cause the total extinction of a regional population, which could lead to a further reduction in the genetic diversity of *A. aurantii*.

The AMOVA suggested the existence of more variation of genetic differentiation within populations than among populations of *A. aurantii*. The pairwise F_{st} results (0.02292–0.29841) suggested that low to moderate degrees of genetic differentiations are present among the *A. aurantii* populations (Wright, 1990). Meanwhile, most of the N_m values generated by gene flow analysis were higher than 1, further indicating the existence of frequent gene flow events among the *A. aurantii* populations (Slatkin and Maddison, 1989). Both the F_{st} and N_m analyses supported the genetically close relationship between P3 of Guangdong and P4 of Fujian. The PCoA also indicated the existence of genetic differentiations in most populations of *A. aurantii*.

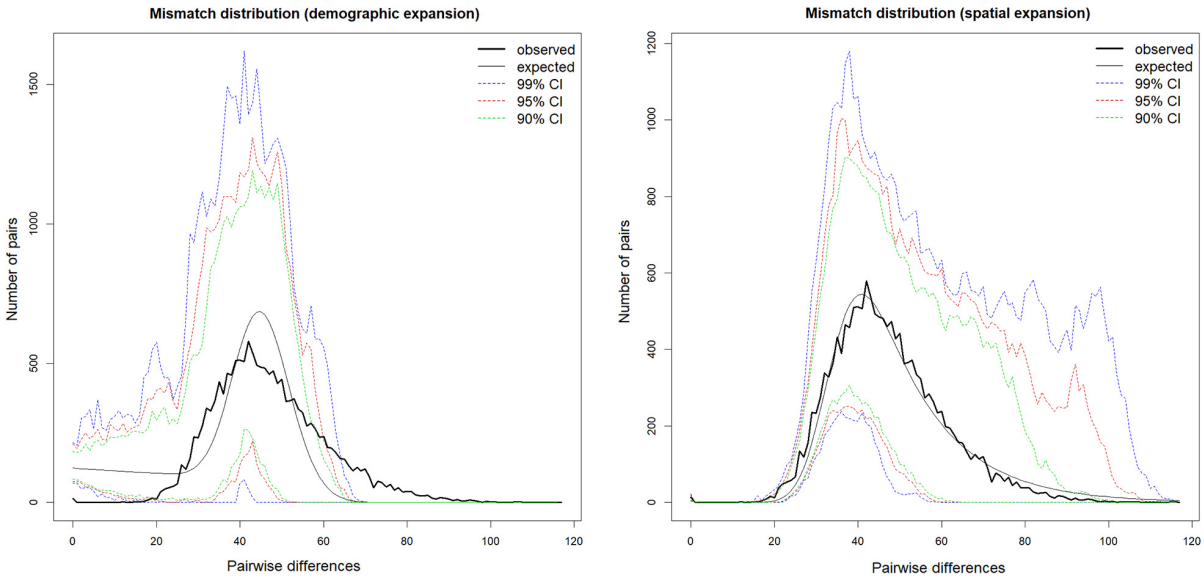


FIGURE 3 | Mismatch distributions of *Aphis aurantii* based on the concatenated *cox1* and *cytb* sequences.

The IBD analysis revealed a highly significant positive correlation ($r = 0.4190$, $p < 0.01$) between the geographic distance and genetic distance of the *A. aurantii* populations. Geographic isolation was supported as the main factor that reduced the gene flow events among the different populations of *A. aurantii*. Despite the restriction by geographic distance, the strong dispersal ability of *A. aurantii* resulted in a certain degree of gene exchange between populations, which was supported by the AMOVA and gene flow analysis. The current cosmopolitan distribution of *A. aurantii* (Piron et al., 2019)

TABLE 7 | Mismatch distribution and the corresponding *p*-values for the entire dataset of *Aphis aurantii* based on the concatenated *cox1* and *cytb* genes.

Population	Model	Sum of squared deviation	Significance (*p*)	Harpending's raggedness index	Significance (*p*)
P1	Sudden expansion	0.0091	0.3300	0.0160	0.1900
	Spatial expansion	0.0078	0.1900	0.0160	0.1500
P2	Sudden expansion	0.0075	0.4400	0.0152	0.3500
	Spatial expansion	0.0058	0.5900	0.0152	0.4800
P3	Sudden expansion	0.0161	0.2500	0.0328	0.1900
	Spatial expansion	0.0162	0.1300	0.0328	0.1300
P4	Sudden expansion	0.0073	0.3800	0.0194	0.0900
	Spatial expansion	0.0073	0.4300	0.0194	0.2300
P5	Sudden expansion	0.0132	0.4200	0.0143	0.5700
	Spatial expansion	0.0097	0.2900	0.0143	0.5000
P6	Sudden expansion	0.0146	0.6800	0.0284	0.7900
	Spatial expansion	0.0146	0.6400	0.0284	0.7100
P7	Sudden expansion	0.0138	0.2900	0.0186	0.2700
	Spatial expansion	0.0136	0.3700	0.0186	0.6000
P8	Sudden expansion	0.0131	0.3200	0.0220	0.4100
	Spatial expansion	0.0096	0.4100	0.0220	0.3300
P9	Sudden expansion	0.0068	0.7300	0.0150	0.3900
	Spatial expansion	0.0071	0.5900	0.0150	0.4600
P10	Sudden expansion	0.0140	0.6800	0.0242	0.5300
	Spatial expansion	0.0109	0.5900	0.0242	0.4700
P11	Sudden expansion	0.0103	0.8200	0.0230	0.8100
	Spatial expansion	0.0103	0.7700	0.0230	0.7600
Whole	Sudden expansion	0.0040	0.2300	0.0003	1.0000
	Spatial expansion	0.0003	0.8500	0.0003	0.9700

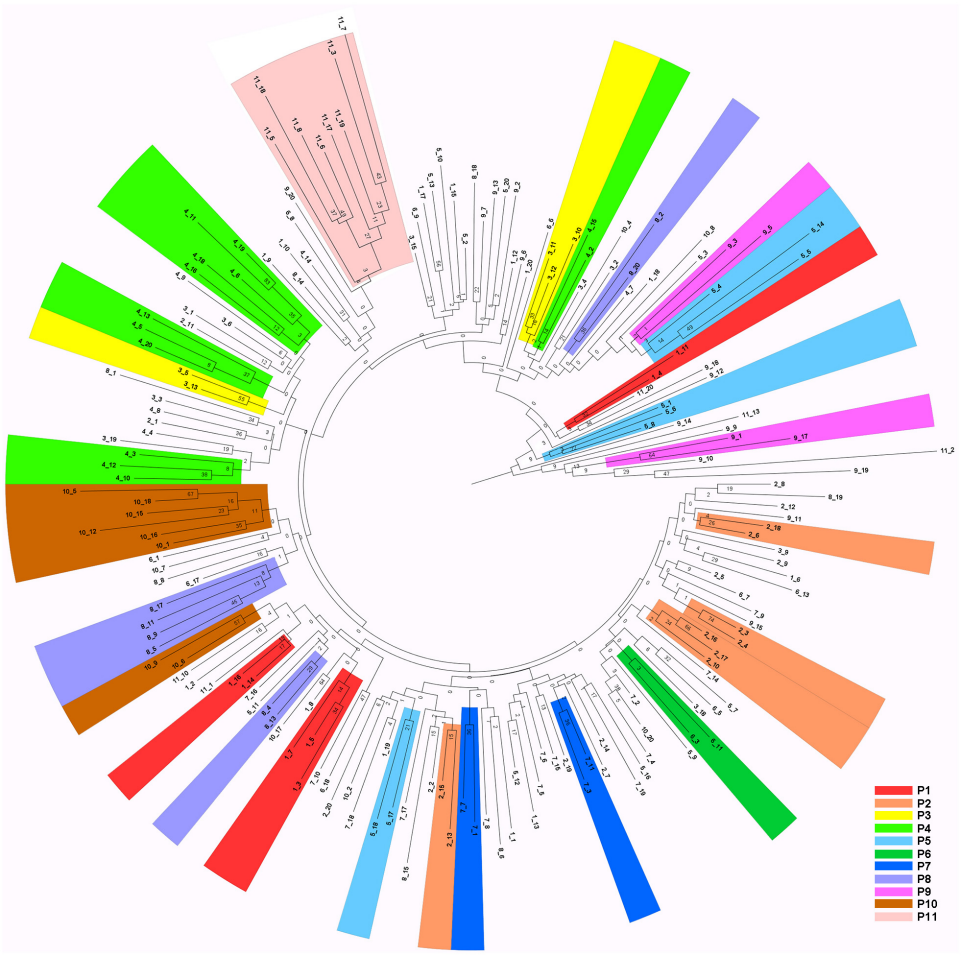


FIGURE 4 | Neighbor-joining tree derived from the concatenated sequences. Clustered two or more individuals from the same location are *painted*. Nodes are named using the sample number. Bootstrap values are shown on the *nodes*.

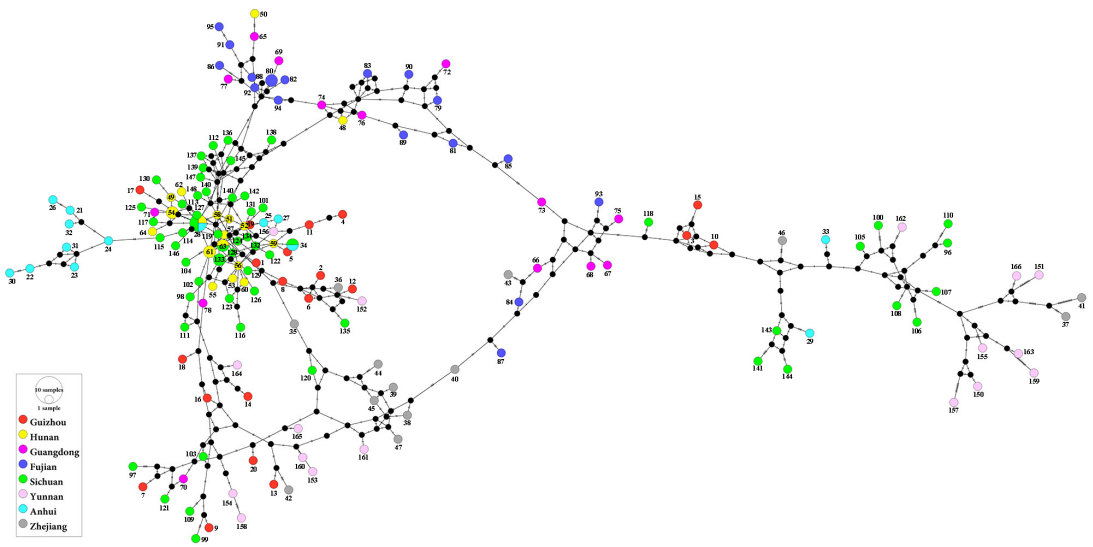


FIGURE 5 | The median-joining network of the haplotypes of *Aphis aurantii*. Haplotypes from each province were, respectively, indicated by *colored circles*.

supported its possession of a strong dispersal ability, which could have promoted the gene exchange between different populations. The ability of inhabiting a wide range of plants also provided intermediate hosts for the dispersal of *A. aurantii*. The continuous transportation of countless tea products or other plants between regions could have also accelerated the process of gene exchange.

In both the NJ tree and haplotype network diagram, no strong genetic structure was detected. Each population was widely distributed on the trees and networks, comprising different haplotypes and not forming distinct lineages. This may be explained by the long period of migration-caused gene exchange between the populations of *A. aurantii*, which reduced the degree of genetic differentiation and made it difficult to establish independently evolved lineages.

Demographic History

For the entire dataset of Chinese *A. aurantii*, the neutrality test showed insignificantly positive Tajima's *D* values and significantly negative Fu's F_s values, suggesting the probable existence of population expansion events in history (Fu, 1997). For each of the 11 sampled populations, most of the *D* values were insignificantly positive and most of the F_s values were insignificantly negative, deviating from neutrality. For the entire dataset and for each sampled population, both the tests of sum of squared deviation and Harpending's raggedness index were not significant, which suggested that the hypothesis of population expansion cannot be rejected (Harpending, 1994). The mismatch distribution of *A. aurantii* was consistent with the results of the neutrality test, being consistent with the simulated mismatch distributions of the expanded populations (Harpending et al., 1993). The obtained low Harpending's raggedness index (<0.04) was also regarded as a significant evidence of ancient expansion events (Harpending, 1994), which further supported the hypothesis of population expansion in Chinese *A. aurantii* populations.

The significant negative values of the neutrality test and the good fitting degree of the mismatch distribution (demographic expansion and spatial expansion models) were the signatures of the population expansion events and supported the hypothesis that Chinese *A. aurantii* populations have undergone a rapid population expansion during their history (Tajima, 1989; Rogers, 1995; Fu, 1997). The result of the demographic history was consistent with that of haplotype and nucleotide diversities, which also indicated that *A. aurantii* underwent a rapid population expansion.

According to the above multiple analyses, Chinese *A. aurantii* populations might have derived from a low effective ancestral population size and then expanded along with their suitable habitats, such as tea gardens or tea factories. The continuous increase of habitats and the transregional transportation of host plants provided favorable conditions for the fast expansion of the tea aphids. The dispersal and spread of *A. aurantii* into different areas generated different geographic populations, and each population gradually accumulated genetic differences due to geographic isolation. Usually, the long geographical distance could reduce the gene exchange and cause different genetic structures within populations (Taguchi et al., 2015; Zhang et al., 2019). However, the excessive use of chemical insecticides continuously caused regional extinction and loss of the genetic diversity of Chinese *A. aurantii* populations. Although geographic isolation was considered the main factor that influenced the genetic diversity of Chinese tea aphids, the gradually increased cultivation areas and frequent transportation apparently promoted the gene exchange between the different *A. aurantii* populations and led to fewer genetic differences. From the aspect of pest control, reduction of population genetic diversity by chemicals or gene exchanges is beneficial to the control of tea aphids.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession numbers can be found below: NCBI (accession: MW265715-MW265910 and MW289592-MW289792).

ETHICS STATEMENT

All experiments and procedures for this study complied with the current animal ethics guidelines and did not involve any protected animals.

AUTHOR CONTRIBUTIONS

H-LL and D-QP conceived and designed the experiments. CL and X-LW collected and identified the samples. H-LL, Z-TC, and K-JX performed the experiments, analyzed the data, and wrote the manuscript. Z-TC prepared the figures. All authors reviewed the manuscript and contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Frontier Discipline Fund of Sichuan Academy of Agricultural Sciences (2019QYXK032). Study on the disaster law and key control technology of main pests in Sichuan characteristic fruits (2021XKJS084), National Key R&D Program of China (2016YFD0200900), and the Sichuan Tea Innovation Team of National Modern Agricultural Industry Technology System (sccxtd-2020-10).

ACKNOWLEDGMENTS

We are grateful to the editor and reviewers for their helpful comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2021.702178/full#supplementary-material

REFERENCES

Bohonak, A. J. (2002). IBD (Isolation By Distance): a program for analyses of isolation by distance. *J. Hered.* 93, 153–154. doi: 10.1093/jhered/93.2.153

Boyer de Fonscolombe, E. L. J. H. (1841). Description des Pucerons qui se trouvent aux environs d'Aix. *Ann. Soc. Entomol. Fr.* 10, 157–198.

Carver, M. (1978). The black citrus aphids, *Toxoptera citricidus* (Kirkaldy) and *T. aurantii* (Boyer de Fonscolombe) (Homoptera: *Aphididae*). *Aust. J. Entomol.* 17, 263–270. doi: 10.1111/j.1440-6055.1978.tb00156.x

Excoffier, L., and Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10, 564–567. doi: 10.1111/j.1755-0998.2010.02847.x

Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147, 915–925. doi: 10.1093/genetics/147.2.915

Grant, W. S., and Bowen, B. W. (1998). Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *J. Hered.* 89, 415–426. doi: 10.1093/jhered/89.5.415

Harpending, H. C. (1994). Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum. Biol.* 66, 591–600.

Harpending, H. C., Sherry, S. T., Rogers, A. R., and Stoneking, M. (1993). The genetic structure of ancient human populations. *Curr. Anthropol.* 34, 483–496. doi: 10.1086/204195

Harry, M., Solignac, M., and Lachaise, D. (1998). Molecular evidence for parallel evolution of adaptative syndromes in Wg-breeding *Lissocephala* (Drosophilidae). *Mol. Phylogenet. Evol.* 9, 542–551. doi: 10.1006/mpev.1998.0508

Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H., and Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc. Natl. Acad. Sci.U. S. A.* 101, 14812–14817. doi: 10.1073/pnas.0406166101

Hedgecock, D., Hutchinson, E. S., Li, G., Sly, F. L., and Nelson, K. (1989). Genetic and morphometric variation in the Pacific sardine, *Sardinops sagax caerulea*: comparisons and contrasts with historical data and with variability in the northern anchovy, *Engraulis mordax*. *Fish. Bull.* 87, 653–671.

Hong, F., Mo, S. H., Liu, Y., and Wei, D. (2020). Transcriptomic profiling of various developmental stages of *Aphis aurantii* to provide a genetic resource for gene expression and SSR analysis. *Front. Physiol.* 11:578939. doi: 10.3389/fphys.2020.578939

Koch, C. L. (1854). *Die Pflanzenläuse Aphiden getreu nach dem Leben abgebildet und beschrieben*. *Nuremberg*: J.L. Lotzbeck, 1854. doi: 10.5962/bhl.title.77523

Leigh, J. W., and Bryant, D. (2015). POPART: full-feature software for haplotype network construction. *Methods Ecol. Evol.* 6, 1110–1116. doi: 10.1111/2041-210X.12410

Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187

Piron, P. G., de Haas, M. C., and Sonnemans, M. A. H. M. (2019). The presence of *Aphis (Toxoptera) aurantii* (Homoptera: Aphididae) in the Netherlands. *Entomol. Ber.* 79, 162–164.

Pu, D. Q., Liu, C., Liu, H. L., Chen, Z. T., Wu, X. L., Xiao, K. J., et al. (2020). Complete mitochondrial genome of Sichuan's population of *Aphis aurantii* (Hemiptera: Aphididae). *Mitochondrial DNA B Resour.* 5, 2119–2120. doi: 10.1080/23802359.2020.1715303

Rogers, A. R. (1995). Genetic evidence for a Pleistocene population expansion. *Evolution* 49, 608–615. doi: 10.1111/j.1558-5646.1995.tb02297.x

Schmitt, T., and Hewitt, G. M. (2004). The genetic pattern of population threat and loss: a case study of butterflies. *Mol. Ecol.* 13, 21–31. doi: 10.1046/j.1365-294x.2004.02020.x

Slatkin, M., and Hudson, R. H. (1991). Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129, 555–562.

Slatkin, M., and Maddison, W. P. (1989). A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* 123, 603–613.

Taguchi, M., King, J. R., Wetklo, M., Withler, R. E., and Yokawa, K. (2015). Population genetic structure and demographic history of Pacific blue sharks (*Prionace glauca*) inferred from mitochondrial DNA analysis. *Mar. Freshw. Res.* 66, 267–275. doi: 10.1071/MF14075

Tahori, A., and Hazan, A. (1970). Rearing of the black citrus aphid *Toxoptera aurantii* on chemically defined diets. *J. Insect Physiol.* 16, 1975–1981. doi: 10.1016/0022-1910(70)90242-8

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595. doi: 10.1093/genetics/123.3.585

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Wang, J. F., and Qiao, G. X. (2009). DNA barcoding of genus *Toxoptera* Koch (Hemiptera: Aphididae): identification and molecular phylogeny inferred from mitochondrial COI sequences. *Insect Sci.* 16, 475–484. doi: 10.1111/j.1744-7917.2009.01270.x

Wang, Y., Ding, M., Du, Y. M., and Huang, A. J. (2019). Phylogenetic relationship and characterization of the complete mitochondrial genome of the black citrus aphid, *Aphis aurantii* (Hemiptera: Aphididae). *Mitochondrial DNA B Resour.* 4, 3567–3568. doi: 10.1080/23802359.2019.1674208

Wright, S. (1990). Evolution in Mendelian populations. *Bull. Math. Biol.* 52, 241–295.

Zhang, X. M., Zhang, X. M., Song, N., Gao, T. X., and Zhao, L. L. (2019). Study on population genetics of *Sillago aeolus* (Perciformes: Sillaginidae) in the coast of China. *Mitochondrial DNA A DNA Mapp. Seq. Anal.* 30, 825–834. doi: 10.1080/24701394.2019.1670820

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 23 July 2021
doi: 10.3389/fevo.2021.683722

Endemism Patterns of Planthoppers (Fulgoroidea) in China

*Zhengxue Zhao[1], Lin Yang[1], Jiankun Long[2], Zhimin Chang[2], Zhengxiang Zhou[1], Yan Zhi[1], Liangjing Yang[1], Hongxing Li[1], Yongjin Sui[1], Nian Gong[1], Xiaoya Wang[1] and Xiangsheng Chen[1]**

[1] Provincial Special Key Laboratory for Development and Utilization of Insect Resources of Guizhou, Guizhou Key Laboratory for Agricultural Pest Management of Mountainous Region, Institute of Entomology, Guizhou University, Guiyang, China, [2] College of Animal Science, Guizhou University, Guiyang, China

Studies on endemism are always of high interest in biogeography and contribute to better understanding of the evolution of species and making conservation plans. The present study aimed to investigate the endemism patterns of planthoppers in China by delimiting centers of endemism and areas of endemism. We collected 6,907 spatial distribution records for 860 endemic planthopper species from various resources. Centers of endemism were identified using weighted endemism values at 1° grid size. Parsimony analysis of endemicity and endemicity analysis were employed to detect areas of endemism at 1°, 1.5°, and 2° grid sizes. Six centers of endemism located in mountainous areas were identified: Taiwan Island, Hainan Island, eastern Yungui Plateau, Wuyi Mountains, western Qinling Mountains, and western Yunnan. We also delimited six areas of endemism, which were generally consistent with centers of endemism. Our findings demonstrated that mountainous areas have an essential role in facilitating the high level of endemism and formation of areas of endemism in planthoppers through the combined effects of complex topography, a long-term stable environment, and geological events. Dispersal ability and distribution of host plants also have important effects on the patterns of planthoppers' endemism.

Keywords: areas of endemism, biogeography, centers of endemism, mountainous areas, planthoppers

OPEN ACCESS

Edited by:
Gengping Zhu,
Tianjin Normal University, China

Reviewed by:
Yujian Li,
Qufu Normal University, China
Gang Feng,
Inner Mongolia University, China

****Correspondence:***
Xiangsheng Chen
chenxs3218@163.com

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *22 March 2021*
Accepted: *24 June 2021*
Published: *23 July 2021*

Citation:
Zhao Z, Yang L, Long J, Chang Z, Zhou Z, Zhi Y, Yang L, Li H, Sui Y, Gong N, Wang X and Chen X (2021) Endemism Patterns of Planthoppers (Fulgoroidea) in China. Front. Ecol. Evol. 9:683722. doi: 10.3389/fevo.2021.683722

INTRODUCTION

The geographical distribution of endemic species (i.e., restriction of a species to a particular area) represents the highest degree of historical and ecological imprint of all biological entities (Casazza et al., 2008). Moreover, endemic areas have been generally recognized as priority areas for biodiversity conservation plans (Myers et al., 2000; Lamoreux et al., 2006; Huang et al., 2010, 2016; Gomes-da-Silva et al., 2017; Zhao et al., 2020a). Therefore, studies related to patterns of endemism have always been a central topic of biogeography and biodiversity conservation (Laffan and Crisp, 2003; Orme et al., 2005; Posadas et al., 2006; Sandel et al., 2011; Huang et al., 2012; Feng et al., 2016; Li et al., 2017).

Studies on the spatial patterns of endemics often delimit the main distribution areas, and these areas are frequently described as centers of endemism (CoEs) and areas of endemism (AoEs). Although the drivers causing formation of CoEs and AoEs may show similar ecological or

historical legacies (Hurdu et al., 2016), they differ because the two concepts are based on different identification methods and assumptions (Linder, 2001a). CoEs are areas including higher endemic richness or endemism than its surroundings (Crisp et al., 2001; Linder, 2001b). AoEs represent areas delimited by the congruent distribution of at least two endemic taxa (spatial homology) (Platnick, 1991; Morrone, 1994; Szumik et al., 2002). AoEs are extremely important because they can be used to design biogeographic regionalization schemes (Morrone, 2014a), to investigate organism-climate dynamics (Gámez et al., 2017), and as a priority target for conservation plans (Huang et al., 2010; Noroozi et al., 2018). Numerous biogeographers and evolutionary biologists have shown interest in evaluating the causes for the presence of CoEs and AoEs (Nelson and Platnick, 1981; Anderson, 1994; López-Pujol et al., 2011; Yuan et al., 2014; Noroozi et al., 2018).

Mountainous areas host a remarkable proportion of Earth's biodiversity, with several species completely restricted to these areas, although these areas only account for 16.5–27% of the land area (Fjeldså et al., 2012). Half of the high biodiversity areas identified so far are mountainous areas (Kohler and Maselli, 2009). The large biodiversity in mountainous areas is associated with their dual role as "species museums" (places of especially long-term persistence) and "species cradles" (places of especially rapid speciation). Therefore, high biodiversity in mountainous areas reflects two mechanisms: enhanced speciation rates and lineage persistence (Rahbek et al., 2019). Mountain biodiversity is characterized by deep-time evolution and ecological processes, which reflect a history worth protecting (Rahbek et al., 2019).

The planthoppers, members of superfamily Fulgoroidea (Insecta: Hemiptera), are a dominant group of herbivorous insects, consisting of 13,844 species in 33 families worldwide (Bourgoin, 2021). Planthoppers are obligatory phytophagous and mainly feed on the phloem tissue of woody or herbaceous plants (Wilson, 2005). Thus, host plants might have highly impacted their distribution. The majority of species in planthoppers lack the ability to disperse over long distances, resulting in a small distribution range. This property makes it a unique and excellent model for studying biogeography (Liang, 1998). China is one of the countries that is the richest in planthoppers, with over 1,300 described species. Unfortunately, although previous studies have investigated species richness patterns in planthoppers (Zhao et al., 2020b), the understanding of patterns of endemism in China remains in its infancy. Thus, to fully understand endemism patterns of Chinese planthoppers, there is a need to compile distributional data for all endemic planthoppers species. Research on spatial patterns of endemism not only contributes to understanding the evolution of planthoppers but also for identification of priority areas of planthopper biodiversity conservation.

The aim of the present study was to propose biogeographical patterns of planthopper endemism in China using the spatial distribution data for endemic species. We focused on the following research areas: (1) identifying the CoEs and delimiting AoEs; and (2) whether or not the CoEs and AoEs in this study are located in mountainous areas, as found in previous studies (e.g., Wang et al., 2017; Noroozi et al., 2018, 2019).

MATERIALS AND METHODS

Species Distribution Data

The 6,907 spatial distribution records for 860 endemic planthoppers species (only recorded in China) were collected from the literature, books, MD/Ph.D. theses, zoological records, China Knowledge Resource Integrated Database, and museum specimens (Zoological Museum, Institute of Zoology, Chinese Academy of Sciences; Institute of Entomology, Guizhou University; the insect collection of China Agricultural University, Hebei University, Nankai University and Dali University; Entomological Museum, Northwestern A & F University; Shanghai Institute of Entomology, Chinese Academy of Sciences; Tianjin Museum of Natural History; Taiwan Agricultural Research Institute). Distribution sites at the city, county, or township levels were obtained from the original source. The distribution sites containing the latitude and longitude information were used directly, and distribution sites without the latitude and longitude information were expressed by the latitude and longitude of the corresponding administrative center.

Assessing Sampling Bias and Mapping Endemism Patterns

The inventory completeness for the study region was accessed using the species accumulation curve from the incidence-based bootstrap estimators (Zhao et al., 2020b). The presence (1) or absence (0) matrix for each endemic species in each 1° grid (∼100 × 100 km) was constructed and was analyzed using EstimateS 9.1 with 100 randomizations (Colwell, 2013). To assess the completeness of the richness of each 1° grid, we fitted a linear regression using the square-root transformed number of records and number of richness per grid. Spatial autocorrelation may inflate the rate of type I error (Diniz-Filho et al., 2003), so the *P*-value for linear regression was reported using geographically effective degrees of freedom (Dutilleul et al., 1993), evaluated using Spatial Analysis in Macroecology 4.0 (Rangel et al., 2010). The patterns of endemism were visualized by calculating the weighted endemism values of each 1° grid. This process was conducted using Biodiverse 2.1 (Laffan et al., 2010). We defined the center of endemism, based on the criterion proposed by Crisp et al. (2001) and Linder (2001b).

Identifying Areas of Endemism

Parsimony analysis of endemicity (PAE) and endemicity analysis (EA) were used to identify AoEs. The performance of the two approaches was compared using the two criteria proposed by Escalante et al. (2009), namely the number of AoEs they delimit and the number of endemic species supporting them. Three grid sizes were used in two approaches: 1° × 1°, 1.5° × 1.5°, and 2° × 2°.

To perform PAE, we constructed presence (1) or absence (0) matrix based on the presence of each endemic species in each grid. A hypothetical outgroup "Root" with all zeros was added to the matrix to root the resulting tree (Morrone, 1994, 2014b). All matrices were performed in TNT v1.1 (Goloboff et al., 2008). New Technology algorithms were based on maximum trees with 1,000

and used sectorial search and tree fusing, with 10 initial addseqs. We obtained a strict consensus tree for each analysis. The relative support for each branch was estimated using bootstrap analysis with 1,000 replicates (Felsenstein, 1985). Branches with bootstrap values ≥ 50% were selected as candidates for AoEs. Finally, AoEs (clades of grids), containing at least two species restricted by these areas, were delimited and mapped.

EA was conducted using NDM/VNDM v 3.1 (Goloboff, 2016) under three grid sizes. This method is used to identify AoEs using an optimality criterion, which explicitly considers the spatial location of the species in a given area (Szumik et al., 2002). The parameters used here were performed by saving temporary sets within 0.99 of the current score; sets containing at least two endemic species and scores above 3.0 were saved. The search was repeated 10 times, keeping overlapping subsets when 60% of their defining species were unique. From the sets obtained, we chose species with a minimum score of 0.5. The strict rule was used to calculate the consensus areas at a cut-off of 40% similarity in species. To obtain the final AoEs, the consensus areas among the different grid sizes were overlapped (do Prado et al., 2015; Gao et al., 2018; Du et al., 2020).

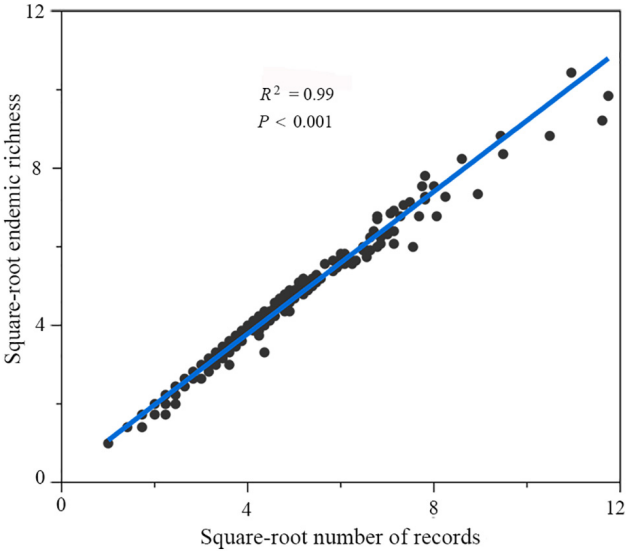


FIGURE 2 | Linear regression (y = 0.905x + 0.173) for number of records and endemic richness in 1° grid. The observed richness in the grids below the regression line is >78.8% of the predicted richness.

RESULTS

The species accumulation curve using bootstrap estimators obtained the "true" species number, namely 1,006 (**Figure 1**), and the data completeness degree is 85.4%. Furthermore, the ratio of observed species richness to those predicted by linear regression models for each grid cell was >78.8% (**Figure 2**). These results show that planthoppers were adequately sampled.

Six CoEs were identified (**Figure 3**), located in Taiwan Island, Hainan Island, eastern Yungui Plateau, Wuyi Mountains, western Qinling Mountains, and western Yunnan. A total of three AoEs were identified using three grid sizes through PAE analysis (**Figure 4**). These areas are supported by single or multiple grid sizes. Taiwan Island and Hainan Island was detected under all three grid sizes. Western Qinling Mountain was detected to be AoEs only at 2°grid size. The AoEs under each grid size and the species supporting them were identified (**Supplementary Table 1**).

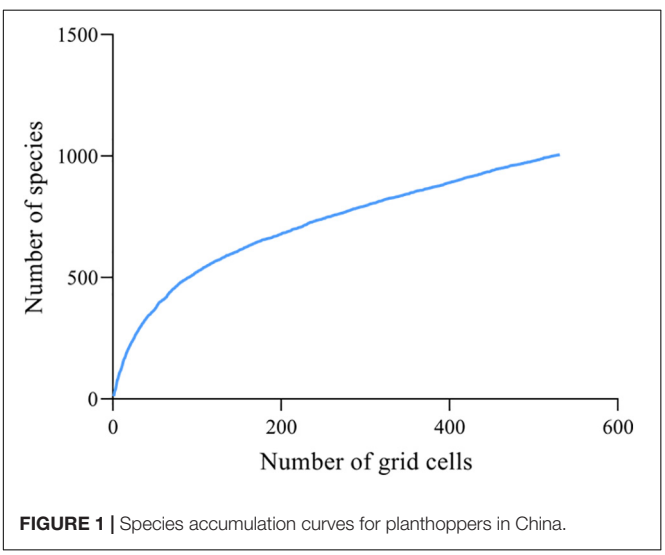


FIGURE 1 | Species accumulation curves for planthoppers in China.

The consensus areas and supported endemic species obtained by EA in each grid size are summarized in **Supplementary Material**. Analysis using 1°, 1.5°, and 2° grid sizes yielded 4, 10, and 9 consensus areas, respectively. Through an overlapping pattern of the consensus areas in different grid sizes, five AoEs were identified (**Figure 5**).

Taiwan Island was delimited under all three grid sizes, containing the consensus areas 2, 3, 4, 11, 14, 18, and 23 (**Supplementary Figures 1–3**). Consensus areas 2, 3, and 4 in the 1° grid size, with the corresponding 35, 12, and 6 species, were found. Consensus areas 11 and 14 were detected in the 1.5° grid size and contained 36 and 9 species, respectively. Consensus areas 18 and 23 were singled out in the 2° grid size and included 38 and 9 species. Two species, *Kuvera taiwana* and *Cixius hopponis*, support consensus areas 2, 3, 11, 14, 18, and 23 (**Supplementary Table 2**). *Cixius wui* supports consensus areas 2, 3, 14, 18, and 23. Four consensus areas with the greatest endemicity values appeared in this area, including consensus areas 2, 3, 11, and 18, with scores of 24.08333, 8.65, 26.4, and 26.3, respectively. Hainan Island was detected to be an AoE in 1° and 1.5° grid sizes and has the highest value (score = 8.4), except Taiwan Island. Consensus area 1 was present in 1° grid size and contained 11 species. Consensus area 9 was detected in 1.5° grid size and was supported by 10 species. These two consensus areas were simultaneously supported by *Deferunda trimaculata*, *Deferunda acuminate*, *Deferunda striat*, *Epora biprolata*, *Gnezdilovius pseudotesselatus*, *Gnezdilovius multipunctatus*, *Neogergithoides tubercularis*, and *Sarimodes clavatus* (**Supplementary Table 2**).

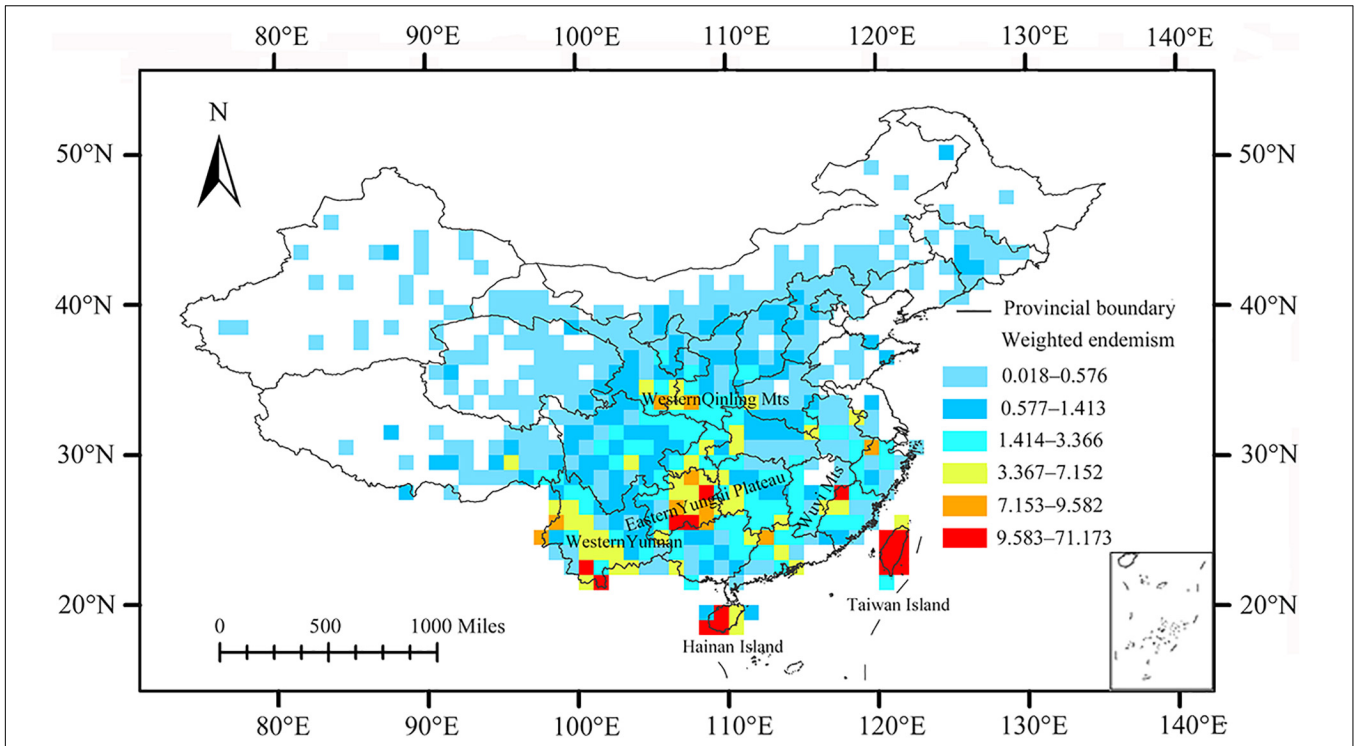


FIGURE 3 | Weighted endemism richness patterns of planthoppers in 1° grid in China.

Western Yunnan, covered by the Ailao, Wuliang, and southern Hengduan Mountains, included consensus areas 6, 7, 10, and 13 in 1.5° grid size and consensus area 15, 19, 21, and 22 in 2° grid size (**Supplementary Figures 2, 3**). *Mongoliana sinuate*, *Paracatonidia webbeda*, *Thabena lanpingensis*, and *Vekunta triprotrusa* provide support of consensus area 15, 21, and 22. Eastern Yungui Plateau (roughly contains the Dalou, Wuling, and Miaoling) was detected in 1.5° and 2° grid sizes (**Supplementary Figures 2, 3**), containing consensus areas 5 and 8 in 1.5° grid size, consensus areas 16 and 17 in 2° grid size. Three species, *Fusiissus frontomaculatus*, *Gergithoides caudospinosus*, and *Usana fissure*, define consensus areas 5, 8, and 17. Wuyi Mountains comprised consensus area 12 in 1.5° grid size and consensus area 20 in 2° grid size. *Errada dimidiate*, *Fortunia belostoma*, *Neoalcathous wuyishanana*, and *Ricanula fujianensi* support these two consensus areas. *Geisha bifurcata* and *Neokodaiana minensis* also support consensus area 20.

DISCUSSION

Causes of High Endemism and AoEs

In this study, we found that all CoEs and AoEs were located in mountainous areas, consistent with the findings in previous studies for insects, such as leafhoppers, scale insects, and aphids (Yuan et al., 2014; Wang et al., 2017; Gao et al., 2018), birds (Lei et al., 2003; Huang et al., 2010), mammals (Tang et al., 2006), and plants (Huang et al., 2012; Noroozi et al., 2019). The topographic complexity, long-term stable environment, and geological events experienced in mountainous areas are generally considered to be the main driving forces for the high level of endemism and formation of AoEs in these areas (Tribsch, 2004; López-Pujol et al., 2011; Wang et al., 2017; Gao et al., 2018; Noroozi et al., 2018, 2019).

With increasing topographic complexity, the amount of habitat diversity is expected to increase, promoting available niche space, and thereby increasing ecological speciation via adaptation to different niches (Rundle and Nosil, 2005; Hendry et al., 2007) and the coexistence of species (Tews et al., 2004; Hortal et al., 2009). Furthermore, topographical complexity can act as a barrier to gene-flow among diverging populations, resulting in supporting reproductive isolation, increasing differentiation, and speciation (Gillespie and Roderick, 2014), and ultimately high endemism. Mountainous areas have long-term climate stability, which is conducive to the persistence of relict lineages, specialization, speciation of small-ranged species, and reduction of extinction probability (Fjeldså and Lovett, 1997; Dynesius and Jansson, 2000; Jansson, 2003; López-Pujol et al., 2011), thus further increasing diversification.

High level of endemism in mountainous areas is also significantly related to the geological events they have experienced. Orogenic processes promote the emergence of new lineages and radiation of species due to the repeated formation, connectivity, and disappearance of habitats within mountain ranges (Rahbek et al., 2019). We found that Taiwan Island and Hainan Island had the first and second highest weighted endemism values and endemicity scores, respectively. Taiwan Island and Hainan Island are China's largest and second

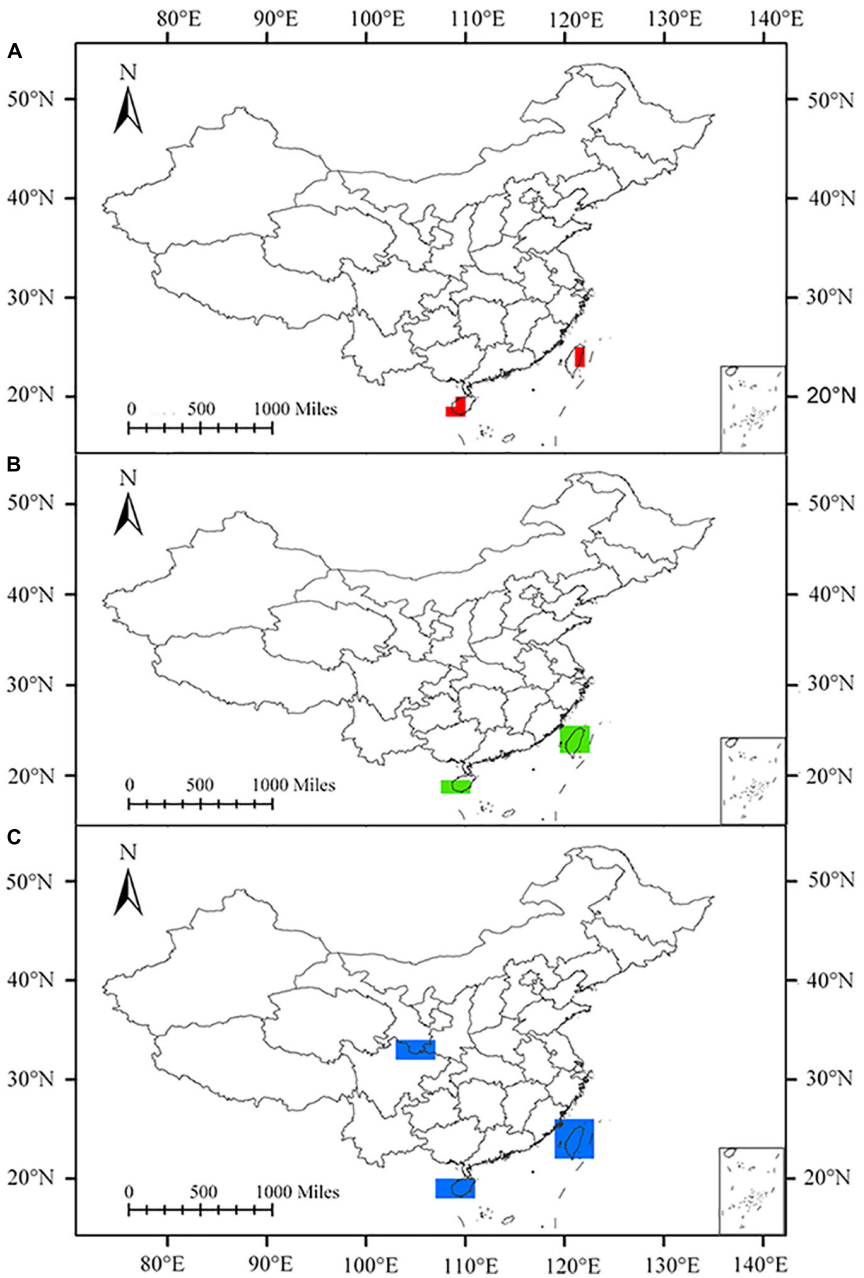


FIGURE 4 | Areas of endemism (AoEs) obtained from planthoppers by parsimony analysis of endemicity (PAE) using three grids in China. **(A)** 1° grid, **(B)** 1.5° grid, and **(C)** 2° grid.

largest islands, respectively, with typical mountain habitats affected by tropical/subtropical climates. The highest mountain in Taiwan is Mt. Yunshan (3,997 m), which is also the highest mountain in southeastern China. The highest mountain of Hainan Island is Mt. Wuzhishan. Taiwan Island and Hainan Island are isolated from the mainland, generating limited interchange between the islands and the mainland (Yuan et al., 2014). Furthermore, volcanoes often erupted on Hainan Island during the Pleistocene (Wang, 1985), causing populations to be isolated on the local and over decadal timescales, thereby increasing the development of different gene pools and the appearance of new genetic combinations (Gillespie and Roderick, 2014). These events were beneficial for promoting the emergence of novel lineages and a high degree of endemism in the two islands. The levels of endemicity in Taiwan Island are significantly higher than Hainan Island, probably because it is farther away from the mainland and had formed and separated earlier.

Planthoppers are very weak flyers and move by short-distance jumping (Liang, 1998). Therefore, planthoppers

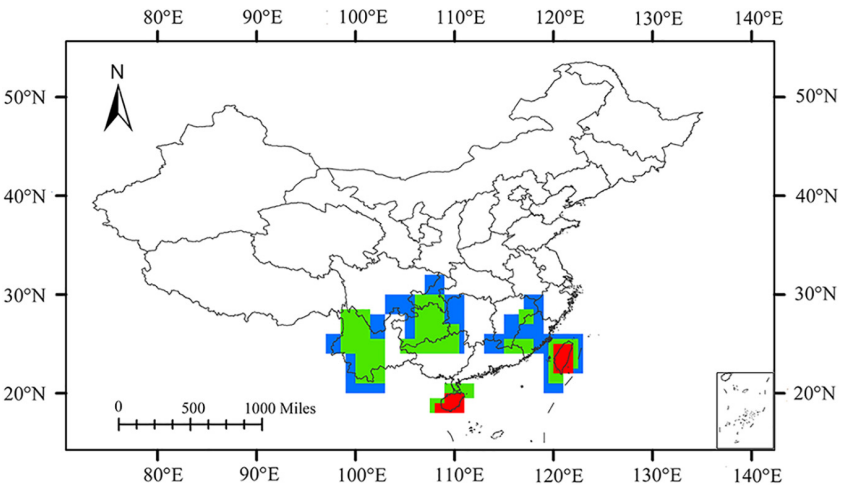


FIGURE 5 | Overlap of the consensus areas of endemism (AoEs) delimited for planthoppers by endemicity analysis (EA) employing 1° (red), 1.5° (green), and 2° (blue) grids in China.

lack a strong ability to disperse. This characteristic may allow them to remain in places of speciation and glacial refuges, particularly when these places have high altitudes and complex terrain, and thus build high endemic species numbers and AoEs over time. The vast majority of various planthoppers feed on woody or herbaceous plants (Wilson, 2005), such as Poaceae, Cyperaceae, Rosaceae, Oleaceae, and Polygonaceae (Ding, 2006; Chen et al., 2014). Host plants, apart from being a food source, are also used by some planthoppers for mating, oviposition, protection in winter, and avoiding their natural enemies (Wilson et al., 1994). Many planthoppers have high degrees of specialization with respect to host plants (Wilson et al., 1994; Liang, 1998). The family Delphacidae, which has the largest number of species among planthoppers, mainly feeds on a small number of host plants species (concentrated to Poaceae and Cyperaceae), and some even feed only on a single host plant species (Wilson et al., 1994; Ding, 2006). Most of the species in the family Caliscelidae feed on bamboo and a few on other plants (Chen et al., 2014). In China, we found that the diversity centers of many host plants are consistent with the CoEs/AoEs identified in the present study. For example, high bamboo diversity is mainly concentrated in southern Yunnan and Wuyi Mountains (Xu et al., 2019). There is high species richness of Rosaceae in northern Yunnan and the Qinling Mountains (Zou et al., 2019). This consistency shows that the distribution pattern of planthoppers is highly associated with the diversity patterns of host plants.

The Effects of Grid Size and Approach on Results

PAE and EA use grids as basic operating units in delimiting AoEs; therefore, the grid size affects the results in a deterministic manner. Grid size can affect the number, dispersion, and range of the localities. Smaller grid sizes will produce narrower AoEs, allowing a small number of endemic species to be included, whereas larger grid sizes identify wider AoEs and accommodate more endemic species (Casagranda et al., 2009; do Prado et al., 2015; DaSilva et al., 2015).

We found the greater number of AoEs and supported endemic species were identified by EA at larger grid cell (1.5° and 2°), as was PAE (2°). EA produced five and four AoEs in 1.5° and 2° grids and PAE delimited three AoEs in 2° grids. Combining different grid sizes is necessary to define the AoEs. Using several grid sizes facilitates the discovery of different AoEs in regions with topographical complexity (Elías and Aagesen, 2016), such as western Yunnan, and eastern Yungui Plateau is only delimited by at 1.5° and 2° grids. AoEs that recovered under different grid sizes were more robustly and clearly supported by data (Aagesen et al., 2009, 2012; Szumik et al., 2012). Additionally, the use of several grids sizes was slightly different in location, area, and boundary of the resulting AoEs, and minimized the effects of sampling biases (DaSilva et al., 2015). The use of grids brings has some drawbacks, such as the inability to identify AoEs smaller than grid sizes and difficulty in representing the fuzzy edges of the AoEs. Geographical Interpolation of Endemism (GIE) (Oliveira et al., 2015), an approach independent of grids, can be valuably used in future work to obtain more detailed information about the AoEs of planthoppers. The performance of PAE and EA has been widely discussed using real or hypothetical distributions (e.g., Moline and Linder, 2006; Carine et al., 2009; Casagranda et al., 2012; Escalante, 2015). However, there is no consensus on which approach is most suitable and their exploration should be continued (Escalante, 2015). Using a total of three grids, we found that EA identified more AoEs than PAE. The observed difference is because species that support the AoEs in PAE require higher and stricter sympatry, resulting in a small number of AoEs. Furthermore, western Qinling Mountains delimited by the PAE are not identified in the EA. In this way, combining the two methods may provide a more comprehensive knowledge of AoEs.

CONCLUSION

The spatial structure of planthoppers' endemism in China was analyzed by identifying CoEs and AoEs. We found that the CoEs and AoEs identified fell in the mountainous areas, mainly due to their complex topography, stable eco-climatic conditions, and related geological events that promote long-term persistence, speciation, and species accumulation. Furthermore, we found that the dispersal ability of planthoppers and diversity pattern of host plants also are responsible for high planthoppers endemism in the mountainous areas. Future studies can add temporal information based on phylogenetic information from endemic species to better interpret AoEs.

DATA AVAILABILITY STATEMENT

The original data of this study are available from the corresponding author upon reasonable request.

AUTHOR CONTRIBUTIONS

ZZha analyzed the data and wrote the article. LNY, JL, ZC, ZZho, YZ, LAY, HL, YS, NG, and XW collected the data. XC revised the article. All authors contributed significantly to the drafts and gave final approval for publication.

FUNDING

This research was funded by the National Natural Science Foundation of China (Nos. 31472033 and 31601886), the Science and Technology Support Program of Guizhou Province (No. 20201Y129), the Program of Excellent Innovation Talents, Guizhou Province, China (No. [2015]4021), the Program of Scientific Research Foundation for Introduced Talent of Guizhou University (No. Guidarenjihezi [2017]18), the Science and Technology Program of Guizhou Province (No. Qiankehejichu [2018]1031), Academic New Cultivation and Innovation Exploration Special Project of Guizhou University in 2018 (No. Qiankehe and Platform for talents [2018]5781-29), and the Science and Technology Program of Guizhou Province (No. Qiankehejichu [2018]1031).

ACKNOWLEDGMENTS

We thank Xiaolei Huang, Chao Gao, and Yang Li for their help in PAE. We are also very grateful to all collectors for collecting specimens.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2021.683722/full#supplementary-material

REFERENCES

Aagesen, L., Bena, M. J., Nomdedeu, S., Panizza, A., López, R. P., and Zuloaga, F. O. (2012). Areas of endemism in the southern central Andes. *Darwiniana* 50, 218–251.

Aagesen, L., Szumik, C. A., Zuloaga, F. O., and Morrone, O. (2009). Quantitative biogeography in the South America highlands-recognizing the Altoandina, Puna and Prepuna through the study of Poaceae. *Cladistics* 25, 295–310. doi: 10.1111/j.1096-0031.2009.00248.x

Anderson, S. (1994). Area and endemism. *Q. Rev. Biol.* 69, 451–471. doi: 10.1086/418743

Bourgoin, T. (2021). *FLOW (Fulgoromorpha Lists on The Web): A World Knowledge Base Dedicated To Fulgoromorpha. Version 8*. Available online at: http://hemiptera-databases. org/flow/ (accessed on February 26, 2021)

Carine, M. A., Humphries, C. J., Guma, I. R., Reyesbetancort, J. A., and Guerra, A. S. (2009). Areas and algorithms: evaluating numerical approaches for the delimitation of areas of endemism in the Canary Islands archipelago. *J. Biogeogr.* 36, 593–611. doi: 10.1111/j.1365-2699.2008.02016.x

Casagranda, M. D., Taher, L., and Szumik, C. A. (2009). Endemismo a diferentes escalas espaciales: un ejemplo con Carabidae (Coleoptera: Insecta) de América del Sur austral. *Rev. Chil. Hist. Nat.* 82, 17–42.

Casagranda, M. D., Taher, L., and Szumik, C. A. (2012). Endemicity analysis, parsimony and biotic elements: a formal comparison using hypothetical distributions. *Cladistics* 28, 645–654. doi: 10.1111/j.1096-0031.2012.00410.x

Casazza, G., Zappa, E., Mariotti, M., Medail, F., and Minuto, L. (2008). Ecological and historical factors affecting distribution pattern and richness of endemic plant species: the case of the Maritime and Ligurian Alps hotspot. *Divers. Distrib.* 14, 47–58. doi: 10.1111/j.1472-4642.2007.00412.x

Chen, X. S., Zhang, Z. G., and Chang, Z. M. (2014). *Issidae and Caliscelidae (Hemiptera: Fulgoroidea) from China*. Guiyang: Guizhou Science and Technology Publishing House.

Colwell, R. K. (2013). *EstimateS: Statistical Estimation of Species Richness and Shared Species from Samples*. Storrs, Connecticut. Available online at: http://purl.oclc.org/estimates (accessed on August 16, 2020)

Crisp, M. D., Laffan, S. W., Linder, H. P., and Monro, A. (2001). Endemism in the Australian flora. *J. Biogeogr.* 28, 183–198. doi: 10.1046/j.1365-2699.2001.00524.x

DaSilva, M. B., Pintodarocha, R., and Desouza, A. (2015). A protocol for the delimitation of areas of endemism and the historical regionalization of the Brazilian Atlantic Rain Forest using harvestmen distribution data. *Cladistics* 31, 692–705. doi: 10.1111/cla.12121

Ding, J. H. (2006). *Fauna Sinica (Insecta Vol. 45, Homoptera, Delphacidae)*. Beijing: Science Press.

Diniz-Filho, J. A. F., Bini, L. M., and Hawkins, B. A. (2003). Spatial autocorrelation and red herrings in geographical ecology. *Global Ecol. Biogeogr.* 12, 53–64. doi: 10.1046/j.1466-822x.2003.00322.x

do Prado, J. R., Brennand, P. G., Godoy, L. P., Libardi, G. S., De Abreujunior, E. F., Roth, P. R., et al. (2015). Species richness and areas of endemism of oryzomyine rodents (Cricetidae. Sigmodontinae) in South America: an NDM/VNDM approach. *J. Biogeogr.* 42, 540–551. doi: 10.1111/jbi.12424

Du, C. C., Chen, J., Jiang, L. Y., and Qiao, G. X. (2020). High correlation of species diversity patterns between specialist herbivorous insects and their specific hosts. *J. Biogeogr.* 47, 1232–1245. doi: 10.1111/jbi.13816

Dutilleul, P., Clifford, P., Richardson, S., and Hemon, D. (1993). Modifying the t test for assessing the correlation between two spatial processes. *Biometrics* 49, 305–314. doi: 10.2307/2532625

Dynesius, M., and Jansson, R. (2000). Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *P. Natl. Acad. Sci. U. S. A.* 97, 9115–9120. doi: 10.1073/pnas.97.16.9115

Elías, G. D., and Aagesen, L. (2016). Areas of vascular plants endemism in the Monte desert (Argentina). *Phytotaxa* 266, 161–182. doi: 10.11646/phytotaxa.266.3.1

Escalante, T. (2015). Parsimony analysis of endemicity and analysis of endemicity: A fair comparison. *Syst. Biodivers.* 13, 413–418. doi: 10.1080/14772000.2015.1046966

Escalante, T., Szumik, C. A., and Morrone, J. J. (2009). Areas of endemism of Mexican mammals: Reanalysis applying the optimality criterion. *Biol. J. Linn. Soc.* 98, 468–478. doi: 10.1111/j.1095-8312.2009.01293.x

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.2307/2408678

Feng, G., Mao, L. F., Sandel, B., Swenson, N. G., and Svenning, J. C. (2016). High plant endemism in China is partially linked to reduced glacial-interglacial climate change. *J. Biogeogr.* 43, 145–154. doi: 10.1111/jbi.12613

Fjeldså, J., and Lovett, J. C. (1997). Geographical patterns of old and young species in African forest biota: the significance of specific montane areas as evolutionary centres. *Biodivers. Conserv.* 6, 325–346.

Fjeldså, J., Bowie, R. C., and Rahbek, C. (2012). The role of mountain ranges in the diversification of birds. *Annu. Rev. Ecol. Evol. S.* 43, 249–265. doi: 10.1146/annurev-ecolsys-102710-145113

Gámez, N., Nihei, S. S., Scheinvar, E., and Morrone, J. J. (2017). A temporally dynamic approach for cladistic biogeography and the processes underlying the biogeographic patterns of North American deserts. *J. Zoolog. Syst. Evol. Res.* 55, 11–18. doi: 10.1111/jzs.12142

Gao, C., Chen, J., Li, Y., Jiang, L. Y., and Qiao, G. X. (2018). Congruent patterns between species richness and areas of endemism of the Greenideinae aphids (Hemiptera: Aphididae) revealed by global-scale data. *Zool. J. Linn. Soc-Lond.* 83, 791–807. doi: 10.1093/zoolinnean/zlx092

Gillespie, R. G., and Roderick, G. K. (2014). Evolution: geology and climate drive diversification. *Nature* 509, 297–298. doi: 10.1038/509297a

Goloboff, P. (2016). *NDM and VNDM: programs for the identification of areas of endemism, version. 3.1. Program and documentation*. Available online at: http://www.lillo.org.ar/ phylogeny (accessed on June 8, 2019)

Goloboff, P. A., Farris, J. S., and Nixon, K. C. (2008). TNT, a free program for phylogenetic analysis. *Cladistics* 24, 774–786. doi: 10.1111/j.1096-0031.2008.00217.x

Gomes-da-Silva, J., Amorim, A. M., and Forzza, R. C. (2017). Distribution of the xeric clade species of pitcairnioideae (Bromeliaceae) in south America: a perspective based on areas of endemism. *J. Biogeogr.* 44, 1994–2006. doi: 10.1111/jbi.12990

Hendry, A. P., Nosil, P., and Rieseberg, L. H. (2007). The speed of ecological speciation. *Funct. Ecol.* 21, 455–464. doi: 10.1111/j.1365-2435.2007.01240.x

Hortal, J., Triantis, K. A., Meiri, S., Thebault, E., and Sfenthourakis, S. (2009). Island species richness increases with habitat diversity. *Am. Nat.* 174, 205–217.

Huang, J. H., Chen, B., Liu, C. R., Lai, J. S., Zhang, J. L., and Ma, K. P. (2012). Identifying hotspots of endemic woody seed plant diversity in China. *Divers. Distrib.* 18, 673–688. doi: 10.1111/j.1472-4642.2011.00845.x

Huang, J. H., Huang, J. H., Lu, X. H., and Ma, K. P. (2016). Diversity distribution patterns of Chinese endemic seed plant species and their implications for conservation planning. *Sci. Rep.* 6:33913.

Huang, X. L., Qiao, G. X., and Lei, F. M. (2010). Use of parsimony analysis to identify areas of endemism of Chinese birds: implications for conservation and biogeography. *Int. J. Mol. Sci.* 11, 2097–2108. doi: 10.3390/ijms11052097

Hurdu, B., Escalante, T., Puşcaş, M., Novikoff, A. V., Bartha, L., and Zimmermann, N. E. (2016). Exploring the different facets of plant endemism in the South−Eastern Carpathians: a manifold approach for the determination of biotic elements, centres and areas of endemism. *Biol. J. Linn. Soc.* 119, 649–672. doi: 10.1111/bij.12902

Jansson, R. (2003). Global patterns in endemism explained by past climatic change. *P. Roy. Soc. B Biol. Sci.* 270, 583–590. doi: 10.1098/rspb.2002.2283

Kohler, T., and Maselli, D. (Eds). (2009). *Mountains and Climate Change. From Understanding to Action*. Bern: Geographica Bernensia.

Laffan, S. W., and Crisp, M. D. (2003). Assessing endemism at multiple spatial scales, with an example from the Australian vascular flora. *J. Biogeogr.* 30, 511–520. doi: 10.1046/j.1365-2699.2003.00875.x

Laffan, S. W., Lubarsky, E., and Rosauer, D. F. (2010). Biodiverse, a tool for the spatial analysis of biological and related diversity. *Ecography* 33, 643–647. doi: 10.1111/j.1600-0587.2010.06237.x

Lamoreux, J. F., Morrison, J. C., Ricketts, T. H., Olson, D. M., Dinerstein, E., McKnight, M. W., et al. (2006). Global tests of biodiversity concordance and the importance of endemism. *Nature* 440, 212–214. doi: 10.1038/nature04291

Lei, F. M., Qu, Y. H., Lu, J. L., Liu, Y., and Yin, Z. H. (2003). Conservation on diversity and distribution patterns of endemic birds in China. *Biodivers. Conserv.* 12, 239–254.

Li, Y., Chen, J., Jiang, L. Y., and Qiao, G. X. (2017). Islands conserve high species richness and areas of endemism of Hormaphidinae aphids. *Curr. ZooL.* 63, 623–632. doi: 10.1093/cz/zox004

Liang, A. P. (1998). Cladistic biogeography of Cercopoidea and Fulgoroidea (Insecta: Homoptera) in China and adjacent regions. *Acta Zootaxon. Sin.* 23, 132–167.

Linder, H. P. (2001a). On areas of endemism, with an example from the African Restionaceae. *Syst. Biol.* 50, 892–912. doi: 10.1080/106351501753462867

Linder, H. P. (2001b). Plant diversity and endemism in sub−Saharan tropical Africa. *J. Biogeogr.* 28, 169–182. doi: 10.1046/j.1365-2699.2001.00527.x

López-Pujol, J., Zhang, F. M., Sun, H. Q., Ying, T. S., and Ge, S. (2011). Centres of plant endemism in China: places for survival or for speciation? *J. Biogeogr.* 38, 1267–1280. doi: 10.1111/j.1365-2699.2011.02504.x

Moline, P. M., and Linder, H. P. (2006). Original article: input data, analytical methods and biogeography of Elegia (Restionaceae). *J. Biogeogr.* 33, 47–62. doi: 10.1111/j.1365-2699.2005.01369.x

Morrone, J. J. (1994). Society of systematic biologists on the identification of areas of endemism. *Syst. Biol.* 43, 438–441. doi: 10.2307/2413679

Morrone, J. J. (2014a). Biogeographical regionalisation of the Neotropical Region. *Zootaxa* 3782, 001–110.

Morrone, J. J. (2014b). Parsimony analysis of endemicity (PAE) revisited. *J. Biogeogr.* 41, 842–854. doi: 10.1111/jbi.12251

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., and Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature* 403, 853–858. doi: 10.1038/35002501

Nelson, G., and Platnick, N. (1981). *Systematics and Biogeography*. New York: Columbia University Press.

Noroozi, J., Talebi, A., Doostmohammadi, M., Rumpf, S. B., Linder, H. P., and Schneeweiss, G. M. (2018). Hotspots within a global biodiversity hotspot-areas of endemism are associated with high mountain ranges. *Sci. Rep.* 8:10345.

Noroozi, J., Zare, G., Sherafati, M., Mahmoodi, M., Moser, D., Asgarpour, Z., et al. (2019). Patterns of endemism in Turkey, the meeting point of three global biodiversity hotspots, based on three diverse families of vascular plants. *Front. Ecol. Evol.* 7:159. doi: 10.3389/fevo.2019.00159

Oliveira, U., Brescovit, A. D., and Santos, A. J. (2015). Delimiting Areas of Endemism through Kernel Interpolation. *PLoS One* 10:e0116673. doi: 10.1371/journal.pone.0116673

Orme, C. D., Davies, R. G., Burgess, M., Eigenbrod, F., Pickup, N., Olson, V. A., et al. (2005). Global hotspots of species richness are not congruent with endemism or threat. *Nature* 436, 1016–1019. doi: 10.1038/nature03850

Platnick, N. I. (1991). On areas of endemism. *Aust. Syst. Bot.* 4, 11–12.

Posadas, P., Crisci, J., and Katinas, L. (2006). Historical biogeography: a review of its basic concepts and critical issues. *J. Arid. Environ.* 66, 389–403. doi: 10.1016/j.jaridenv.2006.01.004

Rahbek, C., Borregaard, M. K., Antonelli, A., Colwell, R. K., Holt, B. G., Noguesbravo, D., et al. (2019). Building mountain biodiversity: Geological and evolutionary processes. *Science* 365, 1114–1119. doi: 10.1126/science.aax0151

Rangel, T. F., Diniz-Filho, J. A. F., and Bini, L. M. (2010). SAM: A comprehensive application for spatial analysis in macroecology. *Ecography* 33, 46–50. doi: 10.1111/j.1600-0587.2009.06299.x

Rundle, H. D., and Nosil, P. (2005). Ecological speciation. *Ecol. Lett.* 8, 336–352. doi: 10.1111/j.1461-0248.2004.00715.x

Sandel, B., Arge, L., Dalsgaard, B., Davies, R. G., Gaston, K. J., Sutherland, W. J., et al. (2011). The influence of Late Quaternary climate change velocity on species endemism. *Science* 334, 660–664. doi: 10.1126/science.1210173

Szumik, C. A., Aagesen, L., Casagranda, D., Arzamendia, V., Baldo, D., Claps, L. E., et al. (2012). Detecting areas of endemism with a taxonomically diverse data set: plants, mammals, reptiles, amphibians, birds, and insects from Argentina. *Cladistics* 28, 317–329. doi: 10.1111/j.1096-0031.2011.00385.x

Szumik, C. A., Cuezzo, F., Goloboff, P. A., and Chalup, A. E. (2002). An optimality criterion to determine areas of endemism. *Syst. Biol.* 51, 806–816. doi: 10.1080/10635150290102483

Tang, Z. Y., Wang, Z. H., Zheng, C. Y., and Fang, J. Y. (2006). Biodiversity in China's mountains. *Front. Ecol. Environ.* 4:347–352. doi: 10.1890/1540-92952006004[0347:BICM]2.0.CO;2

Tews, J., Brose, U., Grimm, V., Tielborger, K., Wichmann, M. C., Schwager, M., et al. (2004). Animal species diversity driven by habitat heterogeneity/ diversity: the importance of keystone structures. *J. Biogeogr.* 31, 79–92. doi: 10.1046/j.0305-0270.2003.00994.x

Tribsch, A. (2004). Areas of endemism of vascular plants in the Eastern Alps in relation to Pleistocene glaciation. *J. Biogeogr.* 31, 747–760. doi: 10.1111/j.1365-2699.2004.01065.x

Wang, F., Jiang, C. Z., Liu, J. Z., and Wei, J. F. (2017). Areas of endemism for scale insects in China. *J. Asia-Pac. Entomol.* 20, 1170–1174. doi: 10.1016/j.aspen.2017.08.024

Wang, H. Z. (1985). *Atlas of the Palaeogeography of China*. Beijing: Cartographic Publishing House.

Wilson, S. W. (2005). Keys to the families of Fulgoromorpha with emphasis on planthoppers of potential economic importance in the southeastern United States (Hemiptera: Auchenorrhyncha). *Fla. Entomol.* 88, 464–481. doi: 10.1653/0015-4040(2005)88[464:kttfof]2.0.co;2

Wilson, S. W., Mitter, C., Denno, R. F., and Wilson, M. R. (1994). "Planthoppers: Their Ecology and Management," in *Evolutionary patterns of host plant use by delphacid planthoppers and their relatives*, eds R. F. Denno and T. J. Perfect (New York: Chapman and Hall), 7–45.

Xu, Y., Shen, Z. H., Ying, L. X., Zang, R. G., and Jiang, Y. X. (2019). Effects of current climate, paleo-climate, and habitat heterogeneity in determining biogeographical patterns of evergreen broad-leaved woody plants in China. *J. Geogr. Sci.* 29, 1142–1158. doi: 10.1007/s11442-019-1650-x

Yuan, S., Huang, M., Wang, X. S., Ji, L. Q., and Zhang, Y. L. (2014). Centers of endemism and diversity patterns for Typhlocybine leafhoppers (Hemiptera: Cicadellidae: Typhlocybinae) in China. *Insect Sci.* 21, 523–526. doi: 10.1111/1744-7917.12040

Zhao, Z. X., Jin, B. C., Zhou, Z. X., Yan, L., Long, J. K., and Chen, X. S. (2020a). Determinants of Delphacidae richness and endemism in China. *Ecol. Entomol.* 45, 1396–1407. doi: 10.1111/een.12924

Zhao, Z. X., Yang, L., Long, J. K., Chang, Z. M., Zhou, Z. X., Zhi, Y., et al. (2020b). Testing seven hypotheses to determine what explains the current planthopper (Fulgoridae) geographical and species richness patterns in China. *Insects* 11:892. doi: 10.3390/insects11120892

Zou, D. T., Wang, Q. G., Luo, A., and Wang, Z. H. (2019). Species richness patterns and resource plant conservation assessments of Rosaceae in China. *Chinese J. Plant. Ecol.* 43, 1–15. doi: 10.17521/cjpe.2018.0091

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 01 September 2021
doi: 10.3389/fevo.2021.730549

Metabarcoding Analysis of Pollen Species Foraged by *Osmia excavata* Alfken (Hymenoptera: Megachilidae) in China

Huanhuan Lu[1†], Feiyue Dou[1†], Youjin Hao[6], Yue Li[1], Ke Zhang[1], Huan Zhang[1], Zeyang Zhou[1], Chaodong Zhu[2,3,4,5], Dunyuan Huang[1] and Arong Luo[2,3*]*

[1] Chongqing Key Laboratory of Vector Insects, College of Life Sciences, Chongqing Normal University, Chongqing, China, [2] Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China, [3] International College, University of Chinese Academy of Sciences, Beijing, China, [4] College of Biological Sciences, University of Chinese Academy of Sciences, Beijing, China, [5] State Key Laboratory of Integrated Pest Management, Institute of Zoology, Chinese Academy of Sciences, Beijing, China, [6] College of Life Sciences, Chongqing Normal University, Chongqing, China

OPEN ACCESS

Edited by:
Shu-Jun Wei, Beijing Academy of Agricultural and Forestry Sciences, China

Reviewed by:
Xiangqun Yuan, Northwest A&F University, China
Huanli Xu, China Agricultural University, China
Xingyuan Men, Institute of Plant Protection, Shandong Academy of Agricultural Sciences, China

***Correspondence:**
Dunyuan Huang
20170054@cqnu.edu.cn
Arong Luo
luoar@ioz.ac.cn

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *25 June 2021*
Accepted: *30 July 2021*
Published: *01 September 2021*

Citation:
Lu H, Dou F, Hao Y, Li Y, Zhang K, Zhang H, Zhou Z, Zhu C, Huang D and Luo A (2021) Metabarcoding Analysis of Pollen Species Foraged by Osmia excavata Alfken (Hymenoptera: Megachilidae) in China. Front. Ecol. Evol. 9:730549. doi: 10.3389/fevo.2021.730549

To meet the pollination need of economic crops, *Osmia excavata* has been successfully used to improve the pollination efficiency of Rosaceae and Brassicaceae plants. As a widely used pollinator of economic crops, a systematic study of flower-visiting species and diversities of *O. excavata* stocked in China was not found. To investigate the foraging pollen species and diversities of *O. excavata*, beebread from 20 experimental plots in China was collected by the trap-nesting method and analyzed by DNA metabarcoding technology. A total of 26 pollen plants in 14 genera and nine families were identified. A further analysis showed that the richness and abundance of the wild flowering plants in orchards and farmlands were lower than those in the nearby semi-natural habitats. The favorite pollen comes from economic crops apple and rape and wild flowering plants *Juncus interior*, *Rosa gymnocarpa*, and *Rosa laevigata*. Through a diversity index analysis, it was found that the Anhui region has the highest pollen plant diversity, while the Liaoning region has the lowest. Our results can provide a basis for flower-visiting species and diversities of *O. excavata*.

Keywords: *Osmia excavata*, metabarcoding method, pollen plants, diversity, species conservation

INTRODUCTION

Pollinators are of great importance in ecosystems and contribute about $136 billion annually to agricultural systems in China (Ouyang et al., 2019). Due to the intensification and expansion of agriculture, the diversity of pollinators has been declining rapidly for decades, which has led to the loss, fragmentation, and degradation of pollinator habitats and a severe decline in global pollination (Foley et al., 2011; Féon et al., 2013; Belsky and Joshi, 2019). In the past 50 years, the dependence of global agriculture on pollinators has increased by more than 300%, while the abundance of pollinators worldwide has continued to decline (Potts et al., 2016). Previous studies have shown that planting of pollen plants in pollinator habitats can significantly increase the number and species of pollinators (Jönsson et al., 2015; M'Gonigle et al., 2016; Warzecha et al., 2018; Williams and Lonsdorf, 2018; Nichols et al., 2019). Therefore, it is necessary to clarify how to choose the best plants for different pollinators.

The internal transcribed spacer II (*ITS2*) barcode based on metabarcoding method can identify pollen plant species from mixed pollen and shows great potential in the analysis of beebread ingredients (Pornon et al., 2016; De-Vere et al., 2017; Lucas et al., 2018). Compared with traditional

palynology, the molecular method not only has a higher sensitivity for the identification of mixed pollen but also effectively reduces time and cost (Lang et al., 2018; Tremblay et al., 2019). The *ITS2* region has proved to be a core barcode for pollen identification (Richardson et al., 2015; Gous et al., 2019; Tremblay et al., 2019). The barcode combination of *rbcL*, *trnH-psbA*, and *ITS2* successfully identified the main pollen plants, *Vitex negundo* and *Vitex quinata*, in the beebread of *Megachile strupigera* by He et al. (2020). Therefore, DNA metabarcode not only helps to identify species and determine the relative abundance of pollen plants in beebread but is also used to construct pollination networks, monitor plant pathogens, assess species diversity, and predict biodiversity patterns (Huang et al., 2020).

To meet the pollination demand of economic crops, pollinators with commercial potential have been found and developed (Ryder et al., 2020). Bumblebees have been successfully used as pollinators for eggplant, tomato, peach, and pear (Wu et al., 2006). *Megachile rotundata* was used for the pollination of alfalfa (Xu et al., 2009). In orchards, the bee of genus *Osmia* was considered as efficient pollinators of some Rosaceae plants, such as apricot, plum, cherry, peach, pear, and apple (Parker, 1981; Torchio et al., 1987; Wei et al., 1991; Xu et al., 1994). Importantly, *Osmia* bee can be artificially stocked due to its advantages such as easy management, low cost, and high pollination efficiency (Wei and Zhao, 1995; Wei et al., 2000b; Dai, 2004). In north and northwest of China, there are mainly five species of *Osmia* (*O. excavata*, *O. cornifrons*, *O. taurus*, *O. jacoti*, and *O. pedicornis*) that play important roles in enhancing pollination, increasing fruit diameter and the number of seeds per fruit, and decreasing the percentage of asymmetrical fruit (Wei et al., 2002). For *O. excavata*, it has many advantages, such as low-temperature tolerance, long daily activity time, high flower-visiting efficiency, centralized pollination range, and easy management (Wei and Yuan, 1997). Therefore, it was considered as an excellent and large-scale stocking pollinator. The pollination of *O. excavata* can create more than $27.8, $9.4, and $3.0 billion for cherries, apples, and rapes planting, respectively, each year (Liu et al., 2018, 2019a,b). Pollination can significantly improve the fruit set rate and fruit quality, and it has broad application prospects in the future of fruit planting industry (He and Zhou, 2009).

However, the short flowering period of fruit trees limits the nutritional requirements of *O. excavata* during the foraging period (Yan et al., 2018). Because of this limitation, survival and large-scale stocking of *O. excavata* face great challenges. A previous study showed that the diversity of pollen plants is the main factor affecting the diversity of wild bee species (Kovács-Hostyánszki et al., 2017). Therefore, the effective solution is to grow other pollen plants in the orchard. This will not only improve the biodiversity of plantations but also provide adequate food sources for bees (Boyle et al., 2020). So, it is necessary to understand the foraging preferences of bees. In this study, beebread of *O. excavata* was collected by trap-nesting from 20 experimental plots in China. The species and abundance of plants were determined by *ITS2* barcode based on metabarcoding method, and the dominant plants that meet the nutrition needs of bees were further analyzed.

MATERIALS AND METHODS

Sample Collection and DNA Extraction

Osmia excavata can build a nest in trap nests made by reed tubes and make beebread to breed its offspring (Wang et al., 2001; Huang et al., 2013). According to the main activity period (March–April) of *O. excavata*, trap nests were placed in similar habitats (e.g., flower composition, competitors, and predators, etc.) in each experimental plot. Trap nests of *O. excavata* were collected from 20 regions of China (March–May 2020), covering three types of land: orchards (monoculture or mixed-species plantation), farmlands, and semi-natural habitats (SNHs; agroforestry ecotone). Detailed geographic information of each sampling plot and surrounding main plant types are listed in **Table 1**. In the laboratory, mixed pollens within five groups of beebread from the trap nests of each site were stored at –80°C for later use. Mixed pollens were extracted using the E.Z.N.A.® Soil DNA Kit according to the manufacturer's protocol (Omega Bio-Tek, Norcross, GA, United States). The concentration and purity of purified DNA were determined by a NanoDrop spectrophotometer (Thermo Fisher Scientific, Inc., United States).

PCR Amplification and Sequencing

The *ITS2* region was amplified using the forward primer ITSS2F: 5′ATGCGATACTTGGTGTGAAT3′ and the reverse primer ITSS4R: 5′TCCTCCGCTTATTGATATGC 3′. PCR reactions were performed in a 20-μl system, containing 4 μl FastPfu Buffer (5×), 2 μl dNTPs (2.5 mM/each), 0.8 μl each primer (5 μM), 0.4 μl FastPfu Polymerase (5 U/μl), 2 μl template DNA (5 ng/μl), and 10 μl ddH_2O. All reactions were performed in an ABI GeneAmp® 9700 (Applied Biosystems, Carlsbad, CA, United States) with the following parameters: an initial denaturation at 95°C for 5 min, followed by 29 cycles of 95°C for 30 s, 55°C for 30 s, 72°C for 45 s, and a final extension at 72°C for 10 min. PCR products were purified using the AxyPrep DNA Gel Extraction Kit (Axygen Biosciences, Union City, CA, United States). The DNA concentrations were measured with a Qubit® 3.0 fluorometer (Life Invitrogen, Waltham, MA, United States) and then paired-end sequenced (2 × 250 bp) on an Illumina MiSeq platform (Illumina, San Diego, CA, United States) at the Biozeron Biological Technology Co. Ltd. (Shanghai, China).

Data Analysis

Raw sequencing data were filtered to delete the tail-base of reads ($Q < 20$) using Trimmomatic v0.38 (Bolger et al., 2014). The chimeric sequences were removed and effective sequences clustered into operational taxonomic units (OTUs; ≥97% similarity) by Usearch v10 (Edgar, 2013). Based on the Unite database v8.2,[1] the OTU representative sequences were analyzed taxonomically to obtain the community composition of each sample using the RDP classifier Bayes algorithm with default parameters (Wang et al., 2007).

[1] http://unite.ut.ee/index.php

TABLE 1 | Geographic information of pollen plants in 20 experimental plots.

Sample	Location	Regions	Coordinates (Lat./Lon.)	Altitude (m)	Main types of plants	Main type of land use
H1	Chengguan, Huaiyuan	Anhui	32.947/117.221	64	Pomegranate	Farmlands
H2	Tangji, Huaiyuan	Anhui	32.866/117.221	24	Pomegranate	Farmlands
H3	Liugusi, Cangzhou	Hebei	38.606/116.273	7	Apple and pear	Mixed orchards
H4	Zishi, Jingzhou	Hebei	30.183/112.423	35	Rape	Farmlands
H5	Lianjiang, Ganzhou	Jiangxi	26.345/115.327	174	Mixed field[1]	SNHs
H6	Wuyun, Ganzhou	Jiangxi	25.993/114.828	142	Mixed field[1]	SNHs
H7	Liantang, Nanchang	Jiangxi	28.504/115.958	25	Rape	Farmlands
H8	Sanrunbao, Dalian	Liaoning	38.940/121.170	30	Multispecies of cherry	Mixed orchards
H9	Manduhu, Shenyang	Liaoning	41.567/122.629	15	Multispecies of apple	Mixed orchards
H10	Niuxintuo, Shenyang	Liaoning	41.571/122.524	13	Apple	Monoculture orchard
H11	Tangwang, Jinan	Shandong	36.819/117.276	26	Apple and pear	Mixed orchards
H12	Cuiping, Qixia	Shandong	37.279/120.844	174	Apple	Monoculture orchard
H13	Guandao, Qixia	Shandong	37.132/120.670	94	Apple	Monoculture orchard
H14	Taocun, Qixia	Shandong	37.179/121.193	90	Apple	Monoculture orchard
H15	Pingyao, Jinzhong	Shanxi	37.282/112.306	780	Apple and pear	Mixed orchards
H16	Nanan, Wenshui	Shanxi	37.492/112.229	754	Apple and pear	Mixed orchards
H17	Qunbake, Luntai	Xinjiang	41.968/84.155	1,498	Multispecies of pear	Mixed orchards
H18	Yiganqi, Aksu	Xinjiang	41.144/80.263	1,097	Multispecies of apple	Mixed orchards
H19	Towankebashi, Aksu	Xinjiang	41.034/80.353	1,081	Multispecies of apple	Mixed orchards
H20	Mongolian	Xinjiang	41.727/86.023	921	Pear	Monoculture orchard

[1]*Mixed field, wild plants, and small-scale economic plants (pear tree and rape; planting area: 2 km^2) were planted. SNHs, semi-natural habitats.*

Alpha diversity indexes (Shannon, Simpson, and Chao1) were calculated using Mothur v1.30 (Schloss et al., 2009). The Fisher's least significant difference (LSD) in SPSS v25.0 (SPSS, Inc., Chicago, IL, United States) was used to analyze the differences of diversities of beebread between different land types and experimental plots. Species accumulation curves (SACs), the abundance, and composition of beebread were prepared and visualized by the "ggplot2" package (Hadley, 2016) and the "pheatmap" package (Raivo, 2019).

RESULTS

Sequencing and Assembly

A total of 4,563,930 clean *ITS2* sequences from all experimental plots were obtained with an average length of 335 ± 9 bp and clustered into 64 OTUs (**Supplementary Table 1**). A statistical analysis showed that the SACs tend to be flat, indicating that the sample sizes reach the conditions that species are fully discovered. The species richness in the environment should not increase significantly with the increase of sample sizes (**Supplementary Figure 1**).

Species and Abundance Analysis of Pollen Plants

Based on OTU representative sequences, a total of nine families (14 genera and 26 species) were identified from all samples combined (**Supplementary Table 1**). The clustering results of plant species showed that the rape and apple pollens were rich in four samples (H1, H2, H4, and H7) and 11 samples (H3, H9–H16, H18, and H19), respectively. Wild plant pollens were dominated in samples H2 (*Juncus interior* and *Papaver rhoeas*), H5 (*J. interior*, *Rosa gymnocarpa*, *Rosa laevigata*, and *Rosa primula*), H6 (*R. gymnocarpa*, *R. laevigata*, and *R. primula*), H7 (*J. interior*), and H16 (*Prinsepia sinensis*; **Figure 1** and **Supplementary Table 2**). A further analysis revealed that plant species in beebread was dominated by economic crops, including apple (*Malus domestica*, *Malus pumila*, and *Malus* sp.), pear (*Pyrus betulifolia*, *Pyrus communis*, *Pyrus cordata*, and *Pyrus spinosa*), plum (*Prunus brachypoda*), cherry (*Cerasus avium*), and rape (*Brassica juncea* and *Brassica* sp.). In addition, nine types of wild plants were detected in beebread. Among them, the abundance of *J. interior*, *R. gymnocarpa*, and *R. laevigata* was higher than others (**Figure 1** and **Supplementary Table 2**).

Impacts of Different Types of Land on the Composition of Beebread

The composition and abundance of pollen in beebread from different types of land use were compared (**Table 1**). The results showed that economic crop pollens mainly came from the single- or mixed-fruit plantations and intensive farmland, and wild plant pollens mainly came from the agroforestry ecotone (**Figure 2A**). Pollen plants in beebread from semi-natural habitat (SNH; agroforestry ecotone) samples have the highest species abundance (4.7 ± 0.95, n = 10), then followed by that from mixed plantations (3.33 ± 0.76, n = 45), farmlands (3.05 ± 0.94, n = 20), and single-fruit plantations (2.08 ± 0.56, n = 25; **Figure 2B**). A further analysis found that the species abundance of pollen plants in beebread from farmlands and mixed-fruit plantations did not differ significantly (p = 0.36), but there were significant differences between other experimental plots ($p < 0.05$; **Supplementary Table 3**).

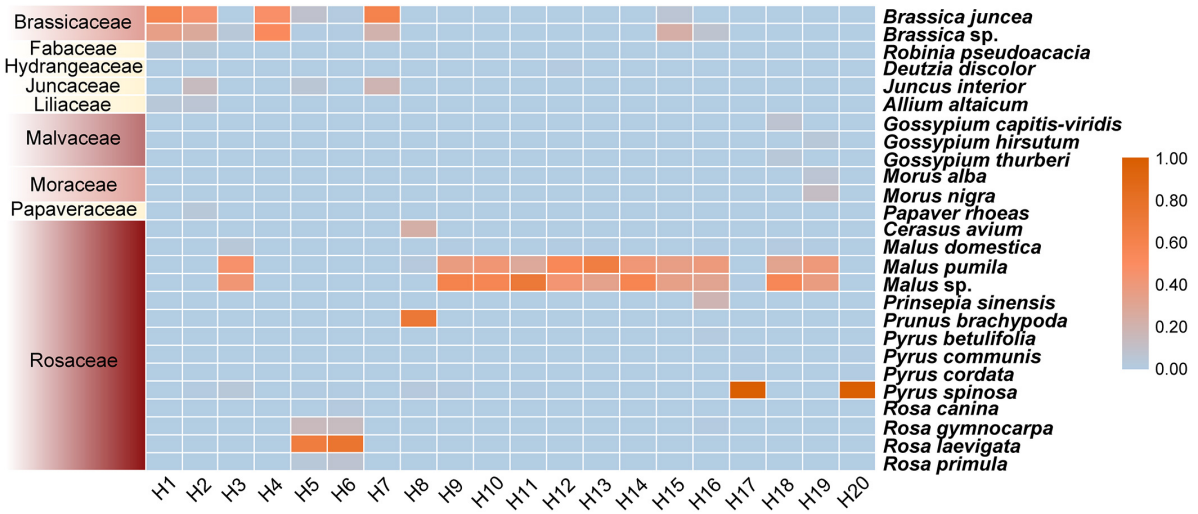


FIGURE 1 | Composition and relative abundance of pollen plants. The left and right of the *Y*-axis and the *X*-axis represent the family and species of pollen plants and sampling plot, respectively. The legend in the upper right represents the proportional abundance of pollen plants.

Diversity of Pollen Plants in Beebread

To better understand the richness and diversity of pollen plants in beebread, 20 experimental plots were divided into seven regions according to the province where *O. excavata* is attracted (**Table 1**). The results showed that the Chao1 index of Anhui (9.40 ± 2.72, n = 10) is the highest, then followed by that of Jiangxi (7.15 ± 2.11, n = 20), Hebei (5.60 ± 2.97, n = 5), Shanxi (5.40 ± 1.07, n = 10), Xinjiang (4.80 ± 0.89, n = 20), Shandong (2.55 ± 1.10, n = 20), and Liaoning (2.33 ± 0.98, n = 15; **Table 2** and **Supplementary Table 4**). The Shannon index of Anhui (1.82 ± 0.24, n = 10), Shanxi (1.37 ± 0.14, n = 10), Jiangxi (1.29 ± 0.38, n = 20), and Hebei (1.06 ± 0.31, n = 5) is significantly higher than that of Xinjiang (0.74 ± 0.45, n = 20), Shandong (0.67 ± 0.10, n = 20), and Liaoning (0.62 ± 0.18, n = 15; all $p < 0.05$), which indicated that the diversity of pollen plants in Anhui, Shanxi, Jiangxi, and Hebei is high (**Table 2** and **Supplementary Table 4**). A large Chao1 and Shannon indexes or a small Simpson index can explain the high species richness and diversity in beebread. Therefore, it can be concluded that the abundance of pollen plants in the beebread from Anhui was significantly higher than that in other regions. In general, the richness and diversity of pollen plants in different regions can be divided into three levels from high to low: Anhui → Hebei, Shanxi, and Jiangxi → Xinjiang, Shandong, and Liaoning (**Table 2**).

DISCUSSION

Wild bees are important pollinators for economic crops and wild plants. They rely on different landscapes to obtain flowering resources and nesting habitats (Kratschmer et al., 2019). In this study, 20 experimental plots in China were set up to collect the beebread of *O. excavata*. A total of 26 pollen plant species were identified, belonging to 14 genera of nine families and including 17 economic crops and nine wild plant species (*Robinia pseudoacacia*, *Deutzia discolor*, *J. interior*, *P. rhoeas*, *P. sinensis*, *Rosa canina*, *R. gymnocarpa*, *R. laevigata*, and *R. primula*). Based on the composition and relative abundance of pollens (**Figures 1**, **2**), *O. excavata* is an effective pollinator for economic crops (*B. juncea*, *Brassica* sp., *M. pumila*, *Malus* sp., *P. brachypoda*, and *P. spinosa*).

Previous studies indicated that the larvae of *O. excavata* released at the early flowering stages of economic crops will grow into effective pollinators, especially Rosaceae and Brassicaceae plants (Yang et al., 1997; Wei et al., 2000a; Hawkins et al., 2015; Gresty et al., 2018). However, the flowering period (7–15 days; **Supplementary Table 5**) (Lech et al., 2008; Liu et al., 2008, 2010; Wang et al., 2012; Lu et al., 2013; Lin et al., 2019; Zhang et al., 2019, 2020; Li et al., 2020) of most economic crops is short and cannot cover the entire adult stage of *O. excavata* (40–55 days; Wei et al., 2002; Men et al., 2018). Through the observation of professional beekeepers and our field work on the stocking of *O. excavata*, when *O. excavata* had no diverse food sources except economic crops, this gap seriously affects the bees to reserve beebread for their offspring, which causes a low recovery rate after artificial stocking. The number of reduced populations needs to be supplemented every year, excluding the situation of unnatural death, human activities, and population migration.

Although when the beebread from H5/H6 experimental plots were collected, the economic crops (pear and rape trees) in the plantation were in the late flowering stage, which led to the lack of pollen resources, *O. excavata* could still obtain pollens from wild plants (*J. interior*, *R. gymnocarpa*, *R. laevigata*, and *R. primula*; **Figure 1**). This means that in order to reduce the impact of the lack of pollen resources of economic crops, *O. excavata* can use the pollen of wild plants as their main source of nutrition for meeting survival needs. Plantations (orchards and farmlands) were dominated by a single type of economic crops and lack of flowering resources of wild plants (**Figure 2A**). A further analysis revealed that more wild plants in beebread were from SHNs (**Figure 2A**). Papanikolaou et al. (2017) found that increasing

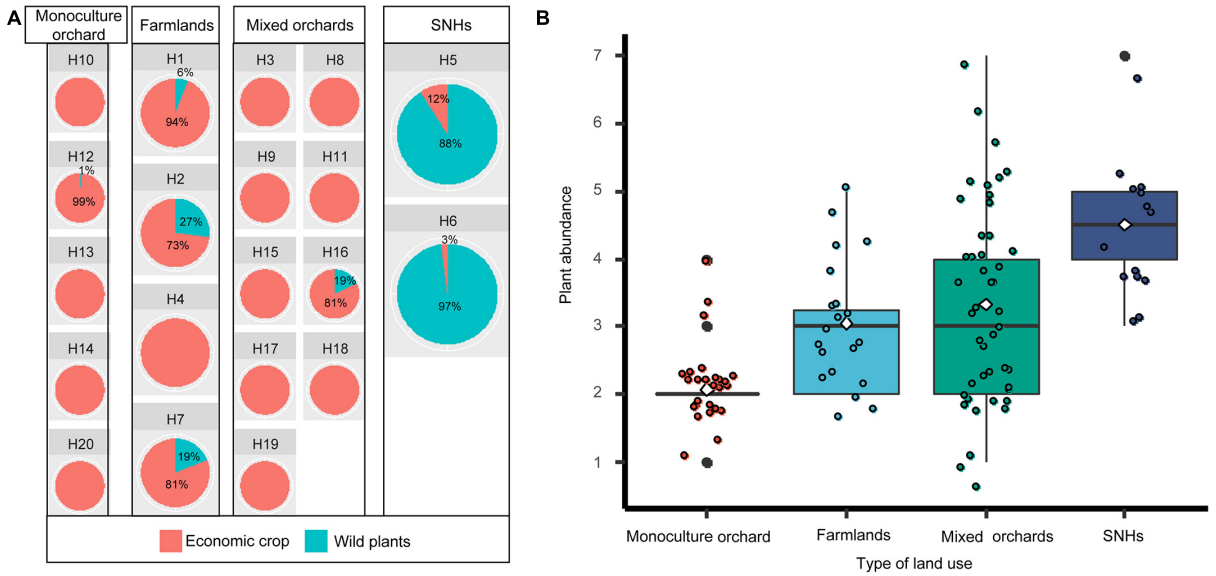


FIGURE 2 | The composition **(A)** and abundance **(B)** of beebread in different types of land use. **(A)** The proportion of economic crops and wild plants in each experimental plot. **(B)** Plant abundance under different land-use types. The *X*-axis and *Y*-axis represent land-use types and plant abundance, respectively.

the richness of plants in the habitat would significantly affect the abundance and diversity of bees. When the pollen resources of economic crops cannot meet the needs of the entire adult stage of *O. excavata*, wild flowering plants can be planted in plantations or introduced into SNHs. The transplanting of wild flowering plants into the plantation helps to attract *O. excavata* and increase food sources. Ideal type of wild flowering plants can be selected from nine wild plants identified in beebread, especially *J. interior*, *R. laevigata*, and *R. gymnocarpa*.

Piko et al. (2021) concluded that the mixed flower field model can effectively provide attractive food resources, thereby greatly increasing the abundance and diversity of bumblebee. Bihaly et al. (2020) also emphasized that planting a variety of flowering herbaceous plants on a small scale could significantly increase the number of bee nests in the orchard. In response to concerns about the food source of the *O. excavata*, combing the previous results, a conjecture is proposed: a mixed flower field model suitable for *O. excavata* (MFF-OE model). We can plant a variety of favorite wild flowering plants around the plantation to meet the needs of pollen resources during the entire adult stage of *O. excavata*. However, the model needs to be further explored in terms of variety allocation, optimum planting area selection, and planting scale, etc. The comprehensive analysis of Chao1 and Shannon indexes reflected lower species richness and diversity in Liaoning. The Liaoning region can be used as a pilot for the exploration and implementation of the MFF-OE model. Taken together, we look forward that the implementation of the MFF-OE model can effectively solve the shortage of pollen resources of economic crops.

TABLE 2 | Alpha diversity indexes of pollen plants.

Region	Chao1 index[1]	Shannon index[1]	Simpson index[1]
Anhui	9.40 ± 2.72	1.82 ± 0.24	0.16 ± 0.04
Jiangxi	7.15 ± 2.11	1.29 ± 0.38	0.38 ± 0.17
Hebei	5.60 ± 2.97	1.06 ± 0.31	0.20 ± 0.23
Shanxi	5.40 ± 1.07	1.37 ± 0.14	0.28 ± 0.04
Xinjiang	4.80 ± 0.89	0.74 ± 0.45	0.62 ± 0.26
Shandong	2.55 ± 1.10	0.67 ± 0.10	0.54 ± 0.07
Liaoning	2.33 ± 0.98	0.62 ± 0.18	0.55 ± 0.14

[1]*Index data format refer to "mean ± SD."*

CONCLUSION

Osmia excavata has been successfully used for pollination in commercial orchards because it can significantly improve fruit setting rate and fruit quality. In this study, beebread of *O. excavata* was collected from various habitats by trap-nesting and analyzed by DNA metabarcoding technology. A total 26 pollen plants in 14 genera and nine families were identified, including 17 economic crops and nine wild plant species. The abundances of wild flowering plants in the plantation were lower than those in SNHs. According to the composition and abundance of pollen plants, *O. excavata* is regarded as an effective pollinator for economic crops in Rosaceae and Brassicaceae from the molecular level.

DATA AVAILABILITY STATEMENT

The sequence data that support the findings of this study are openly available in GenBank of NCBI at https://www.ncbi.nlm.nih.gov/. The associated BioProject, SRA, and

BioSample numbers are PRJNA742556, SRR15011441, and SAMN19967841, respectively.

ETHICS STATEMENT

All experiments procedures for this study complied with the current animal ethics guidelines and did not involve any protected animals.

AUTHOR CONTRIBUTIONS

AL and DH were responsible for the conceptualization and supervision of the study. HL, FD, and YH were responsible for the methodology and writing – original draft preparation. HL, FD, YL, KZ, and HZ were responsible for the software. ZZ, CZ, AL, and DH performed the validation. HL and FD conducted the formal analysis and visualization. FD, YL, KZ, HZ, and DH conducted the investigation. FD and DH were responsible for the resources. HL, FD, YL, KZ, HZ, and DH were responsible for data curation. YH, ZZ, CZ, AL, and DH were responsible for writing – review and editing. DH was responsible for project administration. HL, ZZ, AL, and DH were responsible for funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Program of Ministry of Science and Technology of China (2018FY100405); the Research and Innovation Projects of Graduate Students of Chongqing Normal University (YKC20037); the National Natural Science Foundation of China (31970484 and 32070465); and the Key Laboratory of Animal Evolution and Systematics, Chinese Academy of Sciences (E052G21305).

ACKNOWLEDGMENTS

We are grateful to the editor and reviewers for helpful comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2021.730549/full#supplementary-material

REFERENCES

Belsky, J., and Joshi, N. K. (2019). Impact of biotic and abiotic stressors on managed and feral bees. *Insects* 10, 1–42. doi: 10.3390/insects10080233

Bihaly, ÁD., Kovács−Hostyánszki, A., Szalai, M., and Sárospataki, M. (2020). Nesting activity of cavity−nesting bees and wasps is lower in small−scale apple orchards compared to nearby semi−natural habitats. *Agr. Forest Entomol.* 23, 49–58. doi: 10.1111/afe.12403

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Boyle, N. K., Artz, D. R., Lundin, O., Ward, K., Picklum, D., Wardell, G. I., et al. (2020). Wildflower plantings promote blue orchard bee, Osmia lignaria (*Hymenoptera: megachilidae*), reproduction in California almond orchards. *Ecol. Evol.* 10, 3189–3199. doi: 10.1002/ece3.5952

Dai, Y. G. (2004). The advantage of Osmia pollination for fruit trees of *Rosaceae*. *Apicul. Sci. Technol.* 2:48.

De-Vere, N., Jones, L. E., Gilmore, T., Moscrop, J., Lowe, A., Smith, D., et al. (2017). Using DNA metabarcoding to investigate honey bee foraging reveals limited flower use despite high floral availability. *Sci. Rep.* 7:42838. doi: 10.1038/srep42838

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* 10, 996–998. doi: 10.1038/nmeth.2604

Féon, V. L., Burel, F., Chifflet, R., Henry, M., Ricroch, A., Vaissière, B. E., et al. (2013). Solitary bee abundance and species richness in dynamic agricultural landscapes. *Agr. Ecosyst. Environ.* 166, 94–101. doi: 10.1016/j.agee.2011.06.020

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., et al. (2011). Solutions for a cultivated planet. *Nature* 478, 337–342. doi: 10.1038/nature10452

Gous, A., Swanevelder, D. Z. H., Eardley, C. D., and Willows-Munro, S. (2019). Plant-pollinator interactions over time: pollen metabarcoding from bees in a historic collection. *Evol. Appl.* 12, 187–197. doi: 10.1111/eva.12707

Gresty, C. E. A., Clare, E., Devey, D. S., Cowan, R. S., Csiba, L., Malakasi, P., et al. (2018). Flower preferences and pollen transport networks for cavity−nesting solitary bees: implications for the design of agri-environment schemes. *Ecol. Evol.* 8, 7574–7587. doi: 10.1002/ece3.4234

Hadley, W. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Hawkins, J., de Vere, N., Griffith, A., Ford, C. R., Allainguillaume, J., Hegarty, M. J., et al. (2015). Using DNA metabarcoding to identify the floral composition of honey: a new tool for investigating honey bee foraging preferences. *PLoS One* 10:e0134735. doi: 10.1371/journal.pone.0134735

He, B., Gu, Z. Y., Li, H. Y., and Huang, D. Y. (2020). Analysis of species and diversities in the pollen plants of Megachile strupigera (*Hymenoptera: megachilidae*) by DNA barcoding. *Acta Ecol. Sin.* 40, 2122–2129. doi: 10.5846/stxb201809021867

He, W. Z., and Zhou, W. R. (2009). Study on the pollination effect of Osmia jacoti, Apis mellifera and artificial pollination on apple. *Apicul. China* 60, 9–11.

Huang, D. Y., He, B., Yu, J. F., Gu, P., Peng, F., Huang, G. N., et al. (2013). Nesting biology of Euodynerus nipanicus Schulthess (*Hymenoptera: vespidae*). *J. Environ. Entomol.* 35, 783–791.

Huang, D. Y., Huang, S. C., Li, H. Y., Dou, F. Y., Kou, R. M., Li, Y., et al. (2020). Identification methods of nectar or pollen plants for pollinators. *J. Chongqing Normal U. (Nat. Sci.)* 37, 1–8.

Jönsson, A. M., Ekroos, J., Dänhardt, J., Andersson, G. K., Olsson, O., and Smith, H. G. (2015). Sown flower strips in southern Sweden increase abundances of wild bees and hoverflies in the wider landscape. *Biol. Conserv.* 184, 51–58. doi: 10.1016/j.biocon.2014.12.027

Kovács-Hostyánszki, A., Espíndola, A., Vanbergen, A. J., Settele, J., Kremen, C., and Dicks, L. V. (2017). Ecological intensification to mitigate impacts of conventional intensive land use on pollinators and pollination. *Ecol. Lett.* 20, 673–689. doi: 10.1111/ele.12762

Kratschmer, S., Petroviæ, B., Curto, M., Meimberg, H., and Pachinger, B. (2019). Pollen availability for the Horned mason bee (Osmia cornuta) in regions of different land use and landscape structures. *Ecol. Entomol.* 45, 525–537. doi: 10.1111/een.12823

Lang, D., Tang, M., and Zhou, X. (2018). Qualitative and quantitative molecular construction of plant-pollinator network: application and prospective. *Biodivers. Sci.* 26, 445–456. doi: 10.17520/biods.2018058

Lech, W., Malodobry, M., Dziedzic, E., Bieniasz, M., and Doniec, S. (2008). Biology of sweet cherry flowering. *J. Fruit Ornam. Plant Res.* 16, 189–199.

Li, W. J., Huang, W. W., Li, Q., Sun, Q. Q., and Luo, Y. (2020). Research on the forecast method of rape florescence in the Yangtze River basin. *Chin. J. Agr. Res. Reg. Planning* 41, 101–108.

Lin, X., Hou, K. Q., Chen, L. A., Wei, Z. P., and Fang, Q. Y. (2019). Introduction performance of 8 citrus cultivars in eastern fujian province. *Southeast Horticul.* 7, 15–18.

Liu, H. T., Wang, J. S., Liu, W. J., Li, C. Y., and Guo, G. Z. (2008). Study on the characteristics of apricot florescence. *Chin. Agr. Sci. Bull.* 24, 171–175.

Liu, L., Li, L. L., Ouyang, F., Li, C., Yu, Y., Qu, C. H., et al. (2019a). Fruit-setting and yield increase for apple pollination by Osmia excavata Alfken and evaluation of economic value in Shandong province. *Apicul. China* 70, 65–68.

Liu, L., Li, L. L., Ouyang, F., Li, C., Yu, Y., Qu, C. H., et al. (2019b). Fruit-setting, yield increase and economic value evaluation for cherry pollination by Osmia excavata Alfken in Shandong province. *Shandong Agr. Sci.* 51, 125–128.

Liu, L., Li, L. L., Ouyang, F., Li, L. L., Qu, C. H., Li, C., et al. (2018). Economic value of the pollination services provided by Osmia excavata Alfken to rape seed crops. *Chin. J. Appl. Entomol.* 55, 1016–1022.

Liu, Y. N., He, W. L., Li, Y. L., Bo, Q. F., Liang, Y., and Zhang, T. (2010). A study on the risk index design of agricultural insurance on apple florescence freezing injury in Shaanxi fruit zone. *Chin. J. Agrometeorol.* 31, 125–129.

Lu, X. H., Yang, Y. M., Zhu, Z. Y., Wu, Q. S., Liu, S. H., and Sun, G. M. (2013). Study on florescence characteristics of pineapple. *South China Fruits* 42, 66–68.

Lucas, A., Bodger, O., Brosi, B. J., Ford, C. R., Forman, D. W., Greig, C., et al. (2018). Generalisation and specialisation in hoverfly (*Syrphidae*) grassland pollen transport networks revealed by DNA metabarcoding. *J. Animal Ecol.* 87, 1008–1021. doi: 10.1111/1365-2656.12828

Men, X. Y., Li, L. L., Lu, Z. B., Ouyang, F., Liu, L., Xu, H., et al. (2018). Biological characteristics and pollination service of Mason bee. *Chin. J. Appl. Entomol.* 55, 973–983.

M'Gonigle, L. K., Ponisio, L. C., Cutler, K., and Kremen, C. (2016). Habitat restoration promotes pollinator persistence and colonization in intensively managed agriculture. *Ecol. Appl.* 25, 1557–1565. doi: 10.1890/14-1863.1

Nichols, R. N., Goulson, D., and Holland, J. M. (2019). The best wildflowers for wild bees. *J. Insect Conserv.* 23, 819–830. doi: 10.1007/s10841-019-00180-8

Ouyang, F., Wang, L., Yan, Z., Men, X. Y., and Ge, F. (2019). Evaluation of insect pollination and service value in China's agricultural ecosystems. *Acta Ecol. Sin.* 39, 131–145. doi: 10.5846/stxb201809172030

Papanikolaou, A. D., Ingolf, K., Frenzel, M., Kuhlmann, M., Poschlod, P., Potts, S. G., et al. (2017). Wild bee and floral diversity co-vary in response to the direct and indirect impacts of land use. *Ecosphere* 8:e02008. doi: 10.1002/ecs2.2008

Parker, F. D. (1981). A candidate red clover pollinator Osmia coerulescens (L.). *J. Apicult. Res.* 20, 62–65. doi: 10.1080/00218839.1981.11100474

Piko, J., Keller, A., Geppert, C., Batáry, P., and Hass, A. L. (2021). Effects of three flower field types on bumblebees and their pollen diets. *Basic Appl. Ecol.* 52, 95–108. doi: 10.1016/j.baae.2021.02.005

Pornon, A., Escaravage, N., Burrus, M., Holota, H., Khimoun, A., Mariette, J., et al. (2016). Using metabarcoding to reveal and quantify plant-pollinator interactions. *Sci. Rep.* 6:27282. doi: 10.1038/srep27282

Potts, S. G., Imperatriz-Fonseca, V., Ngo, H. T., Báldi, A., Bartuska, A., Baste, I. A., et al. (2016). *The Assessment Report on Pollinators, Pollination and Food Production: Summary for Policymakers*. Bonn: IPBES.

Raivo, K. (2019). *Pheatmap: Pretty Heatmaps. R package version 1.0.12*.

Richardson, R. T., Lin, C. H., Sponsler, D. B., Quijia, J. O., Goodell, K., and Johnson, R. M. (2015). Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. *Appl. Plant Sci.* 3:1400066. doi: 10.3732/apps.1400066

Ryder, J. T., Cherrill, A., Prew, R., Shaw, J., Thorbek, P., and Walters, K. F. A. (2020). Impact of enhanced *Osmia bicornis* (*Hymenoptera: megachilidae*) populations on pollination and fruit quality in commercial sweet cherry (*Prunus avium* L.) orchards. *J. Apicul. Res.* 59, 77–87.

Schloss, P., Westcott, S., Ryabin, T., Hall, J., Hartmann, M., Hollister, E., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09

Torchio, P. F., Asensio, E., and Thorp, R. W. (1987). Introduction of the European bee, Osmia cornuta, into California almond orchards (*Hymenoptera: megachilidae*). *Environ. Entomol.* 58, 42–52. doi: 10.1093/ee/16.3.664

Tremblay, ÉD., Duceppe, M. O., Thurston, G. B., Gagnon, M. C., Côté, M. J., and Bilodeau, G. J. (2019). High−resolution biomonitoring of plant pathogens and plant species using metabarcoding of pollen pellet contents collected from a honey bee hive. *Environ. DNA* 1, 155–175. doi: 10.1002/edn3.17

Wang, F. H., Guo, Z. H., Xu, X. L., and Chen, C. G. (2001). The main influence factors about retrieving of osmia and its improvement mechanism. *Beijing Forest Sci.* 4, 29–30. doi: 10.1201/9780367801618-7

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267.

Wang, X., Wang, J. Z., Zhao, B. Y., Cao, D., and Zhang, S. L. (2012). Research on the changes of temperature and humidity and phenophase of pear between plastic tunnel and open field. *Chin. Agr. Sci. Bull.* 28, 201–206.

Warzecha, D., Diekötter, T., Wolters, V., Jauker, F., Didham, R., and Batáry, P. (2018). Attractiveness of wildflower mixtures for wild bees and hoverflies depends on some key plant species. *Insect Conserv. Diver.* 11, 32–41.

Wei, S. G., Wang, R., Smirle, M. J., and Xu, H. L. (2002). Release of Osmia excavata and Osmia jacoti (*Hymenoptera: megachilidae*) for apple pollination. *Can. Entomol.* 134, 369–380. doi: 10.4039/Ent134369-3

Wei, S. G., Wei, S. L., Wang, R., Zhou, W. R., and Xu, Z. G. (1991). Morphological and biological study of fruit pollinator *Osmia cornifrons*. *Kun chong Zhi shi* 28, 106–108.

Wei, S. G., and Zhao, L. Y. (1995). Attention should be paid to the use of Osmia to pollinate fruit trees. *Northern Fruits* 2, 15–16.

Wei, Y. P., and Yuan, F. (1997). Applying trap-nesting to collection Osmia jacoti for fruit pollinating. *Shaanxi J. Agr. Sci.* 5, 46–47.

Wei, Y. P., Yuan, F., and Zhang, Y. L. (2000a). Flower visiting habits and the essential number of Osmia excavata Alfken for economic apple production. *Acta U. Agric. Boreali Occidentalis* 28, 76–79.

Wei, Y. P., Yuan, F., Zhang, Y. L., and Wang, Y. H. (2000b). The reproductive characteristics of Osmia excavata Alfken. *Acta Agric. Boreali Occidentalis Sin.* 3, 35–38.

Williams, N. M., and Lonsdorf, E. V. (2018). Selecting cost-effective plant mixes to support pollinators. *Biol. Conserv.* 217, 195–202.

Wu, J., An, J. D., Li, J. L., Huang, J. X., Guo, Z. B., and Tong, Y. M. (2006). Research and application of bumblebee in China. *Proc. 7th Symposium Chin. Beekeeping Soc. Pollination Committee* 11, 124–136.

Xu, H. L., Wu, Y. R., Zhou, W. R., Wei, S. G., and Wang, T. (1994). Biological Study on Pollinators of fruit trees——Osima jacoti, Osima excavate. *J. Fruit Sci.* 3, 153–156.

Xu, H. L., Yang, J. W., and Sun, J. R. (2009). Current status on the study of wild bee-pollinators and conservation strategies in China. *Acta Phytophyl. Sin.* 36, 371–376.

Yan, Z., Wang, L. N., Men, X. Y., Xiao, Y. L., Feng, G. E., and Ouyang, F. (2018). Effect of maintaining surrounding habitat near apple orchards on *Osmia excavata* (Hymenoptera: Megachilidae). *Chinese J. Appl. Entomol.* 55, 1007–1015.

Yang, L. L., Xu, H. L., and Wu, Y. R. (1997). Comparative studies of nesting and foraging behavior and pollination ecology of Osmia excavata Alfken and O. jacoti Cockerel in apple orchards (*Hymenoptera: megachilidae*). *Acta Ecol. Sin.* 17, 1–6.

Zhang, L., Yang, M. F., Liu, Y., and Ji, C. R. (2019). Effects of spring temperature change on flowering period of flat peach in shihezi. *Heilongjiang Agric. Sci.* 12, 42–45.

Zhang, Y., Zhang, F. C., Zhong, H. X., Li, J., Pan, M. G., and Wu, X. Y. (2020). Observation on the growth and development characteristics of 7 table grapes. *Xinjiang Agric. Sci.* 57, 2230–2237.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 01 October 2021
doi: 10.3389/fevo.2021.759496

Population Genetic Structure of the Invasive Spotted Alfalfa Aphid *Therioaphis trifolii* (Hemiptera: Aphididae) in China Inferred From Complete Mitochondrial Genomes

Xinzhi Liu[1], Shuhua Wei[2,3], Zhenyong Du[1], Jia He[1,2,3], Xinyue Zhang[1], Hu Li[1], Rong Zhang[2,3] and Wanzhi Cai[1]**

[1] *Department of Entomology, MOA Key Lab of Pest Monitoring and Green Management, College of Plant Protection, China Agricultural University, Beijing, China,* [2] *Institute of Plant Protection, Academy of Ningxia Agriculture and Forestry Science, Yinchuan, China,* [3] *Ningxia Key Lab of Plant Disease and Pest Control, Yinchuan, China*

OPEN ACCESS

Edited by:
Ai-bing Zhang,
Capital Normal University, China

Reviewed by:
Zhijun Zhou,
Hebei University, China
Yuyu Wang,
Agricultural University of Hebei, China

***Correspondence:**
Wanzhi Cai
caiwz@cau.edu.cn
Rong Zhang
yczhrnx@163.com

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *16 August 2021*
Accepted: *13 September 2021*
Published: *01 October 2021*

Citation:
Liu X, Wei S, Du Z, He J, Zhang X, Li H, Zhang R and Cai W (2021) Population Genetic Structure of the Invasive Spotted Alfalfa Aphid Therioaphis trifolii (Hemiptera: Aphididae) in China Inferred From Complete Mitochondrial Genomes. Front. Ecol. Evol. 9:759496. doi: 10.3389/fevo.2021.759496

Biological invasions represent a natural rapid evolutionary process in which invasive species may present a major threat to biodiversity and ecosystem integrity. Analyzing the genetic structure and demographic history of invaded populations is critical for the effective management of invasive species. The spotted alfalfa aphid (SAA) *Therioaphis trifolii* is indigenous in the Mediterranean region of Europe and Africa and has invaded China, causing severe damages to the alfalfa industry. However, little is known about its genetic structure and invasion history. In this study, we obtained 167 complete mitochondrial genome sequences from 23 SAA populations across China based on high-throughput sequencing and performed population genetic and phylogenomic analyses. High haplotype diversity and low nucleotide diversity were found in SAA populations in China with distinct genetic structures, i.e., all populations diverged into three phylogenetic lineages. Demographic history analyses showed a recent expansion of the SAA population, consistent with the recent invasion history. Our study indicated that SAA may have invaded through multiple introduction events during commercial trades of alfalfa, although this needs further validation by nuclear markers.

Keywords: *Therioaphis trifolii*, biological invasion, population genetic structure, phylogenetics, mitochondrial genome

INTRODUCTION

The spotted alfalfa aphid (SAA), *Therioaphis trifolii* (Hemiptera: Aphididae), was first described by Monell (1882). It is an indigenous species in the Mediterranean region of Europe and Africa, far north to Scandinavia, south to Ethiopia, and eastward through the Middle East to Afghanistan and India (Van den Bosch, 1957; Lambers and Van den Bosch, 1964; Carver, 1978). This species is now widely distributed in Asia, Oceania, and North America following worldwide invasions. SAA occurs

several generations per year via both gamogenesis and parthenogenesis, with eggs and adult females representing the main overwintering stages. SAA's principal host is alfalfa (*Medicago sativa*), but it also infests other Fabaceae species of *Trifolium*, *Trigonella*, and *Melilotus* (Peters and Painter, 1957; Bishop and Crockett, 1961; Lambers and Van den Bosch, 1964; Carver, 1978). This aphid damages its host plants by sucking sap of young leaves and twigs, inducing a toxic reaction which causes deformation of leaves, stunting, and plant death (Giles et al., 2002). The invasion of SAA has caused severe economic losses to alfalfa and husbandry industries in southwestern United States, Australia, and China (Howe and Smith, 1957; Van den Bosch et al., 1957; Zhang and Zhong, 1983; Hughes et al., 1987). In China, although some efforts have been made to predict and control this pest, few studies have been performed to assess its genetic structure and invasion history (Hughes et al., 1987).

The invasion of non-native species continues to occur worldwide with the increasing intercontinental and regional trade favoring entry and dissemination. Biological invasions incur not only economic losses but also dramatic ecological and sociological effects (Simberloff et al., 2013; Roques et al., 2016; Nicolas et al., 2018). Introductions of alien species are accompanied by various genetic consequences. For example, populations of invasive species are usually subjected to the founder and bottleneck effects, and the newly established populations may have lower genetic variations (Dlugosch and Parker, 2010). However, multiple introduction events may increase the genetic diversity of the invaded places, especially when several genetically differentiated source populations contribute to the invasion (Facon et al., 2003; Kolbe et al., 2004; Kang et al., 2007). The potentially higher diversity can counterbalance the deleterious effects of genetic drift or inbreeding depression (Ciosi et al., 2008; Lombaert et al., 2010; Estoup et al., 2016). The genetic structure of an invasive population usually reflects its introduction history, and evaluating the genetic structure of native and invasive populations is a crucial step for managing established populations and preventing further invasions (Estoup and Guillemaud, 2010). More specifically, deciphering the genetic structure of non-native populations provides an insight into the dynamics of the invasion, including the number of introduction events, the associated demographic events, or the relationships between different populations (Roderick and Navajas, 2003; Kolbe et al., 2004; Lavergne and Molofsky, 2007). A better understanding of introduction events also plays an important role in the effective management of invasive pests because different control measures need to be applied to populations based on their distinct genetic makeup.

Whole mitochondrial genome (mitogenome) sequences have been proved as powerful molecular markers for analyzing the genetic structure of species (Drag et al., 2015; Du et al., 2019, 2021). In this study, we sequenced mitogenomes of 167 SAA individuals representing 23 geographic populations in China using high-throughput sequencing and analyzed the genetic diversity, phylogenetic structure, and demographic history of SAA. Based on the population genetic analyses, we discussed the possible invasion history of SAA in China and our study would provide a foundation for further analysis of its global introduction events.

MATERIALS AND METHODS

Sampling and Mitogenome Sequencing

When sampling each population, we chose three to five locations with an interval of more than 5 m and collected five individuals at least for each location. Collected samples were placed in absolute alcohol and stored at −80°C until DNA extraction. A total of 167 SAA adults were selected from 23 natural populations across China (**Table 1** and **Figure 1**). Voucher specimens (NO. VAphA001–VAphA167) were deposited in the Entomological Museum of China Agricultural University. We obtained whole genomic DNA from individual specimen using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. The quality and concentration of extracted DNA were measured using Qubit 3.0 (Invitrogen, United States). Each specimen was identified morphologically (Zhang and Zhong, 1983) and checked by BLAST search in the GenBank database using their *COI* gene sequence (with percent identity $\geq$ 99.00%), which was acquired through PCR sequencing under the following conditions: initial denaturation at 95°C for 1 min; 40 cycles at 98°C for 10 s; 45°C for 50 s; 68°C for 1 min; and a final single extension at 72°C for 2 min (with primers, reagents, and reaction system shown in **Supplementary Tables 1, 2**).

A 350 bp insert size library was constructed from each specimen and sequenced in 150 bp pair-end mode to obtain ∼3 Gb of data using an Illumina NovaSeq 6,000 platform (San Diego, CA, United States) at Berry Genomics (Beijing, China). Raw reads were trimmed off adapters using Trimmomatic (Bolger et al., 2014), and low-quality or short reads were discarded with Prinseq (Edwards, 2011). The remaining high-quality reads were mapped onto the published reference mitogenome (GenBank accession: MK766411; Liu et al., 2019) using Geneious 10.1.3,[1] allowing for mismatches of up to 2% with a maximum gap size of 5 bp and a minimum overlap of 30 bp. The consensus sequences were generated as novel individual mitogenomes under the threshold of 75% uniformity at each site. Sequences obtained from mapping were annotated in Geneious according to the reference sequence. Annotation of protein-coding genes (PCGs) and ribosomal RNA genes (rRNAs) was aligned with homologous genes of the reference mitogenome. Transfer RNA genes (tRNAs) were rechecked using the MITOS Web Server (Bernt et al., 2013). Each PCG was independently aligned using TranslatorX online platform with the MAFFT method, with all stop codons removed (Abascal et al., 2010), and the PCG dataset was finally obtained. Besides, we also performed ABGD (Automatic Barcode Gap Definition) analysis with the aligned PCG dataset (*P*min = 0.001, *P*max = 0.02) and confirmed that all sequences belong to the same species *T. trifolii*. Analyses in this study were all based on the PCG dataset.

[1] http://www.geneious.com/

TABLE 1 | Summary of sampling sites and parameters of genetic diversity of *Therioaphis trifolii*.

Sample locality	Code	Date	Altitude	Longitude	Latitude	N	S	Nh	Hd	π
Beibei, Chongqing	CQBB	2019.8	276 m	106°24′14″	29°49′55″	5	0	1	0	0
Qingyang, Gansu	GSQY	2019.8	1240 m	107°50′25″	35°39′34″	5	48	4	0.9	0.00253
Yuzhong, Gansu	GSYZ	2019.1	1653 m	103°40′15″	35°34′20″	5	48	4	0.9	0.00251
Guiyang, Guizhou	GZGY	2019.8	1150 m	106°39′21″	26°30′9″	5	45	2	0.6	0.00246
Cangzhou, Hebei	HBCZ	2018.7	1 m	117°56′28″	38°49′65″	5	9	4	0.9	0.00044
Zhuozhou, Hebei	HBZZ	2018.8	45 m	115°51′01″	39°28′10″	5	49	3	0.7	0.00184
Shiyan, Hubei	HBSY	2019.4	273 m	110°51′03″	32°55′19″	5	9	4	0.9	0.00044
Harbin, Heilongjiang	HHEB	2018.8	87 m	124°26′00″	47°45′01″	5	4	3	0.8	0.0002
Changde, Hunan	HNCD	2019.7	29 m	112°6′37″	29°6′17″	2	0	1	0	0
Zhengzhou, Henan	HNZZ	2019.4	40 m	113°44′55″	34°53′38″	10	58	8	0.956	0.00128
Baichengshi, Jilin	JBCS	2018.7	155 m	122°50′15″	45°36′05″	5	10	4	0.9	0.00036
Chifeng, Inner Mongolia	MGCF	2018.7	1324 m	120°23′46″	43°32′52″	5	15	4	0.9	0.00064
Hohhot, Inner Mongolia	MGHS	2018.8	1039 m	111°46′34″	40°35′15″	5	10	5	1	0.00044
Xilingol League, Inner Mongolia	MGXL	2018.9	992 m	116°0′97″	43°95′26″	5	10	3	0.7	0.00038
Guyuan, Ningxia	NGNZ	2018.9	1584 m	106°10′42″	36°07′06″	20	56	13	0.884	0.00063
Yinchuan, Ningxia	NSNY	2018.8	1041 m	106°12′49″	38°28′22″	20	70	14	0.937	0.00226
Wuzhong, Ningxia	NWNW	2019.8	1522 m	106°25′32″	37°0′29″	15	59	12	0.962	0.00119
Liangshan, Sichuan	SCLS	2019.8	2240 m	101°27′23	27°29′45″	5	2	2	0.4	0.00007
Qingdao, Shandong	SDQD	2018.8	10 m	120°4′52″	36°26′10″	5	45	4	0.9	0.00164
Shuozhou, Shanxi	SXSZ	2018.8	1032 m	112°36′24″	39°11′34″	5	5	5	1	0.0002
Changji, Xinjiang	XJCJ	2018.8	504 m	86°37′41″	44°14′08″	15	64	8	0.848	0.00172
Binchuan District, Dali, Yunnan	YNBC	2019.4	1400 m	100°34′42″	25°50′6″	5	0	1	0	0
Zhenkang District, Lincang, Yunnan	YNZK	2019.4	996 m	98°55′33″	23°45′22″	5	15	2	0.4	0.00055
Total						167	202	87	0.976	0.00197

N, number of samples; S, the number of segregating sites; Nh, the number of haplotypes; Hd, haplotype diversity; π, nucleotide diversity.

Genetic Diversity and Population Differentiation

The number of polymorphic sites (S), the number of haplotypes (Nh), nucleotide diversity (π), and haplotype diversity (Hd) were calculated using DnaSP 6.0 (Rozas et al., 2017). To check the differentiation levels in different populations and lineages, we calculated pairwise genetic differentiation (F_{ST}) and obtained *p*-values for F_{ST} values with 3,000 permutations using Arlequin 3.5 (Excoffier and Lischer, 2010), and calculated genetic distance (uncorrected *p*-distance) using MEGA7 (Kumar et al., 2016). Mantel tests were used to test the isolation-by-distance model for all geographic populations except Changde, Hunan population (HNCD) including only two samples. Correlations between F_{ST} and linear geographic distance (in km) were plotted with 999 replicates in the ade4 v. 1.7 module of R.

Phylogenetic Analyses and Divergence Time Estimation

Phylogenetic trees were constructed by the Bayesian inference (BI) method using PhyloBayes MPI 1.5a (Lartillot et al., 2013) and the maximum-likelihood (ML) method using IQ-TREE 1.6.5 (Trifinopoulos et al., 2016). PartitionFinder2 (Lanfear et al., 2016) on the CIPRES Science Gateway (Miller et al., 2010) was used to select the optimal partition schemes and substitution models. Thirty-nine (13 × 3) codon partitions were predefined for this dataset, and we used the "greedy" algorithm with branch lengths estimated as "unlinked" and the Akaike information criterion to select the partition schemes and models (summarized in **Supplementary Table 3**). For the ML analysis, the best evolutionary models predicted before were selected under the corrected Bayesian information criterion (BIC) in IQ-TREE, and the ML tree was obtained with IQ-TREE using an ultrafast bootstrap approximation approach with 10,000 replicates. For BI analysis, we chose the site-heterogeneous mixture model (CAT + GTR), with four independent chains starting from a random tree and running for 30,000 cycles. Trees being sampled every 10 cycles and the initial 25% trees of each MCMC run were discarded as burn-in. A consensus tree was computed from the remaining trees combined from two runs which converged at a MaxDiff < 0.1.

Clustering of individuals was performed with BAPS v. 6.0 (Cheng et al., 2013) under the spatial clustering of groups of individuals. Based on the same dataset, the median-joining network of haplotypes was constructed in PopART (Leigh and Bryant, 2015) to further analyze the relationship among different haplotypes. The frequency of haplotypes was obtained using Arlequin software. We also carried out the hierarchical analysis of molecular variance (AMOVA) among populations and between defined lineages (the result of BAPS) with the Arlequin program to gain a better insight of the biogeographical processes shaping the genetic structure. The significance of differentiation between pairs of populations/groups was evaluated by 1,000 bootstrap replications (Excoffier et al., 1992).

We estimated divergence time using BEAST 2.4.8 (Bouckaert et al., 2014). Given that there were no fossils available to calibrate the nodes, we used a molecular clock for divergence time estimation. *COI* gene was evaluated as a separate partition

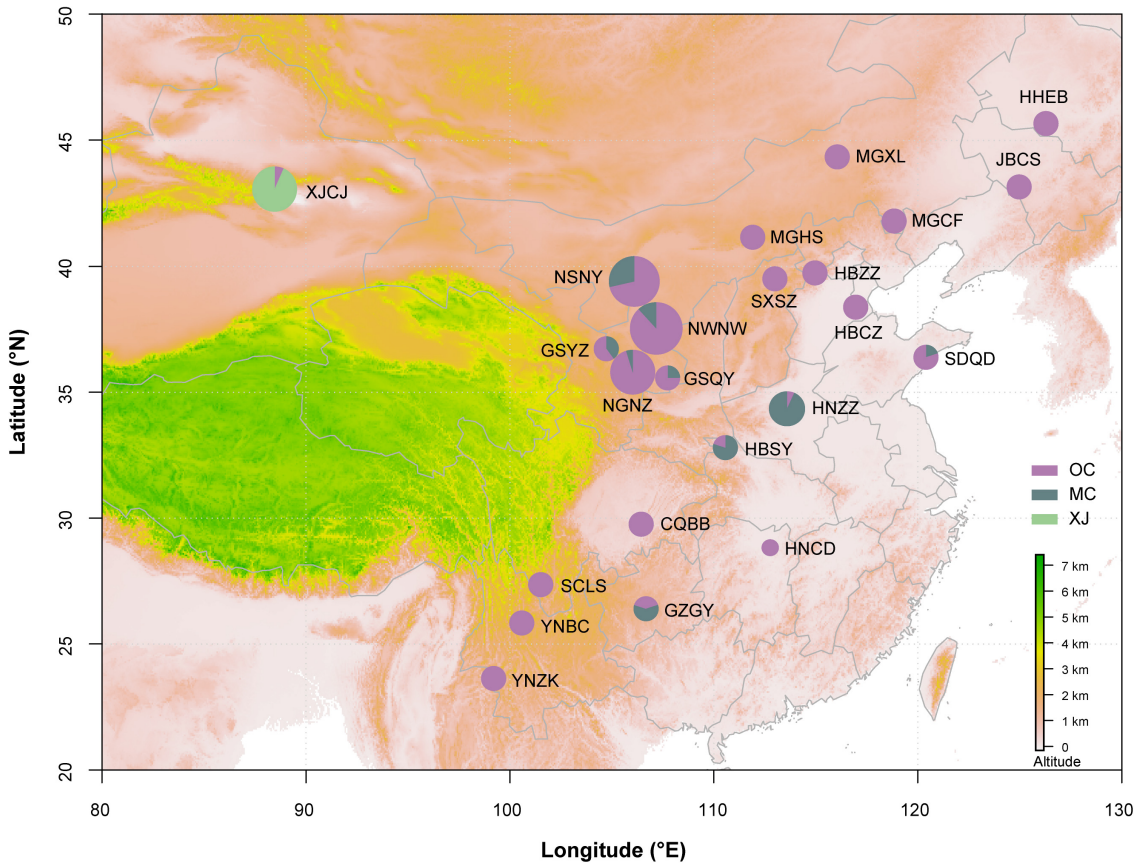


FIGURE 1 | Geographic distribution of samples on the map. Different colors of the circles represent phylogeographic lineages. The information of population code refers to **Table 1**. The area of circle represents the population size.

with the substitution rate fixed at 1.77% per site per million years (Mya) (Papadopoulou et al., 2010). The other 12 PCGs were applied as another partition with the evolutionary rate estimated accordingly. An uncorrelated lognormal relaxed clock model was applied with the birth-death model for the tree prior. Two independent MCMC runs per 500 million generations were performed with tree sampling every 50,000 generations. Tracer v. 1.7 (Rambaut et al., 2018) was used to verify whether the MCMC runs reached a stationary distribution based on the effective sample sizes (ESS) of each estimated parameter. We required ESS for the posterior, prior, and tree likelihood to be at least 200. The resulting trees for each replicate were combined using LogCombiner program in BEAST, and TreeAnnotator program in BEAST was used to calculate the consensus tree and annotate the divergence times with "mean height" after discarding the initial 25% trees as burn-in.

Demographic History

To test whether expansion occurred, we performed mismatch distribution analyses in Arlequin with 1,000 bootstrap replicates to detect expansion through the linear fitting of observed and simulated curves with the significance of two parameters, the sum of squares deviations (*SSD*) and Harpending's raggedness index (*r*) (Excoffier and Lischer, 2010). Bayesian skyline plots (BSPs) were obtained for different lineages in BEAST under Coalescent Bayesian Skyline model for the tree prior. The same partitions and substitution rates were applied consistent with these in the estimation of divergence time. BEAST analysis was conducted for 300 million generations with sampling trees every 30,000 generations. We determined the effective population size through time using Tracer, discarding the initial 10% of generations as burn-in.

RESULTS

Genetic Diversity, Population Differentiation, and Geographical Structure

We obtained the whole mitogenome sequences for 167 individuals from 23 SAA populations across China (**Table 1**), with the size of mitogenome ranging from 16,068 to 16,083 bp. We detected 87 haplotypes in total (**Table 1** and **Supplementary Table 4**), with 202 polymorphic sites (S) in average. Overall, populations showed a high haplotype diversity (Hd = 0.976) and a low nucleotide diversity (π = 0.00197). Among all the different

TABLE 2 | Genetic differentiation (F_{ST}) (lower left) and *p*-distance analysis (upper right) of *T. trifolii* among populations.

Code	XJCJ	HNZZ	HBSY	GZGY	GSYZ	NSNY	GSQY	NWNW	NGNZ	SDQD	CQBB	HBCZ	HBZZ	HHEB	HNCD	JBCS	MGCF	MGHS	MGXL	SCLS	SXSZ	YNBC	YNZK
XJCJ		0.0031	0.0033	0.0032	0.0033	0.0033	0.0032	0.0033	0.0034	0.0032	0.0034	0.0032	0.0032	0.0030	0.0034	0.0032	0.0033	0.0033	0.0032	0.0032	0.0032	0.0032	0.0035
HNZZ	0.5081*		0.0015	0.0020	0.0026	0.0027	0.0020	0.0033	0.0035	0.0031	0.0036	0.0036	0.0036	0.0035	0.0037	0.0036	0.0037	0.0037	0.0036	0.0035	0.0036	0.0035	0.0038
HBSY	0.4671*	0.0134		0.0020	0.0026	0.0028	0.0020	0.0032	0.0033	0.0031	0.0035	0.0035	0.0035	0.0033	0.0034	0.0035	0.0035	0.0034	0.0034	0.0033	0.0034	0.0033	0.0036
GZGY	0.3897*	0.1062	−0.0612		0.0023	0.0025	0.0020	0.0025	0.0026	0.0025	0.0025	0.0026	0.0026	0.0025	0.0026	0.0027	0.0027	0.0027	0.0026	0.0024	0.0026	0.0024	0.0029
GSYZ	0.3852*	0.3038*	0.1596	−0.0748		0.0022	0.0023	0.0019	0.0018	0.0021	0.0018	0.0020	0.0020	0.0018	0.0017	0.0020	0.0020	0.0019	0.0019	0.0017	0.0018	0.0016	0.0021
NSNY	0.3979*	0.3294*	0.2459*	0.0571	−0.0908		0.0025	0.0019	0.0018	0.0021	0.0019	0.0021	0.0021	0.0019	0.0017	0.0021	0.0021	0.0020	0.0019	0.0017	0.0019	0.0017	0.0021
GSQY	0.3814*	0.0871	−0.1152	−0.2189	−0.0975	0.0390		0.0025	0.0026	0.0025	0.0026	0.0027	0.0027	0.0025	0.0026	0.0027	0.0027	0.0026	0.0026	0.0025	0.0026	0.0024	0.0028
NWNW	0.5612*	0.6274*	0.5623*	0.3477*	0.0678	0.0936	0.3303*		0.0009	0.0015	0.0008	0.0011	0.0011	0.0009	0.0007	0.0011	0.0011	0.0010	0.0010	0.0007	0.0009	0.0007	0.0012
NGNZ	0.6667*	0.7556*	0.7251*	0.5540*	0.2739	0.2080*	0.5390*	−0.0062		0.0014	0.0006	0.0008	0.0008	0.0007	0.0004	0.0008	0.0009	0.0007	0.0007	0.0004	0.0006	0.0004	0.0009
SDQD	0.4769*	0.5458*	0.4369*	0.1728	−0.0018	0.0635	0.1774*	0.0993	0.2761*		0.0011	0.0011	0.0011	0.0011	0.0013	0.0012	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0014
CQBB	0.6410*	0.7684*	0.7356*	0.5000	0.3100*	0.2320*	0.5140*	0.1495*	0.2748*	0.2623*		0.0006	0.0006	0.0005	0.0004	0.0007	0.0008	0.0006	0.0006	0.0001	0.0005	0.0002	0.0008
HBCZ	0.5906*	0.7266*	0.6745*	0.4540*	0.2569*	0.2271*	0.4493*	0.1890	0.3284*	0.0240	0.6129*		0.0004	0.0003	0.0007	0.0005	0.0007	0.0006	0.0006	0.0005	0.0005	0.0005	0.0008
HBZZ	0.5904*	0.7266*	0.6745*	0.4540*	0.2569*	0.2271*	0.4493*	0.1890	0.3284*	0.0240	0.6129*	−0.1765		0.0003	0.0007	0.0005	0.0007	0.0006	0.0006	0.0005	0.0005	0.0005	0.0008
HHEB	0.5883*	0.7371*	0.6923	0.4710*	0.2475	0.2062*	0.4565*	0.1401*	0.2720*	0.1293*	0.7963*	0.0885	0.0885		0.0005	0.0003	0.0005	0.0004	0.0004	0.0003	0.0003	0.0003	0.0006
HNCD	0.5682*	0.7044*	0.5983*	0.3216	−0.0170	0.0155	0.2948	−0.1856	−0.2310	0.0977	1.0000	0.5186*	0.5186	0.6893		0.0007	0.0007	0.0005	0.0005	0.0002	0.0004	0.0002	0.0007
JBCS	0.5933*	0.7351*	0.6875*	0.4806*	0.2831*	0.2430*	0.4686*	0.2157*	0.3623*	0.1959*	0.7297*	0.1270	0.1270	0.0549	0.5939*		0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0008
MGCF	0.5784*	0.7127*	0.6469*	0.4356*	0.2136	0.2021	0.4177	0.1445*	0.2630*	0.1344*	0.5833*	0.1940*	0.1940	0.1353	0.3035	0.1667		0.0006	0.0005	0.0006	0.0005	0.0006	0.0008
MGHS	0.5995*	0.7300*	0.6704*	0.4570*	0.2166	0.1938*	0.4372*	0.0932*	0.1801*	0.1618*	0.6364*	0.2857*	0.2941*	0.2708*	0.3113	0.3605*	0.1374		0.0004	0.0005	0.0003	0.0004	0.0006
MGXL	0.5900*	0.7280*	0.6706*	0.4568*	0.2221	0.1937*	0.4366*	0.1276	0.2380*	0.1488*	0.7000*	0.2969*	0.2969*	0.2727*	0.4309	0.3513	0.0210	0.0302		0.0005	0.0003	0.0005	0.0006
SCLS	0.6204*	0.7535*	0.7131*	0.4790	0.2283	0.1736*	0.4704	−0.0148	0.0367	0.2088*	0.7500	0.4964*	0.4964*	0.6053*	0.7436*	0.6078*	0.4265*	0.4400*	0.5370*		0.0004	0.0000	0.0007
SXSZ	0.6096*	0.7456*	0.6966*	0.4866*	0.2460*	0.2007*	0.4651*	0.1183	0.2136*	0.1984*	0.8167*	0.3966*	0.3966*	0.4211*	0.5872*	0.4655*	0.1786*	−0.0058	−0.0390	0.6591*		0.0004	0.0005
YNBC	0.6227*	0.7580*	0.7210	0.4925	0.2333	0.1731	0.4774	−0.0208	0.0273	0.2373*	1.0000*	0.5556*	0.5556*	0.6765*	1.0000	0.6552*	0.4531*	0.4783*	0.5800*	0.0000	0.7250*		0.0007
YNZK	0.6332*	0.7483*	0.6925*	0.5175*	0.3438*	0.3047*	0.5000*	0.3327*	0.4679*	0.3407*	0.7500*	0.5263*	0.5263*	0.5638*	0.5522	0.5614*	0.4196*	0.4490*	0.4575*	0.6731*	0.5119*	0.7000*	

*$P < 0.05$.

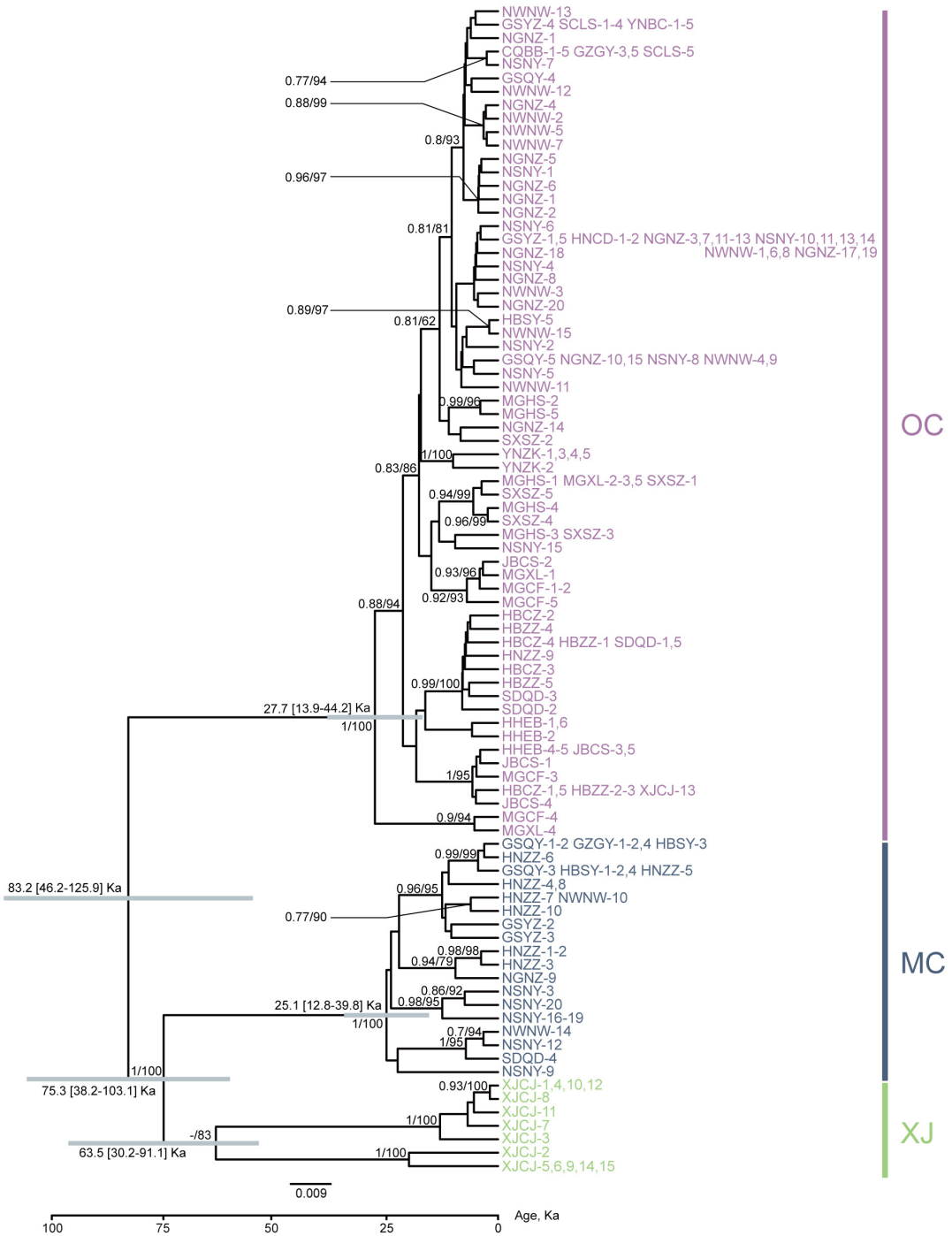


FIGURE 2 | Phylogenetic relationships of major nodes and divergence time inferred from the protein-coding gene (PCG) dataset. Bars in gray show 95% highest posterior density intervals. The node supports shown are posterior probabilities of Bayesian inference and bootstrap percentages in a maximum-likelihood analysis. The information of population code refers to **Table 1** and **Figure 1**. Numbers represent different individuals sharing the same haplotype.

geographic populations, GSQY showed the highest nucleotide diversity (π = 0.00253) and three populations showed lowest (π = 0.00000), including CQBB, HNCD, and YNBC.

We estimated pairwise F_{ST} and uncorrected *p*-distance between 23 different geographical populations (**Table 2**). The F_{ST} value between CQBB, HNCD, and YNBC was the highest (1.0000), and the lowest was between HNCD and NGNZ (−0.2310). For the *p*-distance, the value between HNZZ and YNZK was the highest (0.0038), and the lowest was between SCLS and YNBC (0.0000). Populations from northern China

showed low *p*-distance values from each other. HBSY, HNZZ, and XJCJ populations showed higher F_{ST} and *p*-distance values than those from other populations. We also calculated the *p*-distance between 87 haplotypes, and the results ranged from 0.0001 to 0.0045 (**Supplementary Table 5**). The Mantel test indicated that the correlation between geographic distance and genetic differentiation was not significant ($r = -0.0374$, $p = 0.7120$, **Supplementary Figure 1**).

Phylogenetic Analyses and Divergence Times

Phylogenetic relationships of major nodes in the ML, BI, and BEAST trees showed similar topologies, even though the nodal supports of some pivotal nodes were not high (**Figure 2**). The monophyly of the three lineages was well supported: the first lineage included all samples from XJCJ and one individual from another lineage, the second lineage included samples mainly from middle China, and the last lineage included samples from various populations, ranging from northeast to southwest of China. Therefore, we assigned the name of lineages according to their main geographic region, i.e., Xinjiang (XJ), middle China (MC), and other populations within China (OC) lineages, respectively (**Figure 2**). In MC lineage, most individuals from Ningxia area diverged first followed by samples from nearby cities. Individuals of OC lineage that diverged first were all from north and northeast of China, and individuals that diverged later were mostly from northwest and southwest of China. BAPS result also showed that all samples were divided into three clusters (**Figure 1**); XJ cluster (blue color in **Figure 1**) was only distributed in Xinjiang province, the MC cluster (red color in **Figure 1**) was mainly distributed in the north and middle of China, and the OC cluster (green color in **Figure 1**) was distributed as a majority (the haplotype from this cluster can be found in every geographical population).

Among the lineages revealed by the phylogenetic analyses, OC lineage with the largest sample size showed the lowest diversity, followed by MC lineage (**Table 3**), and XJ lineage exhibited the highest values even though it comprised the fewest individuals (**Table 3**). The *p*-distance among three lineages varied from 0.0031 to 0.0041 (**Table 4**). AMOVA analysis showed that the greatest proportion of variation was among lineages (81.79%) and the genetic differentiation among all of the populations was $F_{ST} = 0.877$, $p < 0.0001$ (**Table 5**).

The median-joining network of haplotypes revealed three haplotype groups (**Figure 3**), consistent with the results of BAPS and phylogenetic trees. For the OC lineage, there were several star-like topologies observed, indicating the recent expansion in this lineage. Haplotype H6 was the most frequently shared haplotype, which included 18 individuals from 10 different populations, followed by haplotype H9, which included 10 individuals from three populations. There were many mutation steps observed between the three lineages in median-joining network.

Our divergence time estimation suggested that the most recent common ancestor (MRCA) of SAA was 83.2 Ka (kilo-annum) and diverged into the MRCA of XJ and MC, and OC lineages with the ages of 75.3 and 27.7 Ka (**Figure 1**). The MRCA ages of XJ and MC were 63.5 and 25.1 Ka, respectively. Thus, all three lineages diverged during the late Pleistocene Quaternary.

TABLE 3 | Parameters of genetic diversity of three lineages.

Lineages	*N*	*S*	Nh	Hd	π
OC	120	97	62	0.961	0.00048
MC	33	53	18	0.936	0.00079
XJ	14	41	7	0.824	0.00146
Total	167	202	87	0.976	0.00197

TABLE 4 | The *p*-distance analysis of *T. trifolii* among lineages.

Lineages	OC	MC	XJ
OC	0	0.0041	0.0035
MC		0	0.0031
XJ			0

Demographic History

In the mismatch distribution analyses (**Figure 4**), the observed and simulated curves did not significantly differ from each other in OC lineage with the parameters of *SSD* and *r* being small and insignificant ($P > 0.05$). Hence, the population expansion process was supported in this lineage. In contrast, XJ and MC lineages exhibited relatively large and significant parameters of *SSD* and *r*. Thus, the population expansion event was not supported in these two lineages. BSP analysis (**Figure 4**) revealed that OC lineage and total samples experienced a significant increase in effective population size about 1500–2000 years ago, which was a very recent expansion event. However, the population size of MC and XJ groups showed a slight decrease after a long steady period.

DISCUSSION

This study investigated the population genetic structure of SAA in China using complete genome sequences. Population genetic

TABLE 5 | Analysis of molecular variance (AMOVA) among populations and between lineages of *T. trifolii*.

Source of variation	*d.f.*	Sum of squares	Percentage of variation	*F* (*p*-value)
Among lineages	2	1234.395	81.79	$F_{CT} = 0.818$ ($p < 0.0001$)
Among populations within lineages	30	240.824	5.95	$F_{ST} = 0.877$ ($p < 0.0001$)
Within populations	134	331.237	12.25	$F_{SC} = 0.327$ ($p < 0.0001$)
Total	166	1806.456		

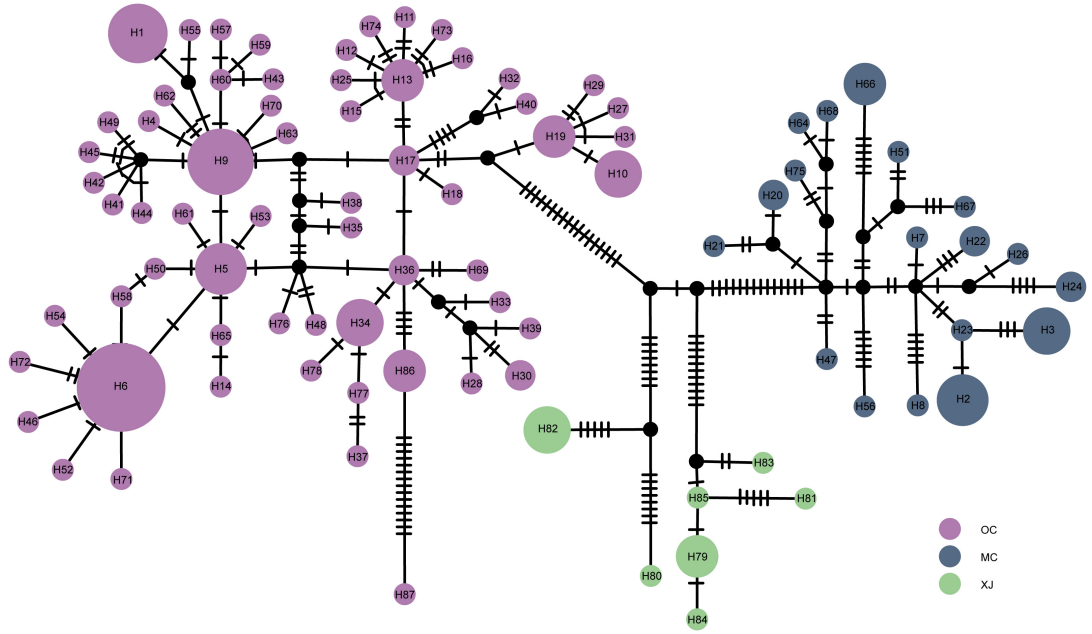


FIGURE 3 | Median-joining network of haplotypes based on the mitogenome. Colored circles represent haplotypes, and black circles represent missing haplotypes that were not observed. The area of the circle is proportional to the number of haplotypes.

structure of species can be inferred from mitochondrial genes and nuclear markers (Zhang et al., 2017; Cao et al., 2021). However, the complete mitogenome may accumulate more variations and provide valuable information for population genetic structure inference (Du et al., 2019, 2021). Although the detected variations are from the same locus, our analyses provided a first insight into genetic diversity and population structure of this economically significant invasive pest.

Overall, the SAA populations in China display relatively low nucleotide diversity but high haplotype diversity. The low nucleotide diversity can be explained by the recent introduction event of SAA, which is associated with the resent expansion and bottleneck effects. Besides, the unique reproduction method of aphid involving parthenogenesis might also lead to the low genetic diversity (Hales et al., 1997). The high haplotype diversity, on the other hand, may be caused by the multiple massive introductions, offsetting the bottleneck effect. There are also some populations, such as GSQY, GSYZ, GZGY, and NSNY, showing a relatively high genetic diversity. This is probably attributed to the mixing OC and MC lineages in those populations.

Based on our analysis, we clarify that the SAA populations in China are composed of three genetically distinct lineages. We inferred that the three lineages are most likely formed by human-mediated introduction instead of natural dispersal, and each lineage possibly has a different invasion source. First of all, the time of the introduction of SAA's host, *M. sativa*, in mainland was around 2,000 years ago (Gao, 2012). The divergent time estimation suggests that three groups have been diverged long before the beginning of alfalfa plantation in China. In addition, the Mantel test also supported that the geographical distances were not the main reason of diversification among populations. Moreover, according to the historical records, China's alfalfa fodder used in animal husbandry, especially the dairy industry, depended heavily on the importation from foreign countries since the early 21st century, including the United States, Canada, Australia, and some European countries such as France, Germany, and Spain.[2] Besides, some Chinese fodder companies also introduced high-quality species of lucerne from other continents and long-distance dispersal of the pest might be carried out as a result.[3] Subsequently, the domestic alfalfa trade in China may further facilitate the expansion of different lineages across the country. The restricted distribution of XJ lineage may be a result of the remote nature of Xinjiang province and limited inter-province transportations between Xinjiang and other provinces.

Biological invasions showed a natural rapid evolutionary process in contemporary time scale and the invasive species may present a major threat to biodiversity and ecosystem integrity (McKinney and Lockwood, 1999; Olden et al., 2004). The multiple invasion event suspected for SAA is common for invasive alien species (Roman and Darling, 2007; Dlugosch and Parker, 2010). Moreover, most invasions of non-indigenous species were the result of commercial trades among different continents and countries in the last decade. For example, the spotted lanternfly (*Lycorma delicatula*) is native to China and invaded some Asian countries and north America possibly through timber transportation (Du et al., 2021); the box tree moth (*Cydalima perspectalis*), native to Asia, spread rapidly

[2]http://www.dairyfarmer.com.cn/

[3]http://www.zgmcwz.com

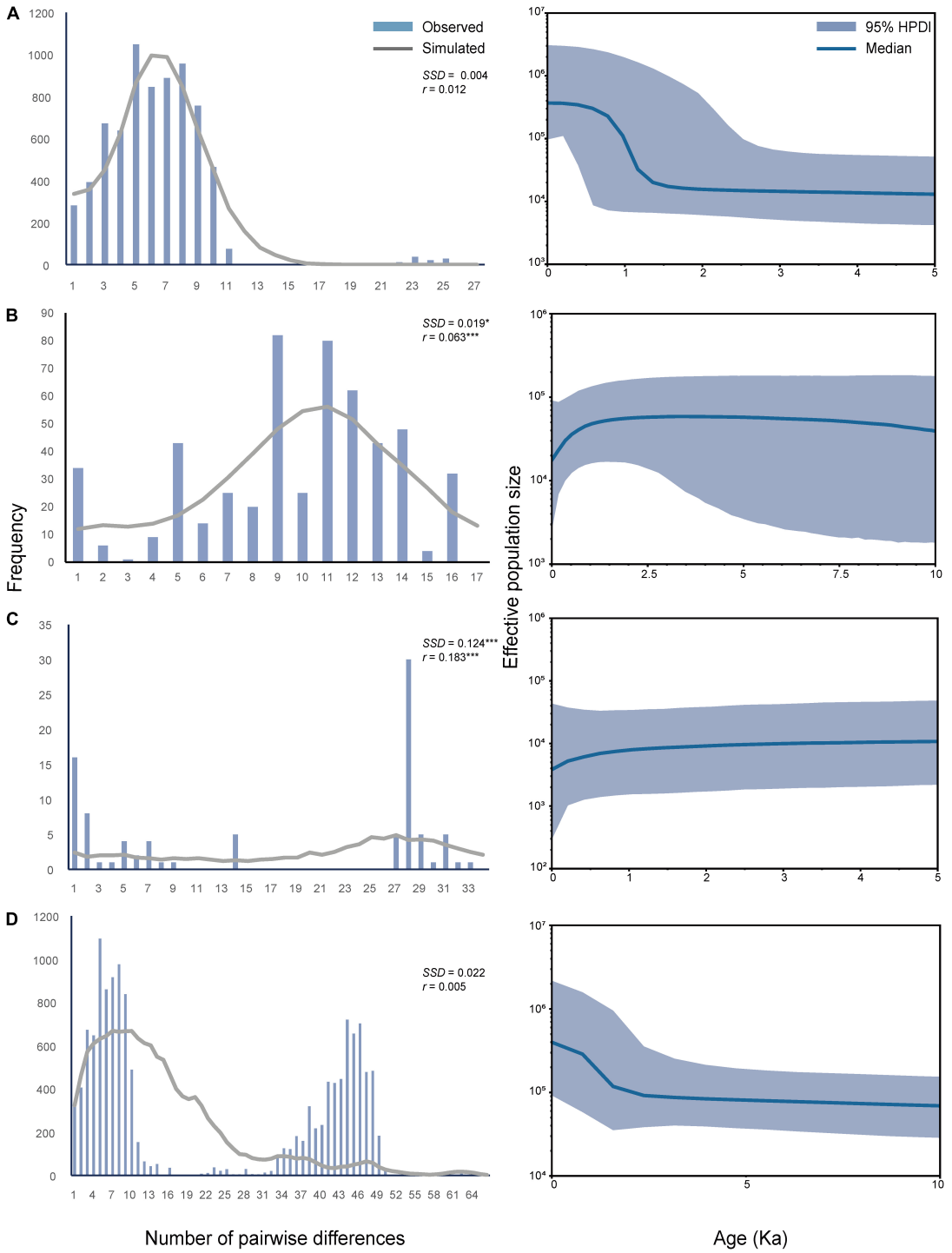


FIGURE 4 | Mismatch distribution analysis and the demographic history of three lineages and whole samples of *T. trifolii* based on the mitogenome sequences. **(A)** OC lineage. **(B)** MC lineage. **(C)** XJ lineage. **(D)** All samples. *SSD*, sum of square deviations; *r*, Harpending's raggedness index; $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$. For Bayesian skyline plots, the median estimated effective population sizes (middle lines) are enclosed within the 95% highest posterior density intervals (HPDI, shaded areas).

because of the trade of ornamental box trees (Bras et al., 2019); the Asian long-horned beetle (*Anoplophora glabripennis*), native to East Asia, was introduced to other countries via solid wood packing material (Javal et al., 2017); and *Drosophila suzukii*, indigenous in eastern Asia, had widely expanded its range and became a serious pest in some areas with its introductions related to the trade of host fruits (Vega et al., 2020). At present, the large-scale alfalfa transportation is still

ongoing between China and other countries. The suspected multiple invasions in this study highlight the importance of developing effective quarantine measures to prevent continuous introductions from source regions.

Our study provides an insight into the genetic structure of SAA populations in China. The results indicated the presumed multiple introductions. This research included populations from China only, and we intend to collect more samples from other continents to obtain a clearer genetic structure and more accurate invasion history on the dynamics of the introductions. Combining the samples collected from overseas, we may trace its origins and invasion pathways and obtain a better understanding of the global distributions of the invasive populations at different scales. Besides, higher resolution molecular markers, such as single nucleotide polymorphisms obtained from whole-genome resequencing method, are expected in the future studies.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, MW885588–MW885754.

AUTHOR CONTRIBUTIONS

WC: conceptualization, methodology, and writing—review and editing. RZ: conceptualization and methodology. XL: investigation, formal analysis, and writing—original draft. SW, JH, and XZ: investigation. ZD: formal analysis and writing—original draft. HL: methodology and writing—review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants from the Ningxia Province Sci-Tech Innovation Demonstration Program of High-Quality Agricultural Development and Ecological Conservation (NGSB-2021-15-04) and the Key Research and Development Program of Ningxia Province (2017BY080).

ACKNOWLEDGMENTS

We are grateful to Guohua Huang, Minglong Yuan, Tao Li, and many other colleagues for their assistance in sample collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2021.759496/full#supplementary-material

REFERENCES

Abascal, F., Zardoya, R., and Telford, M. J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* 38(Suppl. 2), 7–13.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., et al. (2013). MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* 69, 313–319. doi: 10.1016/j.ympev.2012.08.023

Bishop, J. L., and Crockett, D. (1961). *The Spotted Alfalfa Aphid in Virginia*. Glade Spring, VA: Virginia Agricultural Experimental Station, 22.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., et al. (2014). BEAST2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10:e1003537. doi: 10.1371/journal.pcbi.1003537

Bras, A., Avtzis, D. N., Kenis, M., Li, H., Vétek, G., Bernard, A., et al. (2019). A complex invasion story underlies the fast spread of the invasive box tree moth (*Cydalima perspectalis*) across Europe. *J. Pest Sci.* 92, 1187–1202. doi: 10.1007/s10340-019-01111-x

Cao, L. J., Li, B. Y., Chen, J. C., Zhu, J. Y., Hoffmann, A. A., and Wei, S. J. (2021). Local climate adaptation and gene flow in the native range of two co-occurring fruit moths with contrasting invasiveness. *Mol. Ecol.* 30, 4204–4219. doi: 10.1111/mec.16055

Carver, M. (1978). The scientific nomenclature of the spotted alfalfa aphid (Homoptera: Aphididae). *Aust. J. Entomol.* 17, 287–288. doi: 10.1111/j.1440-6055.1978.tb00159.x

Cheng, L., Connor, T. R., Siren, J., Aanensen, D. M., and Corander, J. (2013). Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Mol. Biol. Evol.* 30, 1224–1228. doi: 10.1093/molbev/mst028

Ciosi, M., Miller, N. J., Kim, K. S., Giordano, R., Estoup, A., and Guillemaud, T. (2008). Invasion of Europe by the western corn rootworm, *Diabrotica virgifera virgifera*: multiple transatlantic introductions with various reductions of genetic diversity. *Mol. Ecol.* 17, 3614–3627. doi: 10.1111/j.1365-294X.2008.03866.x

Dlugosch, K. M., and Parker, I. M. (2010). Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Mol. Ecol.* 17, 431–449. doi: 10.1111/j.1365-294X.2007.03538.x

Drag, L., Hauck, D., Bérces, S., Michalcewicz, J., Jelaska, L., Aurenhammer, S., et al. (2015). Genetic differentiation of populations of the threatened saproxylic beetle *Rosalia longicorn*, *Rosalia alpina* (Coleoptera: Cerambycidae) in Central and South-East Europe. *Biol. J. Linn. Soc.* 116, 911–925. doi: 10.1111/bij.12624

Du, Z., Hasegawa, H., Cooley, J. R., Simon, C., Yoshimura, J., Cai, W., et al. (2019). Mitochondrial genomics reveals shared phylogeographic patterns and demographic history among three periodical cicada species groups. *Mol. Biol. Evol.* 36, 1187–1200. doi: 10.1093/molbev/msz051

Du, Z., Wu, Y., Chen, Z., Cao, L., Tadashi, I., Satoshi, K., et al. (2021). Global phylogeography and invasion history of the spotted lanternfly revealed by mitochondrial phylogenomics. *Evol. Appl.* 14, 915–930. doi: 10.1111/eva.13170

Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863–864. doi: 10.1093/bioinformatics/btr026

Estoup, A., and Guillemaud, T. (2010). Reconstructing routes of invasion using genetic data: why, how and so what? *Mol. Ecol.* 19, 4113–4130. doi: 10.1111/j.1365-294X.2010.04773.x

Estoup, A., Ravignné, V., Hufbauer, R., Vitalis, R., Gautier, M., and Facon, B. (2016). Is there a genetic paradox of biological invasion? *Annu. Rev. Ecol. Evol. Syst.* 47, 51–72. doi: 10.1146/annurev-ecolsys-121415-032116

Excoffier, L., and Lischer, H. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10, 564–567. doi: 10.1111/j.1755-0998.2010.02847.x

Excoffier, L., Smouse, P. E., and Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application

to human mitochondrial DNA restriction data. *Genetics* 131, 479–491. doi: 10.1093/genetics/131.2.479

Facon, B., Pointier, J. P., Glaubrecht, M., Poux, C., and David, P. (2003). A molecular phylogeography approach to biological invasions of the new world by parthenogenetic Thiarid snails. *Mol. Ecol.* 12, 3027–3039. doi: 10.1046/j.1365-294X.2003.01972.x

Gao, F. (2012). *Study on Climatic Regionalization of Alfalfa based on Fall Dormancy in China*. Ph.D. dissertation, Beijing: Beijing Forestry University.

Giles, K. L., Berberet, R. C., Zarrabi, A. A., and Dillwith, J. W. (2002). Influence of alfalfa cultivar on suitability of *Acyrthosiphon kondoi* (Homoptera: Aphididae) for survival and development of *Hippodamia convergens* and *Coccinella septempunctata* (Coleoptera: Coccinellidae). *J. Econ. Entomol.* 95, 552–557. doi: 10.1603/0022-0493-95.3.552

Hales, D. F., Tomiuk, J., Wöhrmann, K., and Sunnucks, P. (1997). Evolutionary and genetic aspects of aphid biology: a review. *Eur. J. Entomol.* 94, 1–55.

Howe, W. L., and Smith, O. F. (1957). Resistance to the spotted alfalfa aphid in lahontan alfalfa. *J. Econ. Entomol.* 50, 320–324. doi: 10.1093/jee/50.3.320

Hughes, R. D., Woolcock, L. T., Roberts, J. A., and Hughes, M. A. (1987). Biological control of the spotted alfalfa aphid, *Therioaphis trifolii* f. *maculata*, on lucerne crops in Australia, by the introduced parasitic hymenopteran *Trioxys complanatus*. *J. Appl. Ecol.* 24, 515–537. doi: 10.2307/2403890

Javal, M., Roques, A., Haran, J., Hérard, F., Keena, M., and Roux, G. (2017). Complex invasion history of the Asian long-horned beetle: fifteen years after first detection in Europe. *J. Pest Sci.* 92, 173–187. doi: 10.1007/s10340-017-0917-1

Kang, M., Buckley, Y. M., and Lowe, A. J. (2007). Testing the role of genetic factors across multiple independent invasions of the shrub Scotch broom (*Cytisus scoparius*). *Mol. Ecol.* 16, 4662–4673. doi: 10.1111/j.1365-294X.2007.03536.x

Kolbe, J. J., Glor, R. E., Schettino, L. R., Lara, A. C., Larson, A., and Losos, J. B. (2004). Genetic variation increases during biological invasion by a Cuban lizard. *Nature* 431, 177–181. doi: 10.1038/nature02807

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Lambers, D. R., and Van den Bosch, R. (1964). On the genus *Therioaphis* Walker, 1870 with descriptions of new species (Homoptera: Aphididae). *Zoologische Verhandelingen* 68, 1–47.

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., and Calcott, B. (2016). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34, 772–773. doi: 10.1093/molbev/msw260

Lartillot, N., Rodrigue, N., Stubbs, D., and Richer, J. (2013). PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* 62, 611–615. doi: 10.1093/sysbio/syt022

Lavergne, S., and Molofsky, J. (2007). Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proc. Natl. Acad. Sci. U.S.A.* 104, 3883–3888. doi: 10.1073/pnas.0607324104

Leigh, J. W., and Bryant, D. (2015). Popart: full-feature software for haplotype network construction. *Methods Ecol. Evol.* 6, 1110–1116. doi: 10.1111/2041-210X.12410

Liu, X., Wei, S., He, J., Song, F., and Cai, W. (2019). Complete mitochondrial genome of the spotted alfalfa aphid, *Therioaphis trifolii* (Hemiptera: Aphididae). *Mitochondrial DNA B Resour.* 4, 3260–3261. doi: 10.1080/23802359.2019.1644232

Lombaert, E., Guillemaud, T., Cornuet, J. M., Malausa, T., Facon, B., and Estoup, A. (2010). Bridgehead effect in the worldwide invasion of the biocontrol harlequin ladybird. *PLoS One* 5:e9743. doi: 10.1371/journal.pone.0009743

McKinney, M. L., and Lockwood, J. L. (1999). Biotic homogenization: a few winners replacing many losers in the next mass extinction. *Trends Ecol. Evol.* 14, 450–453. doi: 10.1016/S0169-5347(99)01679-1

Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). "Creating the CIPRES science gateway for inference of large phylogenetic trees," in *Proceedings of the 2010 Gateway Computing Environments Workshop (GCE 2010)* (New Orleans, LA: IEEE Computer Society). doi: 10.1109/GCE.2010.5676129

Monell, J. (1882). Notes on aphididae. *Can. Entomol.* 14, 13–16. doi: 10.4039/Ent1413-1

Nicolas, M., Davide, R., Hurley, B. P., Brockerhoff, E. G., and Haack, R. A. (2018). Common pathways by which non-native forest insects move internationally and domestically. *J. Pest Sci.* 92, 13–27. doi: 10.1007/s10340-018-0990-0

Olden, J. D., Poff, L. R., Douglas, M. R., Douglas, M. E., and Fausch, K. D. (2004). Ecological and evolutionary consequences of biotic homogenization. *Trends Ecol. Evol.* 19, 18–24. doi: 10.1016/j.tree.2003.09.010

Papadopoulou, A., Anastasiou, I., and Vogler, A. P. (2010). Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Mol. Biol. Evol.* 27, 1659–1672. doi: 10.1093/molbev/msq051

Peters, D. C., and Painter, R. H. (1957). A general classification of available small seeded legumes as hosts for three aphids of the "yellow clover aphid complex". *J. Econ. Entomol.* 50, 231–235. doi: 10.1093/jee/50.3.231

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* 67, 901–904. doi: 10.1093/sysbio/syy032

Roderick, G. K., and Navajas, M. (2003). Genes in new environments: genetics and evolution in biological control. *Nat. Rev. Genet.* 4, 889–899. doi: 10.1038/nrg1201

Roman, J., and Darling, J. A. (2007). Paradox lost: genetic diversity and the success of aquatic invasions. *Trends Ecol. Evol.* 22, 454–464. doi: 10.1016/j.tree.2007.07.002

Roques, A., Auger-Rozenberg, M. A., Blackburn, T. M., Garnas, J., Pyšek, P., Rabitsch, W., et al. (2016). Temporal and interspecific variation in rates of spread for insect species invading Europe during the last 200 years. *Biol. Invasions* 18, 907–920. doi: 10.1007/s10530-016-1080-y

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248

Simberloff, D., Martin, J. L., Genovesi, P., Maris, V., Wardle, D. A., Aronson, J., et al. (2013). Impacts of biological invasions: what's what and the way forward. *Trends Ecol. Evol.* 28, 58–66. doi: 10.1016/j.tree.2012.07.013

Trifinopoulos, J., Nguyen, L. T., Haeseler, A. V., and Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* 44, 232–235. doi: 10.1093/nar/gkw256

Van den Bosch, R. (1957). The spotted alfalfa aphid and its parasites in the Mediterranean region, Middle East, and East Africa. *J. Econ. Entomol.* 50, 352–356. doi: 10.1093/jee/50.3.352

Van den Bosch, R., Schlinger, E. I., and Dietrick, E. J. (1957). Imported parasites established: natural enemies of spotted alfalfa aphid brought from Middle East in 1955-56 now established in California. *Calif. Agr.* 11, 11–12.

Vega, G., Corley, J., and Soliani, C. (2020). Genetic assessment of the invasion history of *Drosophila suzukii* in Argentina. *J. Pest Sci.* 93, 63–75. doi: 10.1007/s10340-019-01149-x

Zhang, G., and Zhong, T. (1983). *Economic Insect Fauna of China, Homoptera: Aphids (1)*, Vol. 25. Beijing: Science Press.

Zhang, L., Cai, W., Luo, J., Zhang, S., Li, W., Wang, C., et al. (2017). Population genetic structure and expansion patterns of the cotton pest *Adelphocoris fasciaticollis*. *J. Pest Sci.* 91, 539–550. doi: 10.1007/s10340-017-0939-8

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 22 October 2021
doi: 10.3389/fevo.2021.750703

Environmental Drivers of Diversification and Hybridization in Neotropical Butterflies

*Nicol Rueda-M[1], Fabian C. Salgado-Roa[1,2], Carlos H. Gantiva-Q[3], Carolina Pardo-Díaz[1] and Camilo Salazar[1]**

[1] Department of Biology, Faculty of Natural Sciences, Universidad del Rosario, Bogotá, Colombia, [2] School of BioSciences, University of Melbourne, Melbourne, VIC, Australia, [3] Instituto de Ciencias Naturales, Universidad Nacional de Colombia, Bogotá, Colombia

Studying how the environment shapes current biodiversity patterns in species rich regions is a fundamental issue in biogeography, ecology, and conservation. However, in the Neotropics, the study of the forces driving species distribution and richness, is mostly based on vertebrates and plants. In this study, we used 54,392 georeferenced records for 46 species and 1,012 georeferenced records for 38 interspecific hybrids of the Neotropical *Heliconius* butterflies to investigate the role of the environment in shaping their distribution and richness, as well as their geographic patterns of phylogenetic diversity and phylogenetic endemism. We also evaluated whether niche similarity promotes hybridization in *Heliconius*. We found that these insects display five general distribution patterns mostly explained by precipitation and isothermality, and to a lesser extent, by altitude. Interestingly, altitude plays a major role as a predictor of species richness and phylogenetic diversity, while precipitation explains patterns of phylogenetic endemism. We did not find evidence supporting the role of the environment in facilitating hybridization because hybridizing species do not necessarily share the same climatic niche despite some of them having largely overlapping geographic distributions. Overall, we confirmed that, as in other organisms, high annual temperature, a constant supply of water, and spatio-topographic complexity are the main predictors of diversity in *Heliconius*. However, future studies at large scale need to investigate the effect of microclimate variables and ecological interactions.

Keywords: species distribution models, phylogenetic diversity, species richness, phylogenetic endemism, climatic niches, hybridization

OPEN ACCESS

Edited by:
Gengping Zhu,
Tianjin Normal University, China

Reviewed by:
Vijay Barve,
Florida Museum of Natural History, United States
Pedro Tarroso,
Centro de Investigacao em Biodiversidade e Recursos Geneticos (CIBIO-InBIO), Portugal

****Correspondence:***
Camilo Salazar
camilo.salazar@urosario.edu.co

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *31 July 2021*
Accepted: *04 October 2021*
Published: *22 October 2021*

Citation:
Rueda-M N, Salgado-Roa FC, Gantiva-Q CH, Pardo-Díaz C and Salazar C (2021) Environmental Drivers of Diversification and Hybridization in Neotropical Butterflies. Front. Ecol. Evol. 9:750703. doi: 10.3389/fevo.2021.750703

INTRODUCTION

Understanding how the environment shapes species distribution and affects patterns of biological diversity is still a challenging task, especially in species rich regions, such as the Neotropics (Hawkins et al., 2003; Gotelli et al., 2009; Brown et al., 2020). To date, information on this topic is mostly based on vertebrates and plants, and suggest that the combination of high annual temperature with a constant supply of water and spatio-topographic complexity are the main predictors of species distribution, richness, and endemism (Hawkins et al., 2003; Kreft and Jetz, 2007; Qian, 2010; Vasconcelos et al., 2019). Within the Neotropics, the Amazon and the foothills of

the North-eastern Andes are examples of regions that combine these conditions, and consequently, they exhibit high levels of species richness and phylogenetic diversity in monkeys, snakes, birds, amphibians, palms, and vascular plants (Kreft and Jetz, 2007; Fenker et al., 2014; Vallejos-Garrido et al., 2017; Velazco et al., 2021). Similarly, regions such as the Biogeographic Choco, Costa Rica, and the Amazon show high levels of phylogenetic endemism (e.g., Rosauer and Jetz, 2014; López-Aguirre et al., 2019; Varzinczak et al., 2020). However, these patterns have not been deeply evaluated in Neotropical invertebrates, and particularly butterflies (Pearson and Carroll, 2001; Mullen et al., 2011).

The environment, and especially climatic niche, has also been suggested to have an effect on gene flow. For example, phylogenetic discordance in multiple loci in beetles of the genus *Mesocarabus* seems to be the result of hybridization between species sharing the same climatic niche (Andújar et al., 2014), while in armadillos of the genus *Dasypus*, asymmetric gene flow appears to be facilitated by niche conservatisms at both sides of a geographic barrier (Arteaga et al., 2011). Additionally, climatic-based selection likely plays a role in maintaining mosaic hybrid zones in *Quercus* oaks, where climatic heterogeneity favors the co-occurrence of parental species and their hybrids (Swenson et al., 2008; Ortego et al., 2014).

Heliconius butterflies are a diverse insect group found across southern United States, Central, and South America, where they occupy divergent habitats (Jiggins, 2017). Due to the recent radiation of this butterfly genus, species pairs have different levels of reproductive isolation, which are used as proxies for different stages of speciation (Kronforst et al., 2013; Martin et al., 2013). In total, ∼25% of *Heliconius* species are known to hybridize in nature (Mallet et al., 1998, 2007), but the role of abiotic variables in facilitating or hampering such hybridization has been poorly studied (Mallet et al., 1990; Rosser et al., 2014).

In this study, we combined an extensive database of occurrences of species and hybrids in *Heliconius* as well as environmental data to investigate: (1) how the environment shapes the distribution of *Heliconius* at a regional scale, (2) how the environment molds species richness, phylogenetic diversity, and phylogenetic endemism in these butterflies, and (3) whether niche similarity promotes hybridization.

MATERIALS AND METHODS

Species Data and Environmental Variables

We included occurrence data of 46 species of *Heliconius* and generated a database of the localities where these butterflies have been collected across their entire distribution range. The data were obtained from: (1) entomological collections and (2) the Heliconiinae checklist of Rosser et al. (2012). For those regions in Colombia that we identified as under-sampled, we conducted field trips to improve our geographic coverage. The nomenclature of all records was updated to the most recent taxonomic checklist when needed (Lamas and Jiggins, 2017). We also included occurrence data for all interspecific hybrids documented in *Heliconius*. All individuals were photographed and identified based on their color pattern. We used the point-radius method to georeference specimens with missing coordinates following Wieczorek et al. (2004). Although *Heliconius* is widely represented in databases, such as global biodiversity information Facility (GBIF), we did not include such records to ensure the use of data that have been curated by specialists both in terms of georeference and taxonomy, or that have images of each specimen that would allow us to confirm the taxonomy.

We used the 19 climatological variables from climatologies at high resolution for the earth's land surface areas (CHELSA) at spatial resolution of 1 km (Karger et al., 2017) to characterize climatic variation across the occurrence range of *Heliconius*, and altitude was obtained from Jarvis et al. (2008). Collinearity between variables was avoided by estimating the Pearson correlation coefficient among all 20 variables, and the absolute value of this correlation was used to create a dissimilarity matrix (1-correlation values). We used this matrix to perform a hierarchical clustering analysis with the *hclust* function in R (R Core Team, 2021). We then chose one variable per cluster that had a pairwise distance <0.5. Using the selected variables, we calculated the variance inflation factor (VIF) (Dormann et al., 2013) with the *HH* package in R (Heiberger, 2020) and chose those variables with VIF <5 (Kubota et al., 2015).

Species Distribution Modeling and Environmental Variables Importance

First, we used R pipelines (Assis, 2020) to reduce sampling bias and spatial autocorrelation among occurrences in our species distribution models using the variables that passed the filters mentioned before. The minimum non-significant autocorrelated distances were used to prune species databases. *H. nattereri* and *H. tristero* were not modeled because they had <32 occurrence records.

Then, we generated a second database that included pseudo-absences data following Phillips et al. (2009), Soberón and Nakamura (2009), Barbet-Massin et al. (2012), and Lake et al. (2020). Because *Heliconius* is a very well-sampled genus we had enough information to select pseudo-absences points for each species in places where: (i) *Heliconius* other than the focal species have been collected, (ii) environmental conditions may not be optimal for its occurrence, and (iii) absence is not caused by dispersal limitation. Using these criteria, we defined a minimum convex polygon with a 50 km buffer area for each species and selected 10,000 pseudo-absences only in this buffer.

Then, we estimated the ensemble species distribution models (ESDMs) of *Heliconius* with the R package stacked species distribution models (*SSDM*) (Schmitt et al., 2017), equally weighting presences and pseudo-absences (prevalence weights = 0.5). Individual species distribution models (SDM) were implemented using four algorithms that optimize the use of pseudo-absences in a similar way (Barbet-Massin et al., 2012): (1) Generalized Linear Models (GLMs) (McCullagh and Nelder, 1989), (2) Generalized Boosting Models (GBMs) (Friedman et al., 2000), (3) Maximum Entropy Models (MAXENT)

(Phillips et al., 2006), and (4) Generalized additive model (GAM) (Hastie and Tibshirani, 1990). Each algorithm was run 10 times. In each run, models were calibrated using 75% of the occurrence data and their accuracy was evaluated with the remaining 25%; the "holdout" method was used to ensure independence between training and evaluation sets. The data set randomly changes between runs. An ensemble model (ESDM) was obtained for each species by averaging the best SDM outputs (highest Area Under the Curve—AUC—score), and the ensemble models were evaluated with the AUC score and the Cohen's Kappa coefficient (k). Following Smith and Santos (2020), we did not model species with $n < 32$ or that occupy >70% of the background region (i.e., entire distribution range for the genus).

We used the relative importance values of the variables provided by SSDM to evaluate the influence of each of them within all models. The importance is estimated with a randomization process, where SSDM calculates the correlation between a prediction using all variables and a prediction where the independent variable being tested is randomly removed; this is repeated for each variable. The calculation of the relative importance is made by subtracting this correlation from one, therefore higher values are the best variables for the model (Schmitt et al., 2017).

Diversity Metrics: Species Richness, Diversity, Endemism Phylogenetic Maps, and Environmental Variables Importance

Species richness, phylogenetic diversity, and phylogenetic endemism were calculated by superimposing the distribution maps of all species using the R package *phyloregion* (Daru et al., 2020b). In order to avoid overestimation of the diversity metrics, we created alpha hulls with the R package *rangeBuilder* (Davis Rabosky et al., 2016) and following (Paz et al., 2021). Briefly, we used occurrence data available for all species (54,392 georeferenced records) that had more than 10 locality points, a dynamic selection of alpha for each species, and an alpha that varied in steps of 1 (Meyer et al., 2017). We next generated a community matrix using the alpha hulls of all species with the function *polys2comm* in the R package *phyloregion* (Daru et al., 2020b).

We used the community matrix to calculate species richness by summing all species present in each cell, and also, with this matrix and the best Maximum Likelihood tree estimated with 20 nuclear and 2 mitochondrial loci for *Heliconius* (Kozak et al., 2015), we estimated phylogenetic diversity and phylogenetic endemism (Faith, 1992; Rosauer et al., 2009), with the functions phylogenetic diversity (*PD*) and *phylo_endemism* of the R package *phyloregion* (Daru et al., 2020b). To investigate whether these metrics are scale dependent, we performed the above analyses at three consecutive grain sizes (5, 10, and 20 km). We performed a linear regression model using phylogenetic diversity as response variable and species richness as predictor variable to investigate their relationship and plotted the residuals to highlight areas where these metrics are different.

We also used four machine learning algorithms to generate correlative models and then we created an ensemble prediction of each diversity metric to identify the environmental variables that best explain them (Paz et al., 2021). The algorithms used were: Random Forests (Liaw and Wiener, 2002), Neural Network (Venables and Ripley, 2002), Support Vector Machines (Karatzoglou et al., 2004), and GLM (McCullagh and Nelder, 1989). The models were built with the R package *caret* 6.0-86 (Kuhn, 2008), and we used the *varImp* function to compute the weighted average of the contribution of each variable.

Evaluating the Environmental Effect in the Hybridization on *Heliconius* Butterflies

We estimated the Schoener's niche equivalency test (D) and Warren's niche background test (I) between pairs of hybridizing species to determine if they share environmental niches. We used the R package *humboldt* (Brown and Carnaval, 2019) and we followed the concept of environmental niche *sensu* (Phillips et al., 2006; Soberón and Nakamura, 2009), where the niche consists of the subset of conditions currently occupied and where environmental conditions at the occurrence localities constitute samples from the realized niche. The niche overlap metric *Schoener's D* ranges between 0 and 1, meaning no overlap and complete overlap, respectively (Rödder and Engler, 2011). The environmental overlap was visualized with a principal component analysis (PCA). We tested the significance of this metric by comparing the realized niche overlap against a null distribution of 1,000 overlaps randomly generated from the reshuffled occurrence dataset and tested whether niche background and niche equivalency were different from those expected by chance at $\alpha = 0.05$ (Brown and Carnaval, 2019). This was done using the entire distribution of the entities under comparison (niche overlap test = NOT) and using only the area where they overlap (niche divergence test = NDT) (Brown and Carnaval, 2019).

RESULTS

Species Data, Species Distribution Modeling, and Environmental Variables Importance

We collected a total of 68,877 records for 46 species (n = 67,865), 37 cases interspecific hybrids (n = 164), and 34 cases of intraspecific hybrids (n = 848) in *Heliconius* (**Supplementary Tables 1, 2**).

From the species records we discarded 13,476 records as they could not be reliably georeferenced, thus leaving us with 54,392 records. For species modeling, these were further subject to pruning, which left a total of 13,671 records (**Supplementary Table 3**). There was considerable variation in the sampling effort across the phylogeny. For example, species of the erato and silvaniform clades are well-represented, whereas species from the aoede clade had the lowest number of records (**Supplementary Figure 1**). The variables retained and used to model species distributions and diversity metrics were: (i) minimum temperature of coldest month, (ii) altitude,

(iii) precipitation of coldest quarter, (iv) isothermality, and (v) precipitation seasonality (**Supplementary Figure 2**). The maximum absolute pairwise correlation between minimum temperature of coldest month and precipitation of coldest quarter was 0.436. The four algorithms we implemented were accurate in predicting the distribution of species, but their combination (ensemble) was the most accurate (**Supplementary Figure 3**). In total, we generated 44 species distribution models for *Heliconius* species. These are deposited in ZENODO.[1]

We found that environmental variables are better predictors of the distribution of *Heliconius* compared to topography. For instance, current temperature (isothermality) explains the distribution of 14 species (**Figures 1A,B**) and precipitation explains the distribution of 24 species (**Figures 1C,D**). In contrast, altitude explains the distribution of only five species (**Figure 1E**). No single variable was correlated with the entire distribution of the genus (**Figure 1F**), but we observed some general patterns. For example, isothermality explained the distribution of widely distributed species and *trans*-Andean species (i.e., west of the Andes; **Figures 1A,B**). Also, precipitation of the coldest quarter explains the distribution of species that occur in the biogeographic Choco + Costa Rica while precipitation seasonality explains the distribution of *cis*-Andean species (i.e., east of the Andes) + the Pacific of Ecuador (**Figures 1C,D**). Altitude explains the distribution of species restricted to the eastern foothills of the Andes and highland Andean species (**Figure 1E**). Interestingly, we did not find a single variable that was better correlated with the distribution of *H. charitonia* (**Supplementary Table 4**).

Diversity Metrics: Species Richness, Diversity, Endemism Phylogenetic Maps, and Environmental Variables Importance

We found that higher values of *Heliconius* species richness are concentrated in the foothills of the eastern Andes from Colombia to Ecuador, and into the Amazon basin mainly along the course of the Amazon River (**Figure 2A**). These results were consistent but more striking in the phylogenetic diversity maps (**Figure 2B**). Also, species richness has a strong and significant effect on phylogenetic diversity (adjusted R^2 0.9887, $p \leq$ 2e-16; **Supplementary Figure 4**). Interestingly, the residuals map showed values of phylogenetic diversity below those expected from species richness in the same regions, indicating that phylogenetic diversity, although high, is underestimated (blue grids; **Figure 2C**). In contrast, this metric was overestimated mainly in the Central Andes, the southern Amazon in Brazil, and the northern Chaco in Bolivia (red grids; **Figure 2C**). The highest values of phylogenetic endemism were concentrated in: (i) the Pacific coast of Costa Rica and Panama, (ii) the central foothills of the Eastern Cordillera in Colombia, and (iii) the biogeographic Choco of Colombia (**Figure 2D**). The pattern of these metrics was not scale dependent, and the results were highly congruent at 5, 10 (**Supplementary Figures 5, 6**, respectively), and 20 km (**Figure 2C**).

[1]https://doi.org/10.5281/zenodo.5149294

The ability of the machine learning models to predict species richness, phylogenetic diversity, and phylogenetic endemism varied between algorithms (**Supplementary Figure 7**). The best algorithms for all diversity metrics were the ensemble model followed by random forest, while the GLM algorithm had the lowest predictive accuracy in all metrics (**Supplementary Figure 7**). The best models predicted that altitude and isothermality were the most important variables for species richness and phylogenetic diversity (**Figures 3A,B**). In contrast, the most important variable for phylogenetic endemism was precipitation seasonality, followed by isothermality (**Figure 3C**). Finally, the residuals from the spatial regression between phylogenetic diversity (response variable) and species richness (predictor variable) were explained by isothermality (**Figure 3D**).

Evaluating the Environmental Effect on Hybridization in *Heliconius*

We found 18 pairs of hybridizing species in *Heliconius*. The results of the NOT and NDT tests based on *Schoener's D* revealed that the niches of three of these pairs (*H. melpomene/H. cydno*, *H. melpomene/H. hecale*, and *H. hecalesia/H. hortense*) are equivalent (**Figure 4** and **Table 1**) and overlap climatically ($D > 0.40$). In contrast, 12 of these pairs did not show evidence of niche equivalency. These included both pairs that have extensive geographic overlap (such as *H. ethilla* and *H. numata*) (**Supplementary Figure 8**) and pairs with a narrow overlap (such as *H. erato* and *H. himera*) (**Figure 5**). The remaining three pairs (*H. beskei/H. ethila*, *H. timareta/H. melpomene*, and *H. charitonia/H. peruvianus*) showed inconclusive results (**Figure 4** and **Table 1**). The results of these analyses were deposited in ZENODO (see text footnote 1).

DISCUSSION

We found that *Heliconius* butterflies display five general distribution patterns, namely: (i) wide distribution, (ii) *trans*-Andes, (iii) biogeographic Choco + Costa Rica, (iv) *cis*-Andes + Pacific of Ecuador, and (v) highland Andes. We also found that three variables (isothermality, precipitation and altitude) explain these patterns. Isothermality is a variable that quantifies how daily temperatures oscillate relative to the annual oscillations (O'Donnell and Ignizio, 2012), and its importance as one of the most explanatory variables of species distribution is not without precedent. For example, this variable explains the distribution of frugivorous bats (Chattopadhyay et al., 2019), mealybugs (Heya et al., 2020), Opiliones (Simó et al., 2014), and American monkeys (Vallejos-Garrido et al., 2017). Although all *Heliconius* species are strongly affected by isothermality, its effect seems to be stronger for widely distributed species and those with *trans*-Andean distribution. Interestingly, these species occur in regions with high and medium isothermality (>460%), that is, in regions that experience temperature changes throughout the day but keep a constant temperature throughout the year (O'Donnell and Ignizio, 2012). This suggests that these butterflies are particularly sensitive to long term changes in temperature, thus limiting their range to tropical areas.

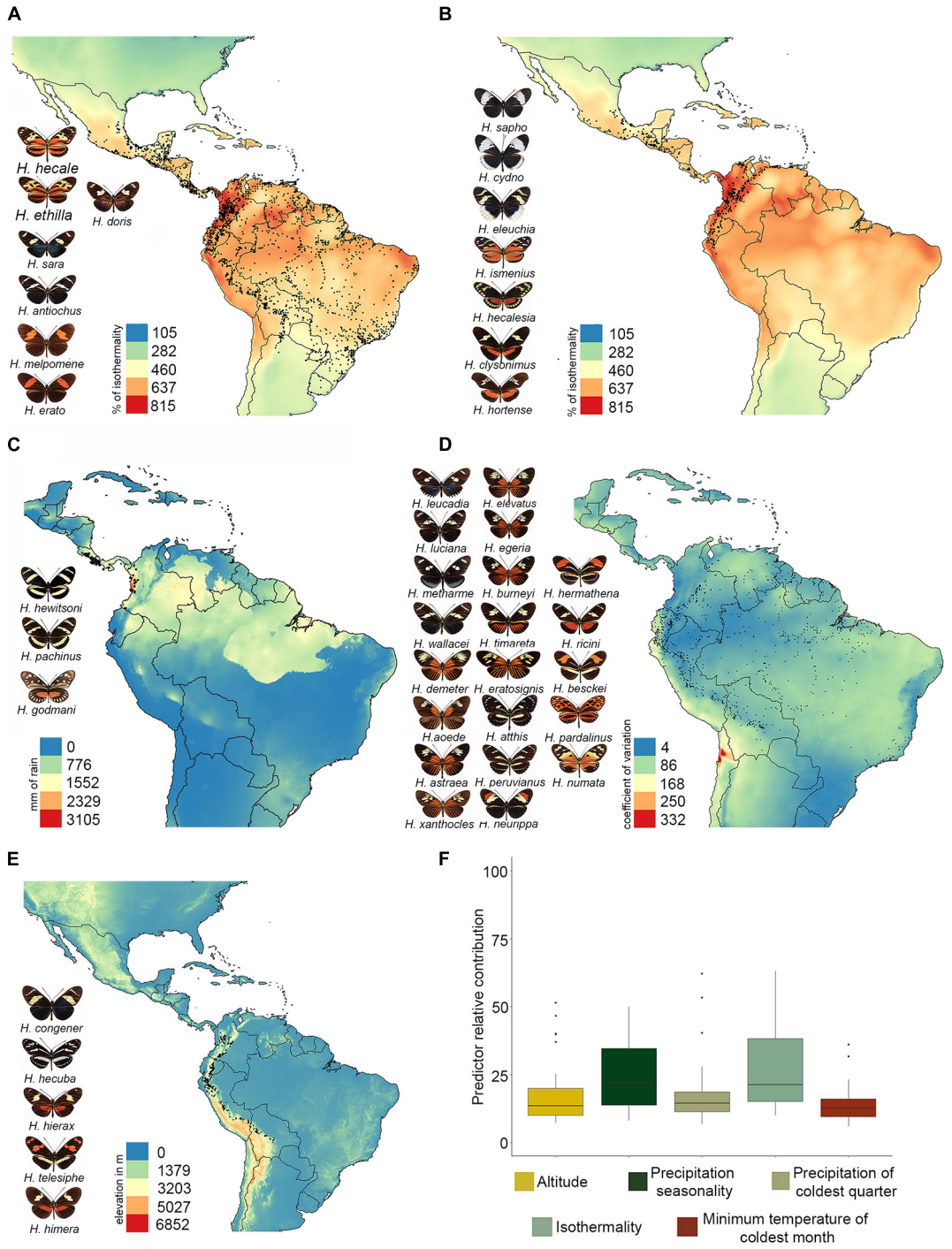


FIGURE 1 | Distribution patterns of *Heliconius* species based on environmental variables. **(A)** Species with wide distribution (explained by isothermality); **(B)** species with *trans*-Andean distribution (explained by isothermality); **(C)** species distributed in the biogeographic Choco + Costa Rica (explained by precipitation of coldest quarter); **(D)** species distributed in the *cis*-Andes + Pacific of Ecuador (explained by precipitation seasonality); **(E)** species distributed in highlands of the Andes (explained by altitude); **(F)** relative importance of environmental variables receiver operating characteristic (ROC) that are predictors of diversity in *Heliconius*. Color scale in panels **(A–E)** indicates the variable gradient. Distribution maps for each of the species can be found at: https://doi.org/10.5281/zenodo.5149294.

The distribution of species occurring in the biogeographic Choco of Colombia, Costa Rica, cis-Andes and the Pacific of Ecuador is also strongly limited by precipitation. Consistently, these regions have either rainforest, monsoon, or savanna climate, and they are the Neotropical regions with the highest precipitation [precipitation in the driest month (Pdry) > 60 mm]

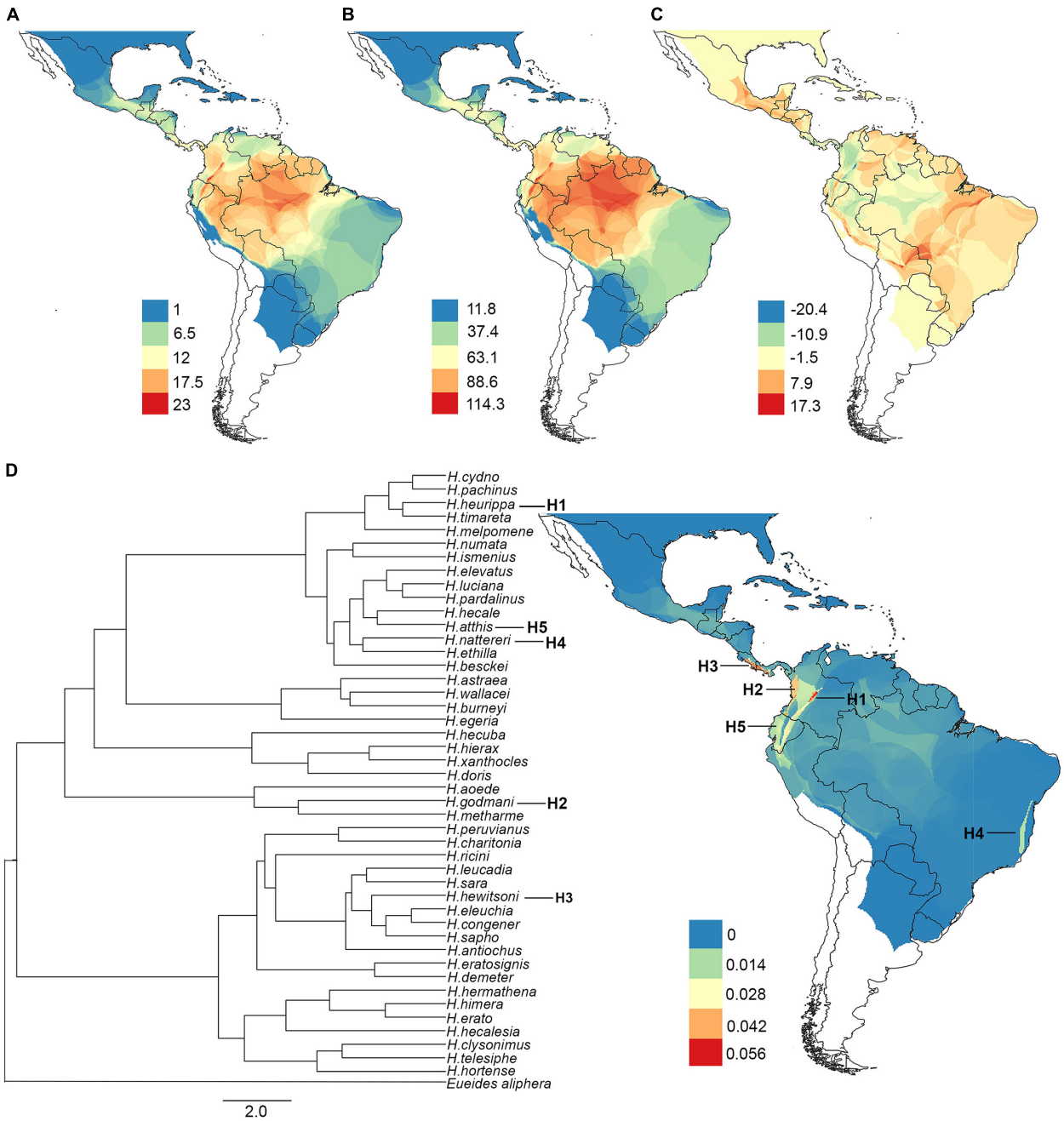


FIGURE 2 | Maps of diversity metrics. **(A)** Species richness, **(B)** phylogenetic diversity, **(C)** residuals of phylogenetic diversity regressed on species richness, and **(D)** phylogenetic endemism. Warm colors indicate higher values, while cold colors are indicative of lower values. The phylogeny shown in panel **(D)** was modified from Kozak et al. (2015), and branches that contribute the most to the phylogenetic endemism are labeled as H1–H5, both in the phylogeny and the map. All maps were plotted in grid cells of 20 km × 20 km.

(Beck et al., 2018). Previous studies have suggested that cloudiness and precipitation decrease flying bout duration in butterflies and, consequently, limit their dispersal (Cormont et al., 2011). Therefore, exceptionally high levels of precipitation in such regions may act as population traps, preventing butterflies from flying over longer distances and keeping them in a single region (Rosser et al., 2014). This finding agrees with previous studies in South America, where precipitation shapes the distribution of multiple vertebrates and invertebrates (Atauchi et al., 2017; Amundrud et al., 2018; Schivo et al., 2019; de Oliveira da Conceição et al., 2020).

In addition, altitude was the best predictor for the distribution of *Heliconius* species that can reach elevations up to 2,600 masl, which is considerably higher than the elevational range occupied

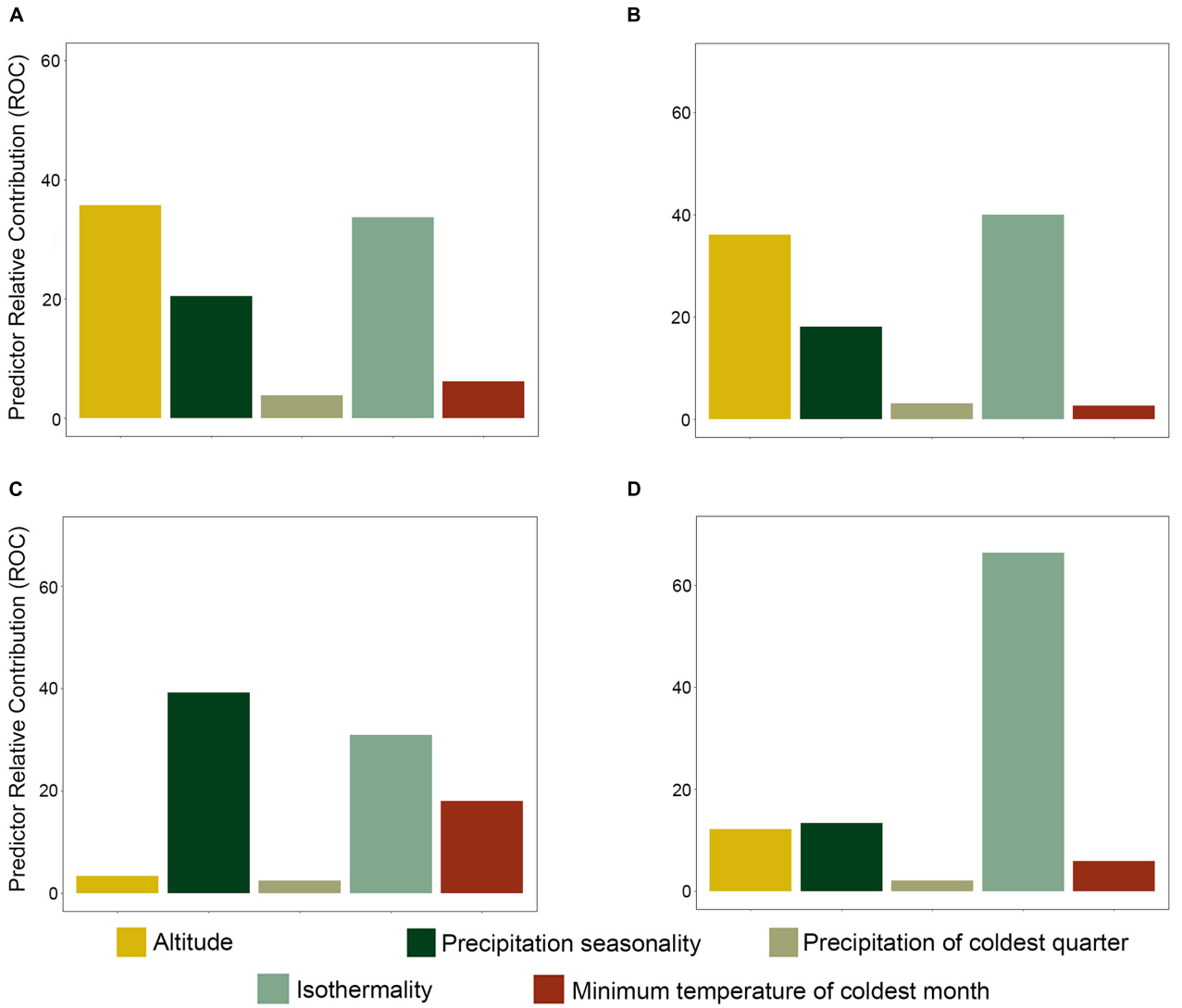


FIGURE 3 | Relative importance of predictors (environmental variables) of diversity. **(A)** Species richness, **(B)** phylogenetic diversity, **(C)** phylogenetic endemism, and **(D)** residuals of the phylogenetic diversity/species richness regression.

by other members of the genus (<2,200) (Rosser et al., 2012). Therefore, it is likely that these highland species have morphological or physiological modifications that allow them to expand their elevational range and occupy new niches. In fact, highland *Heliconius* are known to have rounder wings compared to lowland species, and this has been suggested to aid them flying dense cloud forest or compensate for the lower air pressure found at higher altitudes (Montejo-Kovacevich et al., 2019). Also, comparisons among different populations of *Heliconius* have revealed that highland populations are less tolerant to heat (Montejo-Kovacevich et al., 2020), which may limit their distribution range.

The foothills of the eastern Andes and the Amazon basin appeared as the regions with highest *Heliconius* species richness, which confirms the findings of a previous study done for the subfamily Heliconiine at a higher scale (50 km) (Rosser et al., 2012). Interestingly, both of these regions are known to present unusual concentrations of contact zones and hybrid zones (i.e., suture zones) (Dasmahapatra et al., 2010; Rosser et al., 2021), which may explain the richness they exhibit. Also, altitude, isothermality, and precipitation were the variables best correlated with this metric. This may be due to the elevational gradient found at the foothills of the eastern Andes, which offers multiple ecological niches thus favoring diversification rates (Rahbek and Graves, 2001; Jetz and Rahbek, 2002; Davies et al., 2007; Keppel et al., 2016). Additionally, there are several climate-based hypotheses that seek to explain broad-scale diversity patterns, and water and energy have emerged as crucial influencers of species richness (Silva-Flores et al., 2014). In particular, the water-energy dynamics hypothesis argues that species richness increases in places where liquid water and optimal energy conditions provide the greatest capacity for biotic dynamics

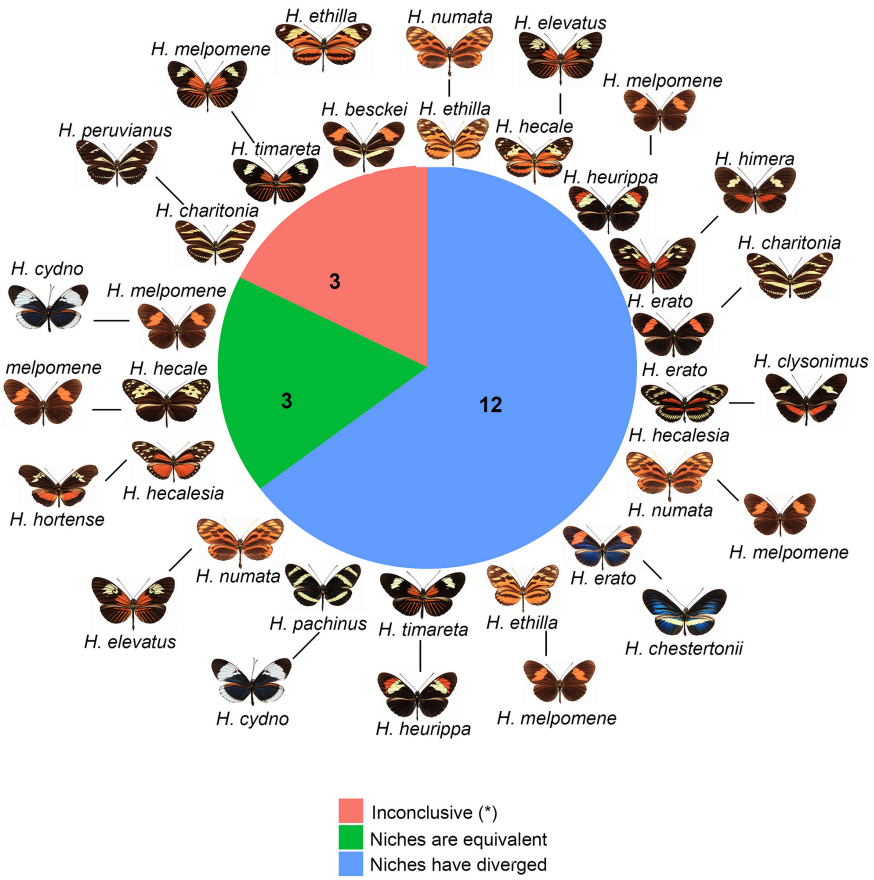


FIGURE 4 | Co-occurring and hybridizing species of *Heliconius*. Green: species pairs with equivalent environmental niches, blue: species pairs with divergent environmental niches, and salmon: species pairs with inconclusive results. Numbers indicate the pairs of species falling into each category.

(Svenning et al., 2008). The Amazon and foothills of the eastern Andes are regions with near constant hot-warm temperature throughout the year and have a permanent liquid water supply (Rosser et al., 2014; Vallejos-Garrido et al., 2017) thus ensuring an optimal water-energy dynamic. The latter translates into constant availability of plants for butterflies, including host plants for immature and pollen for adults, and continual interactions between individuals, which may be correlated with the high species richness we detected.

Similar to other studies, patterns of phylogenetic diversity were similar (although not identical) to richness (Davies Jonathan and Buckley, 2011; Fenker et al., 2014; Mendoza and Arita, 2014; Guedes et al., 2018). Interestingly, areas with highest species richness got low phylogenetic diversity (**Figure 2C**, blue grids), which may be a consequence of the recent increase in diversification rate in *Heliconius* (4.5 Ma) and the consequent co-occurrence of multiple young species in the Amazon and foothills of the eastern Andes (Rosser et al., 2012; Kozak et al., 2015). In agreement with this observation, previous research in both animals and plants have found high phylogenetic diversity in the eastern Andes of Colombia, Peru, and Ecuador (Fenker et al., 2014; Mendoza and Arita, 2014; Guedes et al., 2018; Arango et al., 2021; Velazco et al., 2021).

The highest phylogenetic endemism was found in the central eastern Andes of Colombia, and this result is possibly due to the restricted range of the species *Heliconius heurippa* (**Figure 2D**, area H1). However, we cannot rule out this result as an overestimation since the phylogenetic tree that we used (Kozak et al., 2015) considers this taxon as a separate species and not as part of *H. timareta* (as recently hypothesized). If *H. heurippa* had been included within *H. timareta*, which has a wider distribution range, it is likely this result on phylogenetic endemism does not hold. Additionally, the pacific region of Costa Rica, Panama and Colombia show intermediate values of phylogenetic endemism that resulted from the presence of species that have reduced geographic range and are either long-branch species (e.g., *Heliconius godmani*) or species for which no close relatives are known (e.g., *Heliconius hewitsoni*) (**Figure 2D**, area H2 and H3, respectively). These regions were previously described as highly endemic phylogenetically for plants (Sandel et al., 2020), terrestrial mammals (Rosauer and Jetz, 2014), birds and amphibians (Daru et al., 2020a). Interestingly, there were several species that, although are considered as geographically endemic within *Heliconius*, exhibited low values of phylogenetic endemism. However, it is important to acknowledge that phylogenetic endemism is a concept based on linages rather than

TABLE 1 | Niche overlap test (NOT) and niche divergence test (NDT) between hybridizing species.

Species 1	Species 2	Niche overlap test (NOT)				Niche divergence test (NDT)				Interpretation
		Equivalency test		Background test		Equivalency test		Background test		
		D	*P*-value for D	*P*-value for D (2-1)	*P*-value for D (1-2)	D	*P*-value for D	*P*-value for D (2-1)	*P*-value for D (1-2)	
H. melpomene	*H. cydno*	0.4404	0.7924	0.0036	0.0026	0.4904	0.8643	0.0039	0.0027	**Niches are equivalent**
H. pachinus	*H. cydno*	0.1136	0.2255	0.1264	0.0114	0.1575	0.0100	0.0357	0.0132	Niches have diverged
H. ethilla	*H. numata*	0.3740	0.0020	0.0034	0.0154	0.3733	0.0020	0.0034	0.0182	Niches have diverged
H. melpomene	*H. ethilla*	0.2989	0.0020	0.0154	0.0026	0.3087	0.0020	0.0156	0.0027	Niches have diverged
H. besckei	*H. ethilla*	0.4262	0.9581	0.0149	0.2500	0.6228	0.8104	0.2000	0.1429	Inconclusive
H. melpomene	*H. heurippa*	0.0396	0.0120	0.0536	0.6728	0.0551	0.0020	0.0508	0.8587	Niches have diverged
H. elevatus	*H. numata*	0.3833	0.0579	0.0035	0.0556	0.3939	0.0379	0.0035	0.0526	Niches have diverged
H. timareta	*H. melpomene*	0.0903	0.7425	0.5182	0.0669	0.1106	0.3433	0.3485	0.0669	Inconclusive
H. melpomene	*H. hecale*	0.5138	0.1936	0.0033	0.0025	0.5243	0.2016	0.0038	0.0026	**Niches are equivalent**
H. hecale	*H. elevatus*	0.2655	0.0020	0.0385	0.0110	0.2731	0.0020	0.0333	0.0042	Niches have diverged
H. erato	*H. chestertonii*	0.0405	0.0019	0.6593	0.7804	0.0571	0.0001	0.6087	0.6926	Niches have diverged
H. erato	*H. charitonia*	0.0170	0.0020	0.0050	0.0021	0.2426	0.0100	0.0050	0.0022	Niches have diverged
H. charitonia	*H. peruvianus*	0.0172	0.7226	0.1429	0.8894	0.2404	0.9541	0.1250	0.0135	Inconclusive
H. hecalesia	*H. hortense*	0.4213	1.0000	0.0161	0.0130	0.4799	0.9940	0.0182	0.0323	**Niches are equivalent**
H. hecalesia	*H. clysonimus*	0.2464	0.0060	0.0032	0.0118	0.2359	0.0020	0.0032	0.0132	Niches have diverged
H. melpomene	*H. numata*	0.4600	0.0200	0.0036	0.0026	0.4843	0.0180	0.0034	0.0026	Niches have diverged
H. timareta	*H. heurippa*	0.2432	0.0998	0.0448	0.1111	0.1821	0.0020	0.0185	0.3103	Niches have diverged
H. erato	*H. himera*	0.0340	0.002	0.02632	0.0322	0.1100	0.0019	0.0333	0.0129	Niches have diverged

Bolded text means that the occupied niches by these two species are not statistically different.

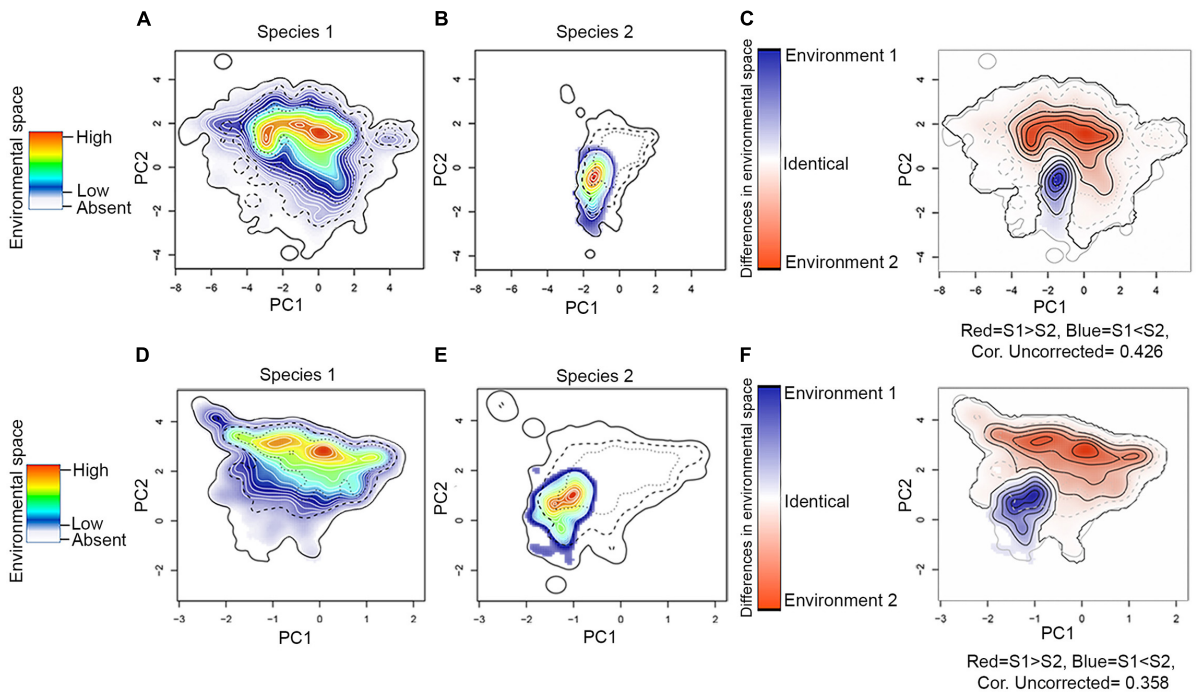


FIGURE 5 | Assessment of niche similarity. As an example, we show the Niche Overlap Test (NOT–top row) and Niche Divergence Test (NDT–bottom row) between *H. erato* (species 1) and *H. himera* (species 2), but other comparisons can be found in ZENODO (https://doi.org/10.5281/zenodo.5149294). **(A,D)** Environmental space of species 1; **(B,E)** environmental space of species 2; and **(C,F)** difference in the environmental space (E-space) of two species and Niche E-space Correlation Index (NECI). When NECI was higher than 0.5, we corrected species occupied niches by the frequency of E-space in accessible environments. Significance of NOT and NDT can be found in **Table 1**. Equivalency statistic and niche background statistic for each NOT and NDT can be found in https://doi.org/10.5281/zenodo.5149294.

species, and thus, if an endemic species has a narrow range but it is closely related to a widespread species, its phylogenetic endemism will not necessarily be low (Rosauer et al., 2009). An example of this is *Heliconius nattereri*, an endemic species from Brazil's Atlantic Forest that, despite having a narrow distribution, is sister to the widely distributed *Heliconius ethilla* (**Figure 2D**, area H4). Similarly, *Heliconius atthis* is restricted to the Ecuadorian and Peruvian Pacific, but it is sister to the widely distributed *Heliconius hecale* (**Figure 2D**, area H5). In our study we found that high precipitation and near constant hot-warm temperature throughout the year are strongly correlated with phylogenetic endemism, which agrees with studies that point a role for temperature in promoting endemism by reducing extinction rates and increasing population sizes in small areas (Jetz et al., 2004; Rosauer and Jetz, 2014; Varzinczak et al., 2020).

Our environmental niche analysis showed that hybridizing species do not necessarily share the same climatic space despite some of them having largely overlapping geographic distributions. This is the case of *H. ethilla* and *H. numata*, which frequently co-occur throughout their distribution, but there are some regions with an extreme climate, such as the Pacific coast of Colombia (a humid jungle) and the Colombian Magdalena valley (which has a marked precipitation gradient, being humid in the north and dry in the south), where *H. ethilla* but not *H. numata* occur (**Supplementary Figure 9**). This suggests that the former species has a broader climatic tolerance. We also detected differences in the environmental niche between pairs of hybridizing species that rarely overlap geographically, but when they do, they hybridize. For example, *H. erato* and *H. himera* occupy contrasting environmental niches in Ecuador (Jiggins et al., 1997), where *H. himera* lives in dry forests while *H. erato* inhabits wet forests of the Andes (**Figure 5**). Similarly, the hybridizing *H. erato* (*H. e. venus*) and *H. chestertonii* meet in an environmental transition zone between wet and dry forest in the Colombian Andes (Muñoz et al., 2010; **Supplementary Figure 8**).

In summary, we confirmed that, at large scales, the distribution of *Heliconius*, its richness, diversity, and phylogenetic endemism are mainly shaped by a combination of high annual energy (i.e., hot-warm temperature), constant water supply, and an extraordinary topographic complexity. However, species distributions are thought to result from dynamics occurring at multiple spatial scales. Therefore, including microclimate variables and ecological interactions would provide an in-depth understanding of the multiscale drivers of distribution, niche range and phylogenetic processes (Montejo-Kovacevich et al., 2020; Paz and Guarnizo, 2020). Our study confirms the richness and diversity of areas already identified in other taxa, thus strengthening the importance for their conservation as strategic hotspots of biodiversity.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://doi.org/10.5281/zenodo.5149294.

AUTHOR CONTRIBUTIONS

NR-M and CS conceived the study. NR-M and FS-R analyzed the data. NR-M and CG-Q curated the databases. NR-M, CS, CP-D, FS-R, and CG-Q collected the specimens in regions poorly sampled in Colombia. NR-M, CS, and CP-D drafted the manuscript. All authors read and approved the final version.

FUNDING

NR-M was funded by the Ministerio de Ciencia Tecnología e Innovación (MinCiencias) Grant 727-2016, and the Universidad del Rosario Doctoral Student Fellowships. CS and CP-D were funded by the Universidad del Rosario Big-Grant IV-FGD005/IV-FGI006.

ACKNOWLEDGMENTS

We thank Chris Jiggins for providing us with photos of some specimens. We also thank "Autoridad Nacional de Licencias Ambientales—ANLA" for granting Universidad del Rosario the collecting permit 530. Also, the High-Performance Computing Service of Universidad del Rosario provided free computing time.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2021.750703/full#supplementary-material

REFERENCES

Amundrud, S. L., Videla, M., and Srivastava, D. S. (2018). Dispersal barriers and climate determine the geographic distribution of the helicopter damselfly *Mecistogaster modesta*. *Freshw. Biol.* 63, 214–223. doi: 10.1111/fwb.13054

Andújar, C., Arribas, P., Ruiz, C., Serrano, J., and Gómez-Zurita, J. (2014). Integration of conflict into integrative taxonomy: fitting hybridization in species delimitation of Mesocarabus (Coleoptera: Carabidae). *Mol. Ecol.* 23, 4344–4361. doi: 10.1111/mec.12793

Arango, A., Villalobos, F., Prieto-Torres, D. A., and Guevara, R. (2021). The phylogenetic diversity and structure of the seasonally dry forests in the Neotropics. *J. Biogeogr.* 48, 176–186. doi: 10.1111/jbi.13991

Arteaga, M. C., Mccormack, J. E., Eguiarte, L. E., and Medell, R. A. (2011). Genetic admixture in multidimensional enviromental space : asymetrical niche similarity promotes gene flow in Armadillos (*Dasypus Novemcinctus*). *Evolution* 65, 2470–2480. doi: 10.1111/j.1558-5646.2011.01329.x

Assis, J. (2020). *R Pipelines to Reduce the Spatial Autocorrelation in Species Distribution Models. the MarineDataScientist*. Available online at: https://github.com/jorgeassis/spatialAutocorrelation (accessed April 20, 2021).

Atauchi, P. J., Peterson, T., and Flanagan, J. (2017). Species distribution models for *Peruvian plantcutter* improve with consideration of biotic interactions. *J. Avian Biol.* 49:jav-01617.

Barbet-Massin, M., Jiguet, F., Albert, C. H., and Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and

how many? *Methods Ecol. Evol.* 3, 327–338. doi: 10.1111/j.2041-210X.2011.00172.x

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F. (2018). Present and future köppen-geiger climate classification maps at 1-km resolution. *Sci. Data* 5, 1–12. doi: 10.1038/sdata.2018.214

Brown, J. L., and Carnaval, A. C. (2019). A tale of two niches: methods, concepts, and evolution. *Front. Biogeogr.* 11:e44158. doi: 10.21425/F5FBG44158

Brown, J. L., Paz, A., Reginato, M., Renata, C. A., Assis, C., Lyra, M., et al. (2020). Seeing the forest through many trees: multi-taxon patterns of phylogenetic diversity in the Atlantic forest hotspot. *Divers. Distrib.* 26, 1160–1176. doi: 10.1111/ddi.13116

Chattopadhyay, B., Garg, K. M., Ray, R., and Rheindt, F. E. (2019). Fluctuating fortunes: genomes and habitat reconstructions reveal global climate-mediated changes in bats' genetic diversity. *Proc. R. Soc. B Biol. Sci.* 286, 1–10. doi: 10.1098/rspb.2019.0304

Cormont, A., Malinowska, A. H., Kostenko, O., Radchuk, V., Hemerik, L., WallisDeVries, M. F., et al. (2011). Effect of local weather on butterfly flight behaviour, movement, and colonization: significance for dispersal under climate change. *Biodivers. Conserv.* 20, 483–503. doi: 10.1007/s10531-010-9960-4

Daru, B. H., Karunarathne, P., and Schliep, K. (2020b). phyloregion: R package for biogeographical regionalization and macroecology. *Methods Ecol. Evol.* 11, 1483–1491. doi: 10.1111/2041-210X.13478

Daru, B. H., Farooq, H., Antonelli, A., and Faurby, S. (2020a). Endemism patterns are scale dependent. *Nat. Commun.* 11, 1–11. doi: 10.1038/s41467-020-15921-6

Dasmahapatra, K. K., Lamas, G., Simpson, F., and Mallet, J. (2010). The anatomy of a "suture zone" in Amazonian butterflies: a coalescent-based test for vicariant geographic divergence and speciation. *Mol. Ecol.* 19, 4283–4301. doi: 10.1111/j.1365-294X.2010.04802.x

Davies, R. G., Orme, C. D. L., Storch, D., Olson, V. A., Thomas, G. H., Ross, S. G., et al. (2007). Topography, energy and the global distribution of bird species richness. *Proc. R. Soc. B Biol. Sci.* 274, 1189–1197. doi: 10.1098/rspb.2006.0061

Davies Jonathan, T., and Buckley, L. B. (2011). Phylogenetic diversity as a window into the evolutionary and biogeographic histories of present-day richness gradients for mammals. *Philos. Trans. R. Soc. B Biol. Sci.* 366, 2414–2425. doi: 10.1098/rstb.2011.0058

Davis Rabosky, A. R., Cox, C. L., Rabosky, D. L., Title, P. O., Holmes, I. A., Feldman, A., et al. (2016). Coral snakes predict the evolution of mimicry across New World snakes. *Nat. Commun.* 7, 1–9. doi: 10.1038/ncomms11484

de Oliveira da Conceição, E., Mantovano, T., de Campos, R., Rangel, T. F., Martens, K., Bailly, D., et al. (2020). Mapping the observed and modelled intracontinental distribution of non-marine ostracods from South America. *Hydrobiologia* 847, 1663–1687. doi: 10.1007/s10750-019-04136-6

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36, 27–46. doi: 10.1111/j.1600-0587.2012.07348.x

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* 61, 1–10. doi: 10.1016/0006-3207(92)91201-3

Fenker, J., Tedeschi, L. G., Pyron, R. A., Nogueira, C., and de, C. (2014). Phylogenetic diversity, habitat loss and conservation in South American pitvipers (Crotalinae: Bothrops and Bothrocophias). *Divers. Distrib.* 20, 1108–1119. doi: 10.1111/ddi.12217

Friedman, J. H., Hastie, T., and Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting. *Ann. Stat.* 28, 337–407.

Gotelli, N. J., Anderson, M. J., Arita, H. T., Chao, A., Colwell, R. K., Connolly, S. R., et al. (2009). Patterns and causes of species richness: a general simulation model for macroecology. *Ecol. Lett.* 12, 873–886. doi: 10.1111/j.1461-0248.2009.01353.x

Guedes, T. B., Sawaya, R. J., Zizka, A., Laffan, S., Faurby, S., Pyron, R. A., et al. (2018). Patterns, biases and prospects in the distribution and diversity of Neotropical snakes. *Glob. Ecol. Biogeogr.* 27, 14–21. doi: 10.1111/geb.12679

Hastie, T. J., and Tibshirani, R. J. (1990). *Generalized Additive Models*. London: Chapman and Hall.

Hawkins, B. A., Field, R., Cornell, H. V., Currie, D. J., Guégan, J. F., Kaufman, D. M., et al. (2003). Energy, water, and broad-scale geographic patterns of species richness. *Ecology* 84, 3105–3117. doi: 10.1890/03-8006

Heiberger, R. M. (2020). *HH: Statistical Analysis and Data Display: Heiberger and Holland. R Package Version 3*. Available online at: https://CRAN.R-project.org/package=HH (accessed February 15, 2021).

Heya, H. M., Khamis, F. M., Onyambu, G. K., Akutse, K. S., Mohamed, S. A., Kimathi, E. K., et al. (2020). Characterization and risk assessment of the invasive papaya mealybug, *Paracoccus marginatus*, in Kenya under changing climate. *J. Appl. Entomol.* 144, 1–17. doi: 10.1111/jen.12748

Jarvis, A., Reuter, H. I., Nelson, A., and Guevara, E. (2008). *Hole-Filled SRTM for the Globe Version 4, the CGIAR-CSI SRTM 90m Database*. Available online at: http://srtm.csi.cgiar.org (accessed January 30, 2021).

Jetz, W., and Rahbek, C. (2002). Geographic range size and determinants of avian species richness. *Science* 297, 1548–1551. doi: 10.1126/science.1072779

Jetz, W., Rahbek, C., and Colwell, R. K. (2004). The coincidence of rarity and richness and the potential signature of history in centres of endemism. *Ecol. Lett.* 7, 1180–1191. doi: 10.1111/j.1461-0248.2004.00678.x

Jiggins, C. (2017). *The Ecology and Evolution of Heliconius Butterflies*. Oxford: Oxford University Press.

Jiggins, C., Mcmillan, W. O., and Mallet, J. (1997). Host plant adaptation has not played a role in the recent speciation of *Heliconius himera* and *Heliconius erato*. *Ecol. Entomol.* 22, 361–365. doi: 10.1046/j.1365-2311.1997.00067.x

Karatzoglou, A., Hornik, K., Smola, A., and Zeileis, A. (2004). kernlab - An S4 package for kernel methods in R. *J. Stat. Softw.* 11, 1–20. doi: 10.18637/jss.v011.i09

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., et al. (2017). Climatologies at high resolution for the earth's land surface areas. *Sci. Data* 4:170122. doi: 10.1038/sdata.2017.122

Keppel, G., Gillespie, T. W., Ormerod, P., and Fricker, G. A. (2016). Habitat diversity predicts orchid diversity in the tropical south-west Pacific. *J. Biogeogr.* 43, 2332–2342. doi: 10.1111/jbi.12805

Kozak, K. M., Wahlberg, N., Neild, A. F., Dasmahapatra, K. K., Mallet, J., and Jiggins, C. D. (2015). Multilocus species trees show the recent adaptive radiation of the mimetic *Heliconius* butterflies. *Syst .Biol.* 64, 505–524.

Kreft, H., and Jetz, W. (2007). Global patterns and determinants of vascular plant diversity. *Proc. Natl. Acad. Sci. U.S.A.* 104, 5925–5930. doi: 10.1073/pnas.0608361104

Kronforst, M. R., Hansen, M. E., Crawford, N. G., Gallant, J. R., Zhang, W., Kulathinal, R. J., et al. (2013). Hybridization reveals the evolving genomic architecture of speciation. *Cell Rep.* 14, 666–677. doi: 10.1016/j.celrep.2013.09.042

Kubota, Y., Shiono, T., and Kusumoto, B. (2015). Role of climate and geohistorical factors in driving plant richness patterns and endemicity on the east Asian continental islands. *Ecography* 38, 639–648. doi: 10.1111/ecog.00981

Kuhn, M. (2008). Building predictive models in R using the caret package. *J. Stat. Softw.* 28, 1–26. doi: 10.18637/jss.v028.i05

Lake, T. A., Briscoe Runquist, R. D., and Moeller, D. A. (2020). Predicting range expansion of invasive species: pitfalls and best practices for obtaining biologically realistic projections. *Divers. Distrib.* 26, 1767–1779. doi: 10.1111/ddi.13161

Lamas, G., and Jiggins, C. D. (2017). "Taxonomic list," in *The Ecology and Evolution of Heliconius Butterflies*, ed. C. D. Jiggins (Oxford: Oxford University Press), 214–244.

Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. *R News* 2, 18–22.

López-Aguirre, C., Hand, S. J., Laffan, S. W., and Archer, M. (2019). Zoogeographical regions and geospatial patterns of phylogenetic diversity and endemism of New World bats. *Ecography* 42, 1188–1199. doi: 10.1111/ecog.04194

Mallet, J., Barton, N., Lamas, G., Santisteban, J., Muedas, M., and Eeley, H. (1990). Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *Heliconius* hybrid zones. *Genetics* 124, 921–936.

Mallet, J., Beltrán, M., Neukirchen, W., and Linares, M. (2007). Natural hybridization in heliconiine butterflies: the species boundary as a continuum. *BMC Evol. Biol.* 7:28. doi: 10.1186/1471-2148-7-28

Mallet, J., McMillan, W. O., and Jiggins, C. D. (1998). "Mimicry and warning color at the boundary between races and species," in *Endless Forms: Species and Speciation*, eds D. Howard and S. Berlöcher (New York, NY: Oxford University Press), 390–403.

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., et al. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res.* 23, 1817–1828. doi: 10.1101/gr.159426.113

McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman & Hall.

Mendoza, A. M., and Arita, H. T. (2014). Priority setting by sites and by species using rarity, richness and phylogenetic diversity: the case of neotropical glassfrogs (Anura: Centrolenidae). *Biodivers. Conserv.* 23, 909–926. doi: 10.1007/s10531-014-0642-5

Meyer, L., Diniz-Filho, J. A. F., and Lohmann, L. G. (2017). A comparison of hull methods for estimating species ranges and richness maps. *Plant Ecol. Divers.* 10, 389–401. doi: 10.1080/17550874.2018.1425505

Montejo-Kovacevich, G., Martin, S. H., Meier, J. I., Bacquet, C. N., Monllor, M., Jiggins, C. D., et al. (2020). Microclimate buffering and thermal tolerance across elevations in a tropical butterfly. *J. Exp. Biol.* 223:jeb220426. doi: 10.1242/jeb.220426

Montejo-Kovacevich, G., Smith, J. E., Meier, J. I., Bacquet, C. N., Whiltshire-Romero, E., Nadeau, N. J., et al. (2019). Altitude and life-history shape the evolution of *Heliconius* wings. *Evolution* 73, 2436–2450. doi: 10.1111/evo.13865

Mullen, S. P., Savage, W. K., Wahlberg, N., and Willmott, K. R. (2011). Rapid diversification and not clade age explains high diversity in neotropical Adelpha butterflies. *Proc. R. Soc. B Biol. Sci.* 278, 1777–1785. doi: 10.1098/rspb.2010.2140

Muñoz, A. G., Salazar, C., Castaño, J., Jiggins, C. D., and Linares, M. (2010). Multiple sources of reproductive isolation in a bimodal butterfly hybrid zone. *J. Evol. Biol.* 23, 1312–1320. doi: 10.1111/j.1420-9101.2010.02001.x

O'Donnell, M. S., and Ignizio, D. A. (2012). Bioclimatic predictors for supporting ecological applications in the conterminous United States. *U.S. Geol. Surv. Data Ser.* 691, 1–17.

Ortego, J., Gugger, P. F., Riordan, E. C., and Sork, V. L. (2014). Influence of climatic niche suitability and geographical overlap on hybridization patterns among southern Californian oaks. *J. Biogeogr.* 41, 1895–1908. doi: 10.1111/jbi.12334

Paz, A., Brown, J. L., Cordeiro, C. L. O., Aguirre-Santoro, J., Assis, C., Amaro, R. C., et al. (2021). Environmental correlates of taxonomic and phylogenetic diversity in the Atlantic forest. *J. Biogeogr.* 48, 1377–1391. doi: 10.1111/jbi.14083

Paz, A., and Guarnizo, C. E. (2020). Environmental ranges estimated from species distribution models are not good predictors of lizard and frog physiological tolerances. *Evol. Ecol.* 34, 89–99. doi: 10.1007/s10682-019-10022-3

Pearson, D. L., and Carroll, S. S. (2001). Predicting patterns of tiger beetle (Coleoptera : Cicindelidae) species richness in Northwestern South America. *Stud. Neotrop. Fauna Environ.* 36, 125–136.

Phillips, S. J., Anderson, R. P., and Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190, 231–259. doi: 10.1016/j.ecolmodel.2005.03.026

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., et al. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol. Appl.* 19, 181–197. doi: 10.1890/07-2153.1

Qian, H. (2010). Environment-richness relationships for mammals, birds, reptiles, and amphibians at global and regional scales. *Ecol. Res.* 25, 629–637. doi: 10.1007/s11284-010-0695-1

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team.

Rahbek, C., and Graves, G. R. (2001). Multiscale assessment of patterns of avian species richness. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4534–4539. doi: 10.1073/pnas.071034898

Rödder, D., and Engler, J. O. (2011). Quantitative metrics of overlaps in Grinnellian niches: advances and possible drawbacks. *Glob. Ecol. Biogeogr.* 20, 915–927. doi: 10.1111/j.1466-8238.2011.00659.x

Rosauer, D., Laffan, S. W., Crisp, M. D., Donnellan, S. C., and Cook, L. G. (2009). Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. *Mol. Ecol.* 18, 4061–4072. doi: 10.1111/j.1365-294X.2009.04311.x

Rosauer, D. F., and Jetz, W. (2014). Phylogenetic endemism in terrestrial mammals. *Glob. Ecol. Biogeogr.* 24, 168–179. doi: 10.1111/geb.12237

Rosser, N., Dasmahapatra, K. K., and Mallet, J. (2014). Stable *Heliconius* butterfly hybrid zones are correlated with a local rainfall peak at the edge of the Amazon basin. *Evolution* 68, 3470–3484.

Rosser, N., Phillimore, A., Huertas, B., Willmott, K., and Mallet, J. (2012). Testing historical explanations for gradients in species richness in heliconiine butterflies of tropical America. *Biol. J. Linn. Soc.* 105, 479–497. doi: 10.1111/j.1095-8312.2011.01814.x

Rosser, N., Shirai, L. T., Dasmahapatra, K. K., Mallet, J., and Freitas, A. V. L. (2021). The Amazon river is a suture zone for a polyphyletic group of co-mimetic heliconiine butterflies. *Ecography* 44, 177–187. doi: 10.1111/ecog.05282

Sandel, B., Weigelt, P., Kreft, H., Keppel, G., van der Sande, M. T., Levin, S., et al. (2020). Current climate, isolation and history drive global patterns of tree phylogenetic endemism. *Glob. Ecol. Biogeogr.* 29, 4–15. doi: 10.1111/geb.13001

Schivo, F., Bauni, V., Krug, P., and Quintana, R. D. (2019). Distribution and richness of amphibians under different climate change scenarios in a subtropical region of South America. *Appl. Geogr.* 103, 70–89. doi: 10.1016/j.apgeog.2019.01.003

Schmitt, S., Pouteau, R., Justeau, D., de Boissieu, F., and Birnbaum, P. (2017). ssdm: an r package to predict distribution of species richness and composition based on stacked species distribution models. *Methods Ecol. Evol.* 8, 1795–1803. doi: 10.1111/2041-210X.12841

Silva-Flores, R., Pérez-Verdín, G., and Wehenkel, C. (2014). Patterns of tree species diversity in relation to climatic factors on the Sierra Madre Occidental, Mexico. *PLoS One* 9:e105034. doi: 10.1371/journal.pone.0105034

Simó, M., Guerrero, J. C., Giuliani, L., Castellano, I., and Acosta, L. E. (2014). A predictive modeling approach to test distributional uniformity of Uruguayan harvestmen (Arachnida: Opiliones). *Zool. Stud.* 53, 1–13. doi: 10.1186/s40555-014-0050-2

Smith, A. B., and Santos, M. J. (2020). Testing the ability of species distribution models to infer variable importance. *Ecography* 43, 1801–1813. doi: 10.1111/ecog.05317

Soberón, J., and Nakamura, M. (2009). Niches and distributional areas: concepts, methods, and assumptions. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19644–19650. doi: 10.1073/pnas.0901637106

Svenning, J. C., Borchsenius, F., Bjorholm, S., and Balslev, H. (2008). High tropical net diversification drives the New World latitudinal gradient in palm (Arecaceae) species richness. *J. Biogeogr.* 35, 394–406. doi: 10.1111/j.1365-2699.2007.01841.x

Swenson, N. G., Fair, J. M., and Heikoop, J. (2008). Water stress and hybridization between *Quercus gambelii* and *Quercus grisea*. *West. North Am. Nat.* 68, 498–507.

Vallejos-Garrido, P., Rivera, R., Inostroza-Michae, O., Rodríguez-Serrano, E., and Hernández, C. E. (2017). Historical dynamics and current environmental effects explain the spatial distribution of species richness patterns of New World monkeys. *PeerJ* 5, 2–27. doi: 10.7717/peerj.3850

Varzinczak, L. H., Zanata, T. B., Moura, M. O., and Passos, F. C. (2020). Geographical patterns and current and short-term historical correlates of phylogenetic diversity and endemism for New World primates. *J. Biogeogr.* 47, 890–902. doi: 10.1111/jbi.13767

Vasconcelos, T. S., da Silva, F. R., dos Santos, T. G., Prado, V. H. M., and Provete, D. B. (2019). *Biogeographic Patterns of South American Anurans*. Berlin: Springer, doi: 10.1007/978-3-030-26296-9

Velazco, S. J. E., Svenning, J. C., Ribeiro, B. R., and Laureto, L. M. O. (2021). On opportunities and threats to conserve the phylogenetic diversity of Neotropical palms. *Divers. Distrib.* 27, 512–523. doi: 10.1111/ddi.13215

Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics With S*. Berlin: Springer.

Wieczorek, J., Guo, Q., and Hijmans, R. J. (2004). The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *Int. J. Geogr. Inf. Sci.* 18, 745–767. doi: 10.1080/13658810412331280211

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 15 November 2021
doi: 10.3389/fevo.2021.778686

Elevational Diversity Patterns of Green Lacewings (Neuroptera: Chrysopidae) Uncovered With DNA Barcoding in a Biodiversity Hotspot of Southwest China

*Yan Lai[1], Yunhui Liu[2] and Xingyue Liu[1]**

[1] Department of Entomology, College of Plant Protection, China Agricultural University, Beijing, China, [2] Beijing Key Laboratory of Biodiversity and Organic Farming, College of Resources and Environmental Sciences, China Agricultural University, Beijing, China

OPEN ACCESS

Edited by:
Shu-Jun Wei,
Beijing Academy of Agricultural and Forestry Sciences, China

Reviewed by:
Xiaolei Huang,
Fujian Agriculture and Forestry University, China
Xubo Wang,
Southwest Forestry University, China

***Correspondence:**
Xingyue Liu
xingyue_liu@yahoo.com

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *17 September 2021*
Accepted: *14 October 2021*
Published: *15 November 2021*

Citation:
Lai Y, Liu Y and Liu X (2021) Elevational Diversity Patterns of Green Lacewings (Neuroptera: Chrysopidae) Uncovered With DNA Barcoding in a Biodiversity Hotspot of Southwest China. Front. Ecol. Evol. 9:778686. doi: 10.3389/fevo.2021.778686

Elevational diversity patterns can reflect the responses of biodiversity to climate change spatially. We investigate the species diversity patterns of green lacewings (an important predatory group of insects) along the gradient of elevation from the Shaluli Mountains (Mts. Shaluli), which belong to the Hengduan Mountains in southwestern China, one of the important hotspots of global biodiversity. We combined multiple approaches, including Automatic Barcode Gap Discovery (ABGD), Assemble Species by Automatic Partitioning analysis (ASAP), General Mixed Yule Coalescent (GMYC), Poisson tree processes (bPTP), multi-rate Poisson tree processes (mPTP), to delimit the green lacewings species based on the standard barcoding region of cytochrome c oxidase subunit I (COI). The α-diversity and β-diversity patterns of green lacewings from the Mts. Shaluli along the gradient of elevation were analyzed, with further exploration on how the temperature effect elevational-diversity pattern on broad-scale (county scale) elevational gradients. The DNA barcoding reference library consisted of 40 green lacewing species from the Mts. Shaluli. The α-diversity of green lacewings decreased with the increasing elevation. The temperature was found to have a significant effect on the abundance and Shannon-Wiener diversity index but not on the species richness. Nestedness replaced turnover as the main component of Sørensen's dissimilarity with the increasing elevation, and greater nestedness occurred at low temperature areas. The combination of a reliable DNA barcoding database could improve the accuracy and efficiency to investigate the species diversity patterns of green lacewings. Temperature, resource, and resultant interspecific competitions may have important roles in explaining the species diversity patterns of green lacewings from the Mts. Shaluli. Priority of conservation should be given to the species at low elevation, middle elevation, and relatively high temperature regions under the background of global climate warming.

Keywords: DNA barcoding, α-diversity pattern, β-diversity pattern, elevational gradient, natural enemy

INTRODUCTION

Elucidating the factors that drive global biodiversity is the fundamental goal in ecology and conservation (McCain, 2007) considering the complexity of biotic (i.e., interspecific interactions, food availability, and thermal tolerance) and abiotic (i.e., temperature, precipitation, atmospheric pressure, and wind velocity) factors in nature's ecosystem (Sundqvist et al., 2013). Temperature is commonly considered to be a main causal factor of species diversity dynamics (McCain, 2007; Deutsch et al., 2008; Diamond et al., 2016; Birkett et al., 2018; Barbarossa et al., 2021). The elevational gradients, which may reflect their responses to climate change spatially, also play an important role in providing powerful information on how the ranges of biodiversity are restricted by the environmental conditions (Grinnell, 1914; Sundqvist et al., 2013; Supriya et al., 2019). Although a variety of elevational diversity patterns are found among the various taxa and from different regions, exploration of the unknown patterns of diversity along the altitude gradients and their formation mechanisms are still ongoing projects. So far, the elevation patterns of overall community variation (α-diversity) are known as follows: (1) α-diversity declines with increasing elevation, (2) α-diversity shows a hump-shaped or bimodal-shaped distributions with increasing elevation, (3) α-diversity increases with increasing elevation, and (4) α-diversity has no linear relation with elevation (Rahbek, 2005; Peters et al., 2016; Noriega and Realpe, 2018; Supriya et al., 2019; Uhey et al., 2021). The community composition variation (β-diversity) has been interpreted as the following two different features. First, the nestedness (species are lost or gained) is prevalent at high elevation and low elevation, or only present at one of them, while the turnover (a species replaced by other species) is common at mid-elevation (da Silva et al., 2018; Uhey et al., 2021). Second, only turnover is present in all the elevational gradients (Paknia and Sh, 2015; Young et al., 2019). Combining both the α-diversity and β-diversity gives a more comprehensive understanding of the biotic heterogeneity of a certain area (Wang et al., 2000; Baselga, 2010).

The insects represent a major component of global biodiversity and are widely used as an indicator of biodiversity in ecological studies. The previous studies on the insect biodiversity along the elevational gradients are confined to a small number of groups with ecological significance, such as geometrid moths (Axmacher et al., 2004; Beck and Chey, 2008), butterflies (Despland et al., 2012), pollinating hymenopterans (Jackson et al., 2018; Perillo et al., 2021), ants (Reymond et al., 2013; Bishop et al., 2014), dung beetles (da Silva et al., 2018; Stanbrook et al., 2021), and ground beetles (Zou et al., 2014). Moreover, the studied taxa have distinctly differed distribution patterns along elevation (Rahbek, 2005). The green lacewings (Neuroptera: Chrysopidae), a group of important predatory biocontrol agents with a positive role in the management of aphids, coccids, thrips, whiteflies, and mites in the ecosystem (Alford, 2019; Lai and Liu, 2020), had rarely been investigated. The previous studies on the diversity of green lacewings mainly focused on the species richness and/or Shannon-Wiener diversity index in agro-ecosystem (Deutsch et al., 2005; Thierry et al., 2005; Martins et al., 2019) or forest ecosystem (Yi et al., 2018), but rarely investigated on the pattern of β-diversity. For example, Bozdoğan (2020) preliminarily explored the richness patterns of a green lacewing community along the elevational gradients from the East Mediterranean but did not document the β-diversity of green lacewing community along the elevational gradients and correlated temperature variation up to now. On the other hand, understanding the taxonomic diversity, distribution pattern, and driving factors of the green lacewings are all critical for the exploration of the dominant species as the natural enemies in the different agricultural ecosystems (Liu et al., 2020) and developing conservation strategies (Zou et al., 2014). However, the diversity of green lacewings had been rarely taxonomically known, and their distribution patterns and responses to the environmental factors were rarely investigated, particularly, in the remote mountainous region of Southwestern China (e.g., the Hengduan Mountains), one of the most important global diversity hotspots, where the complex topography and climate were thought to be in favor of high level of biodiversity (Wu et al., 2013; Wen et al., 2016).

In the recent two decades, owing to the molecular species delimitation accelerated by DNA barcoding (Hebert et al., 2003; Ruiter et al., 2013; Hendrich et al., 2015), the large-scale biodiversity investigation has broken through the limit for analyzing the diversity at the family level (Negi and Gadgil, 2002; Zou et al., 2020), but is turning to focus further on the species richness (Smith et al., 2005; Woodcock et al., 2013; Young et al., 2019). Nevertheless, accurate molecular species identification relies on the establishment of the DNA barcoding library of the specific groups and the development of an appropriate analytical tool for species delimitation (Ruiter et al., 2013; Puillandre et al., 2021). So far, various approaches [e.g., General Mixed Yule Coalescent (GMYC), Automatic Barcode Gap Discovery (ABGD), Poisson tree processes (bPTP), multi-rate Poisson tree processes (mPTP), and Assemble Species by Automatic Partitioning analysis (ASAP)] are mainly applied in the species delimitation based on DNA barcoding. GMYC and bPTP estimate the absolute time and mutational time, respectively, at different nodes of the phylogenetic tree (Pons et al., 2006; Zhang et al., 2013; Puillandre et al., 2021). ABGD automatically identifies the limit between the smaller intraspecific distances and the larger interspecific distances based on the distribution of pairwise genetic distance within *priori* maximal genetic intraspecific divergence *P* (Puillandre et al., 2012, 2021). mPTP alleviates the theoretical and technical shortcomings of bPTP and takes account of divergent intraspecific variation. ASAP proposes the species partitions are ranked by a scoring system that does not need the biological prior insight of intraspecific diversity (Puillandre et al., 2021). However, the resulted divisions of potential species inferred by these methods are sometimes inconsistent, which thus calls an integrative approach combining these methods and morphological evidence as well (Ducasse et al., 2020; Puillandre et al., 2021).

Based on a 2-year survey on the lacewing diversity from the Shaluli Mountains (Mts. Shaluli), we aimed to (i) explore the species diversity of green lacewings from the Mts. Shaluli by using multiple approaches of species delimitation (GMYC,

ABGD, bPTP, mPTP, and ASAP) based on the DNA barcodes, and to (ii) reveal the community diversity patterns of green lacewings along elevation and the effect of temperature on their broad-scale (county scale) community diversity. We predict that (i) α-diversity may reach the peak from the mid-elevation areas, as the green lacewings are apt to suffer human disturbance from the low-elevation areas (Bozdoğan, 2020; Dupont and Strohm, 2020; Lai and Liu, 2020), and the fitness to the habitats from the high-elevation areas is low; (ii) α-diversity may increase with the increasing temperature; (iii) greater nestedness component of β-diversity may occur at sites with a high elevation and low temperature as many thermophilic organisms are limited at higher elevations by their inability to withstand cold temperature (Fu et al., 2007).

MATERIALS AND METHODS

Study Area and Sampling

The Mts. Shaluli is one of the major parts of the Hengduan Mountains, located at the west of Sichuan, southwestern China, and is featured in a complicated topography, with striking vertical zonation, ranging from a dry-hot valley canopy to alpine permafrost. The climates vary dramatically, with the difference of mean annual temperature reaching up to 20.2°C (Wu et al., 2013; Wen et al., 2016). The high floral diversity and wide range of habitat types make the Hengduan Mountains form a hotspot of biodiversity (Yao et al., 2015). Additionally, the Mts. Shaluli is a key area producing apple and pear in Sichuan Province (Xie et al., 2011). With conspicuous elevation gradients and rich species diversity of the green lacewings, the Mts. Shaluli is an ideal candidate for the regional ecosystem for exploring the responses of green lacewings to environmental change, and these findings will contribute to the bio-control application using the local green lacewings.

All the green lacewing specimens herein studied were collected from July to August of 2019 and 2020 from 48 sampling sites of 113 sampling sites (27.3094–30.2623 N, 99.0108–101.7959 E) in eight counties within the Mts. Shaluli (**Figure 1** and **Supplementary Table 1**), containing six important reserves (Zhubalong Nature Reserve, Batang; Xiayong Nature Reserve, Derong; Fozhuxia Nature Reserve, Xiangcheng; Gexigou Nature Reserve, Yajiang; Haizishan Nature Reserve, Litang; and Qialangduoji Nature Reserve, Muli) and its adjacent regions. The number of sampling sites of every county varied from 11 to 22 depending on the size of the area of these nature reserves. The specimens were collected by sweep-nets with a diameter of 50 cm from 9 am to 6 pm within a day, and also by light trap with 450 W high-pressure mercury lamp from 8 pm to midnight. No green lacewings were collected from the remaining sites. The habitat types of the sampling sites comprise (1) croplands, (2) broadleaved deciduous forests, (3) residential areas, (4) mixed coniferous broad leaved forests, (5) mixed coniferous broad leaved forests and residential areas, (6) coniferous forests, (7) alpine shrubs, (8) grasslands, (9) wetland, and (10) rivers, ranging from 1,338 to 4,525 m in altitude. All the specimens are preserved in 95% ethanol or pinned, and preserved at −20°C or room temperature, respectively. All the specimens are deposited in

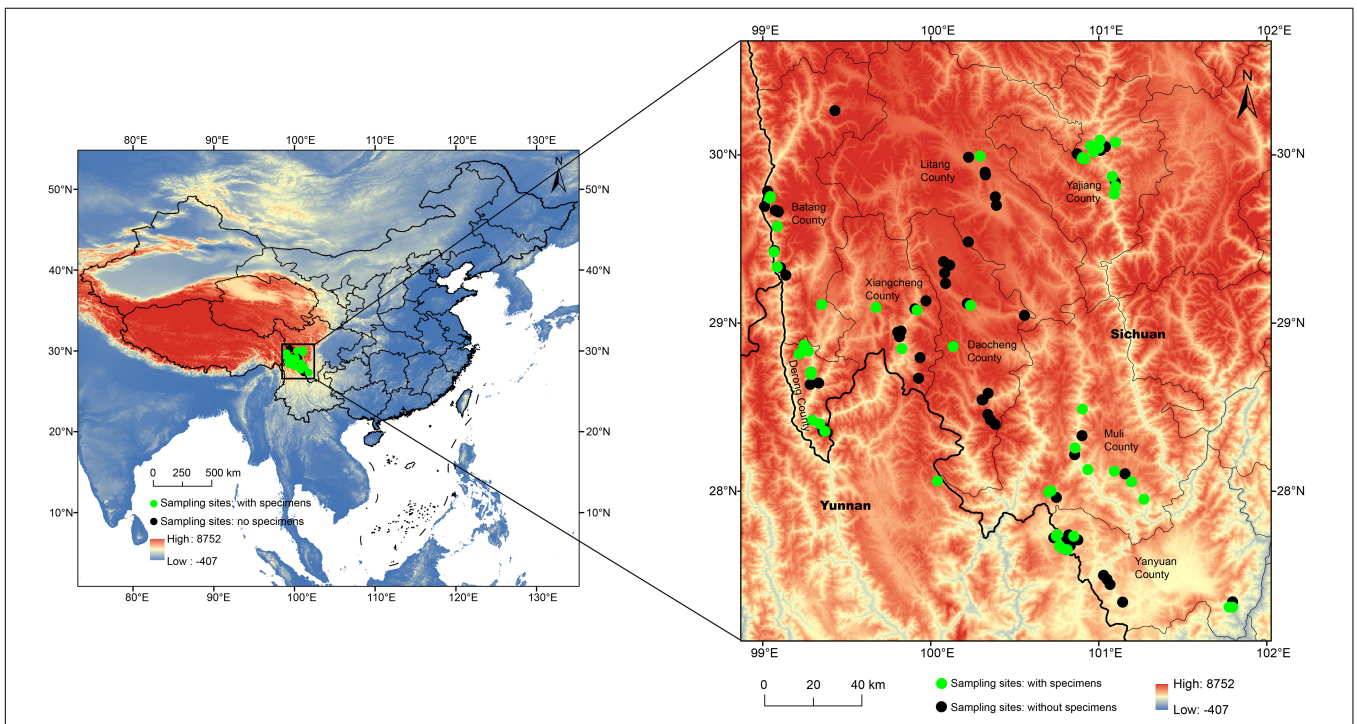


FIGURE 1 | The locations of the 113 sampling sites in the Shaluli Mountains. A green circle represents the site with green lacewings collected, while a black circle represents the site without green lacewings collected.

the Entomological Museum of China Agricultural University (CAU), Beijing, China.

Preliminary Species Identification Based on Morphology

The green lacewing specimens were identified to species or morphospecies based on the morphological characters by using the published keys to genera and species (Brooks and Barnard, 1990; Yang et al., 2005). The habitus photographs of the identified specimens are deposited and can be accessed in the BOLD database (dx.doi.org/10.5883/DS-CXB1).

DNA Barcoding

We selected 3–5 individuals per species or morphospecies to obtain the DNA barcodes. For the species that are difficult to be distinguished or have distinct potential intraspecific variations, over five individuals were selected for the analysis. The genomic DNA was extracted from the hind leg of each specimen using the TIANamp Genomic DNA Kit (TIANGEN Inc., Beijing, China). Because the universal DNA barcoding primers usually applied for the insects (i.e., LCO1490 and HCO2198; as shown in the reference Folmer et al., 1994) did not work well for the green lacewings in our study, the barcoding region of the mitochondrial *COI* gene (Hebert et al., 2003) was amplified using primers COI1-F (5′-ATTCAACCAATCATAAAGATATTGG-3′) and COI1-R (5′-TAAACTTCTGGATGT-CCAAAAAATCA-3′), which were accessed from the data named as "mitochondrion *Chrysopa oculata* (golden-eye lacewing)" in NCBI[1]. In case of unsuccessful amplification, an alternative forward primer COI4-F (5′-TGTAAAACGACG-GCCAGTAAACTAATARCCT TCAAAG-3′)[2] (reference to Park et al., 2010) was applied to amplify the selected region. Each 25 μl volume contained 12.5 μl 2 × EasyTaq PCR SuperMix (TransGen Biotech, Beijing, China), 9.5 μl ddH_2O, 1 μl of each primer, and 1 μl DNA template. The thermal cycling conditions followed Yi et al. (2018), but the annealing temperature was 52°C for COI4-F/COI1-R and 50°C for COI1-F/R. The PCR products were sequenced on ABI3730XL Genetic Analyzer by Beijing Ribio BioTech Co., Ltd., Beijing, China.

The sequence reads of *COI* in forward and reverse directions were assembled by ContigExpress (NY, United States), and we trimmed the tRNA-W region that owned stop codon at <200 bp of the 5′ end sequence based on the translation of amino acid. The sequence of standard DNA barcoding region of *Chrysoperla nipponensis* (trimmed from the complete mitochondrial genome DNA of this species, GenBank accession number: AP011623.1) was selected to join in the multiple sequence alignment in MEGA v.7.0 (Kumar et al., 2016) to confirm the tRNA-W region completely trimmed.

To assess the efficiency of our DNA barcoding database, the DNA barcoding gap was explored by calculating the nearest neighbor (NN) distance and maximum intraspecific (MI) distance based on a Kimura 2-parameter (K2P) model in MEGA v.7.0. The sequences with at least 500 bp were retained for the species delimitation.

The species delimitation based on the DNA barcodes was conducted by using the five methods as follows. First, based on the pairwise genetic distances, we performed the ABGD analysis (Puillandre et al., 2012) using the webserver[3]. The settings were as follows: relative gap width X = 1.0, K2P distance, and the prior maximum divergence of intraspecific diversity (P) values ranged from 0.001 to 0.100, other parameter values employed defaults. Second, we performed the Assemble species by ASAP (available at)[4], which is also a K2P distance-based approach. Partition with the smallest score is considered as the final result of the species delimitation (Puillandre et al., 2021). The remaining three analyses, i.e., GMYC (Pons et al., 2006), PTP (Zhang et al., 2013), and mPTP (Kapli et al., 2017), are phylogeny-based approaches. The maximum-likelihood (ML) tree was generated by using an IQ-TREE webserver (Nguyen et al., 2015; Trifinopoulos et al., 2016) based on the *COI* sequences under the GTR model[5]. The trees were saved as Newick format in the software FigTree v.1.4.3 (Rambaut and Drummond, 2016), then imported into the PTP[6] and mPTP[7] web server to delimit the species. For the GMYC analysis, the ultrametric trees were reconstructed with BEAST v.2.6.3 (Bouckaert et al., 2014). The relaxed lognormal clock model, Coalascant Constant Population tree priors were used to estimate the relative divergence times (Monaghan et al., 2009; Michonneau, 2015). The Markov Chain Monte Carlo (MCMC) generations were set to 200,000,000. TRACER v.1.7.0 (Rambaut et al., 2018) was utilized to check whether all the effective sample size (ESS) values have exceeded 200 for assessing the convergence of runs. The consensus trees were generated after discarding 25% of trees as burn-in (Puillandre et al., 2021).

Diversity Analysis

For the analysis of the diversity patterns of green lacewings along elevation, to avoid environmental heterogeneity among the different counties, 14 sampling sites within the same county (i.e., Derong County) with obvious elevational gradient (2093–4118 m) were selected to analyze (**Supplementary Table 2**). The elevational gradients of Derong County (range from 2,093 to 4,188 m) were partitioned into low-elevation (below 2,500 m), mid-elevation (2,501–3,500 m), and high-elevation (more than 3,501 m). For the analysis of the temperature effect on broad-scale (county scale) community diversity of the green lacewings, as it was considered as a key driver of an elevational diversity pattern (Bishop et al., 2014), we analyzed the data from the 33 sampling sites among the six sampling counties, because their historical daily maximum temperature, daily minimum temperature, and daily mean temperature were available in the China Meteorological Data Service Centre[8] (**Table 1**), and the mean elevation of these

[1]https://www.ncbi.nlm.nih.gov/nuccore/HQ552716

[2]https://www.ncbi.nlm.nih.gov/nuccore/GU013582

[3]https://bioinfo.mnhn.fr/abi/public/abgd/

[4]https://bioinfo.mnhn.fr/abi/public/asap/

[5]http://iqtree.cibiv.univie.ac.at/

[6]https://species.h-its.org/ptp/

[7]http://mptp.h-its.org/

[8]http://data.cma.cn/

TABLE 1 | Temperature data of six temperature blocks in the Shaluli Mountains.

Temperature block	Daily maximum temperature (°C)	Daily minimum temperature (°C)	Daily mean temperature (°C)	Mean elevation (m)
Batang County	28.50	15.64	20.07	3,208
Derong County	31.72	18.48	24.26	3,167
Litang County	19.10	6.80	12.10	3,990
Muli County	25.05	14.87	18.98	2,723
Yanyuan County	25.46	14.70	18.88	2,477
Daocheng County	21.30	7.00	13.40	3,167

The historical temperature data are from China Meteorological Data Service Centre.

counties form an elevational gradient on broad-scale. Then, we calculated the mean of these temperature data, respectively, for the final analysis. The temperature data from the sampling site 7–6 (27.7333 N, 100.7473 E) belonging to the Ninglang County was used as same as that from the sampling site 7–9 (27.7418 N, 100.7539 E) belonging to the Yanyuan County, because they are geographically near to each other (1,151 m), and the Ninglang County lack climate station.

The correlation between α-diversity (abundance, species richness, and Shannon-Wiener diversity index) (Kaltsas et al., 2018) of green lacewings community and the variate we analyzed (elevation and temperature) were assessed in SPSS v.19.0 (IBM, NY, United States) with Pearson's correlation coefficient. If there is a correlation, the linear regression model ($y = bx + a$) or quadratic regression model [($y = cx^2 + a$) or ($y = cx^2 + bx + a$)] will be used to fit (Werenkraut and Ruggiero, 2014). The goodness of fit was evaluated based on the adjusted R^2 (Yang et al., 2020). The species component difference between pairs of sites was assessed by β-diversity. The Sørensen dissimilarity index (βsor) based on the species richness (Baselga, 2010) was used to measure the overall β-diversity (Fu et al., 2019). Following Baselga (2010), βsor was partitioned into Simpson dissimilarity index (βsim) and nestedness-resultant dissimilarity (βnes). To determine the most important component, the comparisons between βnes and βsor were performed for pairs of sites (Dobrovolski et al., 2012). The formulas are as follows:

$$\beta\text{sor} = \frac{b + c}{2a + b + c}$$

$$\beta\text{sim} = \frac{\min(b, c)}{a + \min(b, c)}$$

$$\beta\text{nes} = \beta\text{sor} - \beta\text{sim}$$

where a is the number of species common to both sites, b is the number of species that only occur in the first site, and c is the number of species that only occur in the second site (Baselga, 2010).

To evaluate whether the α-diversity and β-diversity between the different elevational gradients are significantly different, the ANOVA of the Shannon-Wiener diversity index and βsor among the different elevational gradients would be conducted in SPSS v.19.0 with the generalized linear models, a paired comparison was performed by Tukey's significant difference *post-hoc* test (Yang et al., 2020).

RESULTS

Species Delimitation

A total of 1,168 specimens of Chrysopidae, belonging to 12 genera and 40 morphospecies (**Supplementary Table 1**), were collected from the Mts. Shaluli. In total, 246 specimens of 40 morphospecies were selected and their DNA barcodes (547–705 bp) were successfully obtained (**Table 2**; GenBank accession number: MZ557095–MZ557340). There are 181 (73.58%) barcodes from 22 morphospecies identified as the described species. The remaining 18 morphospecies were identified only to the genus.

The results from the different species delimitation methods generally correspond to the identifications of the morphospecies, but with some discrepancies (**Figure 2**). mPTP was the most conservative method on the species delimitation, with the fewest mOTUs (=36) recovered. Both the genetic distance-based approaches, i.e., ABGD and ASAP, resulted in more mOTUs (=37) than that from mPTP. However, GMYC and bPTP delimitated much more mOTUs (41 and 42, respectively). Specifically, each of the following four pairs of morphospecies, namely *Chrysoperla furcifera* (Okamoto)/*Mallada* sp., *Nineta vittata* (Wesmael)/*Nineta* sp. 1, *Cunctochrysa* sp. 1/*Cunctochrysa* sp. 2, and *Chrysoperla annae* Brooks/*Chrysoperla nipponensis* (Okamoto), was delimited to be the same mOTU by mPTP. Similarly, ABGD and ASAP, respectively, recognized three pairs of morphospecies [i.e., *Ch. annae*/*Ch. nipponensis*, *Anachrysa* sp. 1/*Anachrysa* sp. 2, and *Chrysopidia* (*Chrysopidia*) *remanei* Hölzel/*Chrysopidia* (*Chrysopidia*) sp. 1] as the same mOTU, with prior maximal distance value as 0.022. *Retipenna parvula* Yang et al. and *Apertochrysa jiuzhaigouana* (Yang et al.), respectively, were separated into two mOTUs by bPTP. *Cunctochrysa albolineatoides* Tsukaguchi was divided into two mOTUs by GMYC.

After alignment and trimming, there are 245 barcodes with 557 bp used for the calculation of the genetic distances. A distinct barcode gap was presented in 93.10% of all the analyzed species (**Figure 3**). The minimal barcode gaps were presented in *Ch. annae* and *Ch. nipponensis*. The MI distance was close to the interspecific distance between NN species (MI vs. NN: 0.016 vs. 0.013 for *Ch. annae*; 0.016 vs. 0.009 for *Ch. nipponensis*) (**Supplementary Tables 3, 4**).

Species Diversity

Among the 40 species herein delimitated, the four species are newly recorded in China, and 11 species are newly recorded in Sichuan Province (**Supplementary Table 5**). *Apertochrysa* owns the highest richness (22.50% of species), followed by *Chrysopidia* (17.50% of species). Additionally, *Apertochrysa* owns the highest abundance (33.73% of individuals), followed by *Chrysopa* (17.51% of individuals), and *Cunctochrysa* (16.35% of individuals). Three dominant species from the Mts. Shaluli was recognized and listed as follows (**Figure 4**). *Apertochrysa barkamana* (Yang et al.), accounting for 20.98% of the total number of individuals, is a dominant species in Yajiang County with elevation

TABLE 2 | The DNA barcodes of Chrysopidae from the Shaluli Mountains.

Species		Number of barcodes	GenBank accession number
Belonopterygini	*Italochrysa ludingana*	1	MZ557095
	Italochrysa sp.	5	MZ557096–MZ557100
Ankylopterygini	*Ankylopteryx* (*Ankylopteryx*) *octopunctata*	2	MZ557101-MZ557102
	Retipenna parvula	3	MZ557103–MZ557105
Chrysopini	*Anachrysa* sp. 1	9	MZ557106–MZ557114
	Anachrysa sp. 2	1	MZ557115
	Apertochrysa barkamana	48	MZ557116–MZ557163
	Apertochrysa eumorpha	3	MZ557164–MZ557166
	Apertochrysa jiuzhaigouana	15	MZ557167–MZ557181
	Apertochrysa prasina	20	MZ557182–MZ557201
	Apertochrysa sp. 1	2	MZ557214–MZ557215
	Apertochrysa sp. 2	1	MZ557216
	Apertochrysa sp. 3	1	MZ557217
	Apertochrysa sp. 4	2	MZ557218–MZ557219
	Apertochrysa sp. 5	12	MZ557202–MZ557213
	Chrysopa formosa	3	MZ557220–MZ557222
	Chrysopa pallens	4	MZ557223–MZ557226
	Chrysopa sp. 1	11	MZ557227–MZ557237
	Chrysoperla annae	9	MZ557238–MZ557246
	Chrysoperla furcifera	1	MZ557247
	Chrysoperla nipponensis	6	MZ557248–MZ557253
	Chrysoperla sp. 1	1	MZ557254
	Chrysopidia (*Chrysopidia*) *remanei*	3	MZ557255–MZ557257
	Chrysopidia (*Chrysopidia*) *zhaoi*	12	MZ557258–MZ557269
	Chrysopidia (*Chrysopidia*) sp. 1	7	MZ557270–MZ557276
	Chrysopidia (*Chrysopidia*) sp. 2	1	MZ557277
	Chrysopidia (*Chrysotropia*) *ciliata*	3	MZ557278–MZ557280
	Cunctochrysa albolineatoides	8	MZ557281–MZ557288
	Cunctochrysa shuenica	13	MZ557289–MZ557301
	Cunctochrysa sp. 1	2	MZ557302–MZ557303
	Cunctochrysa sp. 2	5	MZ557304–MZ557308
	Himalochrysa bhandarensis	5	MZ557309–MZ557313
	Himalochrysa sp. 1	4	MZ557314–MZ557317
	Mallada krakatauensis	3	MZ557318–MZ557320
	Mallada sp.	1	MZ557321
	Nineta vittata	1	MZ557322
	Nineta sp. 1	1	MZ557323
	Nineta sp. 2	2	MZ557324–MZ557325
	Tumeochrysa tibetana	10	MZ557326–MZ557335
	Tumeochrysa yunica	5	MZ557336–MZ557340

ranging from 2,087 to 3,579 m; *Cunctochrysa albolineatoides*, accounting for 13.61% of the total number of individuals, is distributed in all the sampling counties, ranging from 2,087 to 4,202 m in elevation; *Chrysopa* sp. 1 accounting for 10.36% of the total number of individuals, is dominated in Batang County and Derong County with elevation ranging from 2,093 to 2,409 m.

Elevational Diversity Pattern

Given the obvious elevational gradient (2,093–4,118 m) and sufficient sampling, Derong County was selected for the analysis of the elevational diversity pattern of the green lacewings. Based on the above species delimitation, there are 17 species (a total of 249 specimens) from Derong (**Supplementary Table 2**). The abundance, richness, and Shannon-Wiener diversity index of the green lacewing communities from Derong County were significantly correlated with elevation (R^2 = 0.66, df = 11, $P < 0.01$; R^2 = 0.53, df = 12, $P < 0.01$; and R^2 = 0.34, df = 12, $P < 0.05$, respectively), and have the same elevational-diversity pattern, being decreased with the increasing elevation, as shown by the better linear regression model (**Figure 5A**), but unlike richness and Shannon-Wiener diversity index, the abundance of the green lacewing communities was not monotonically decreasing, and its rate of decrease gradually flattened out. The ANOVA result (**Table 3**) showed that the Shannon-Wiener diversity index of the green lacewing communities in high-elevation was significantly different from that in the low-elevation and mid-elevation, and there was no significant difference between the low-elevation and mid-elevation. There was a difference in the Sørensen dissimilarity index (βsor) based on the species richness of the paired sites in the different elevational gradients, but their differences among the elevational gradients were not significant (**Table 3**). Nestedness-resultant dissimilarity (βnes) became a major component of β-diversity with the increasing elevation (**Figures 5B,C**).

Species Diversity and Temperature Variables

The abundance of the green lacewing communities from six counties (each herein defined as a temperature block; **Table 1**) located at the Mts. Shaluli significantly increased with the increasing temperature (**Figure 6**). However, the richness was not significantly related to temperature in the present results. Shannon-Wiener diversity index of the green lacewing communities was positively correlated with daily minimum temperature and daily mean temperature, but not daily maximum temperature (**Figure 6**).

Compared with the other temperature blocks (**Table 1** and **Supplementary Table 6**), the ratio of βnes to βsor from the low-temperature blocks (i.e., Daocheng County and Litang County; daily maximum temperature below 21.40°C, daily minimum temperature below 7.10°C, and daily mean temperature below 13.50°C) was 1 or 0, indicating that low temperature could lead to the nesting of communities or the absence of common species shared among the communities. Besides, the results show that when the difference of daily minimum temperature or daily mean temperature between the paired temperature blocks exceeds 6°C, the difference of species components or nestedness between the communities reaches the greatest level.

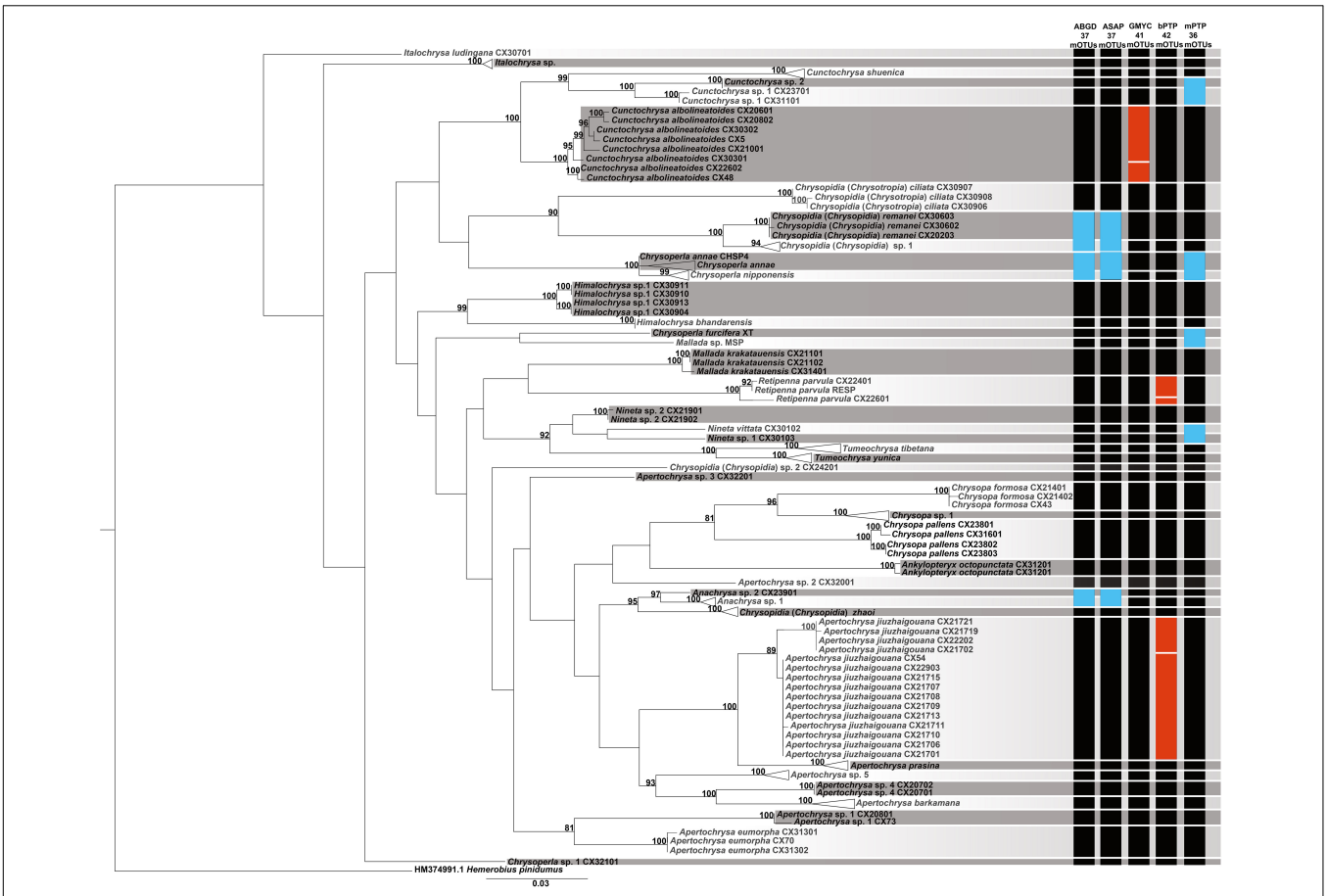


FIGURE 2 | Molecular species delimitation of the green lacewings from the Shaluli Mountains through multiple approaches based on DNA barcodes. A maximum-likelihood (ML) tree was based to summarize the results. Only the bootstrap values > 80 were given at relevant nodes. The vertical colored bars indicate putative species delimitated by the different approaches. A black bar indicates congruent results between the molecular and morphological identifications, a red bar indicates putative species divided within morphospecies, and a blue bar indicates a single putative species comprising two or more morphospecies.

DISCUSSION

Barcoding of Green Lacewings

The green lacewings represent one of the most diverse lineages of Neuroptera. Although they are commonly found in various ecosystems, such as forests, orchards, and croplands, the identifications of many green lacewing species, even of some frequently used natural enemy species, are difficult to practice solely based on the morphological characters because their interspecific differences are indistinct (Pappas et al., 2011; Lai and Liu, 2020). So far, there have been few studies that attempted to identify the green lacewings using the DNA barcodes (Morinière et al., 2014; Yi et al., 2018). Our analysis of the green lacewing barcodes from the Mts. Shaluli shows a distinct barcode gap presented in 93.10% of all the analyzed species (**Figure 3**), allowing unambiguous identification of 92.5% species. It suggests that DNA barcoding is promising for the accurate identification of the green lacewing species. Nevertheless, a well-established library of the DNA barcodes is an important precondition, which requires species delimitation facilitated by integrative data from morphology, molecule, and distribution.

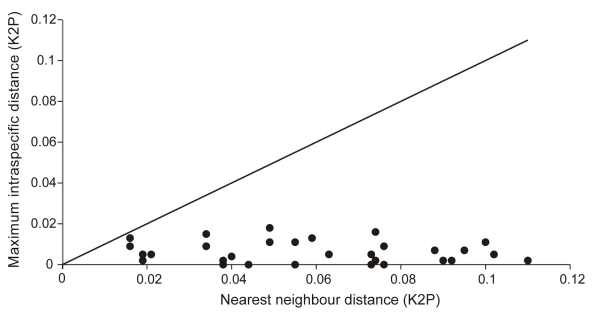


FIGURE 3 | A plotting chart showing the barcode gap for 30 green lacewing species. The dots below the 1:1 line indicate the presence of a barcode gap.

Our results (**Figure 2**) suggest that mPTP is too conservative to be implemented in species delimitation of the green lacewings based on the DNA barcodes. In addition, ABGD and ASAP similarly performed as mPTP, being relatively conservative to delimitate closely related species. Here, GMYC and bPTP

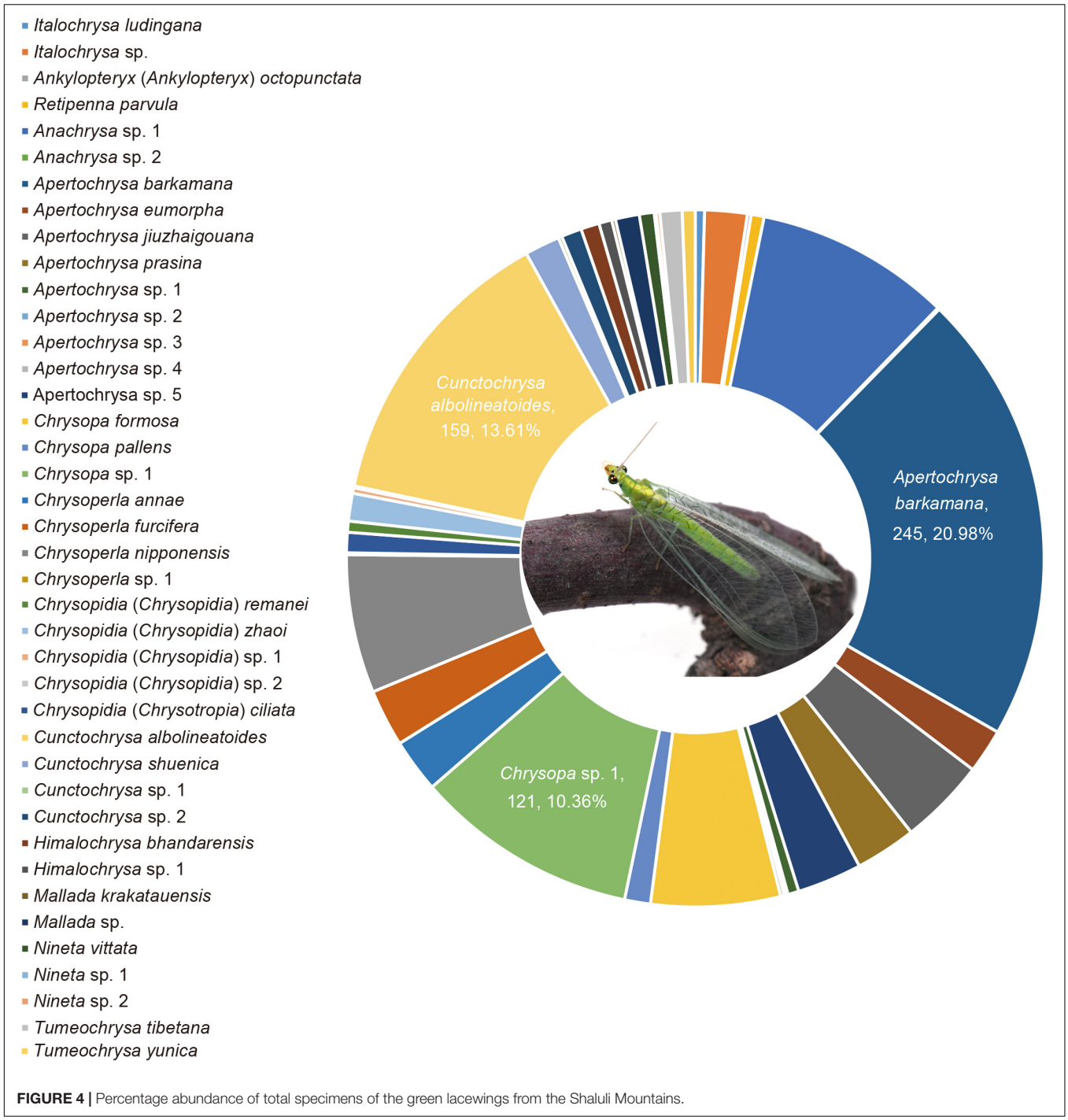


FIGURE 4 | Percentage abundance of total specimens of the green lacewings from the Shaluli Mountains.

performed better because they could correctly distinguish the most closely related species or even cryptic species. A similar case is found for the species delimitation of Psylloidea (Hemiptera: Sternorrhyncha) (Martoni et al., 2018). Accordingly, for the common and species-rich genera *Chrysoperla*, *Apertochrysa*, and *Mallada*, GMYC and bPTP are recommended for molecular species delimitation even though they are time-consuming (7 s for ABGD and ASAP, 7 h for bPTP, and 33 h for GMYC in this study). When delimitating closely related species, however, approaches combining morphological and ecological evidence are needed (Puillandre et al., 2012; Huang et al., 2020).

Patterns of Elevational Diversity

The previous studies on some organisms from the Hengduan Mountains (Gong et al., 2005; Rahbek, 2005; Fu et al., 2006, 2007; Li et al., 2009; Wu et al., 2013; da Silva et al., 2018; Uhey et al., 2021) found that the hump-shaped elevational α-diversity patterns are most prevalent, while a monotonic

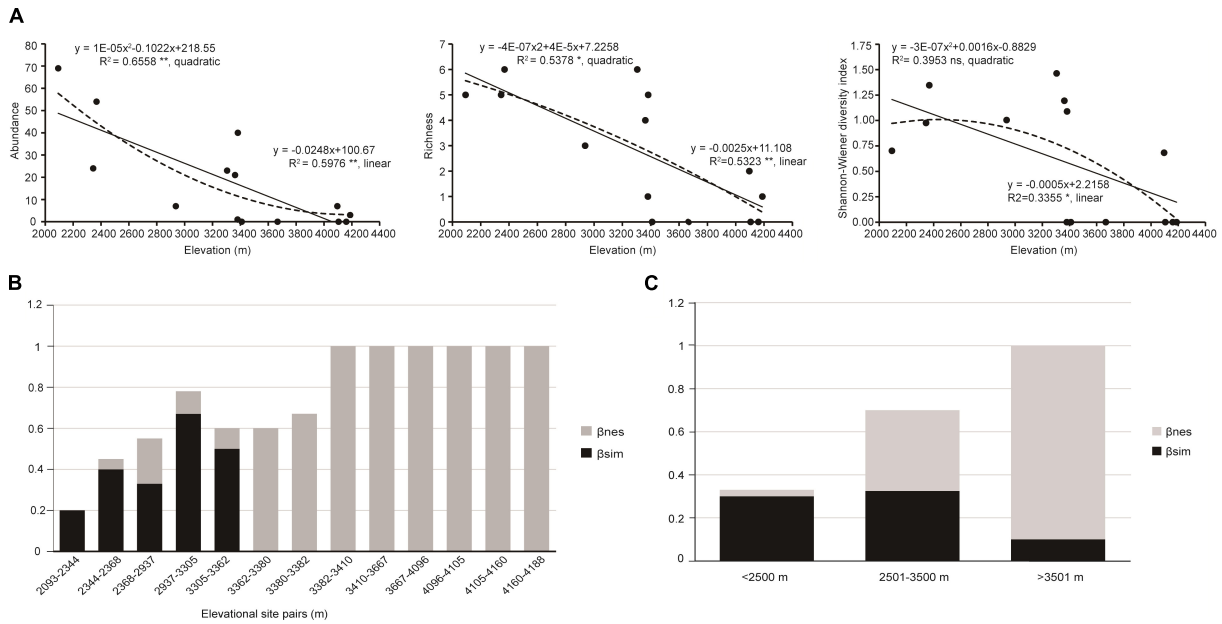


FIGURE 5 | The elevational diversity patterns of the green lacewing community from Derong County. **(A)** Relationships between elevation and α-diversity. The linear and quadratic regression were used to fit the elevation pattern. The goodness of fit was assessed based on the adjusted R^2. **, *, and ns indicate significance at $P \leq 0.01$, $P \leq 0.05$, and non-significant, respectively. The solid lines indicate linear regression and the dashed lines indicate quadratic regression. Dissimilarity analysis (Sørensen dissimilarity index, βsor) of the green lacewings from Derong Country partitioned into Simpson dissimilarity index (βsim) and nestedness-resultant dissimilarity (βnes) of **(B)** elevational site pairs (i.e., each site and its immediately adjacent uphill site), and **(C)** elevational gradients.

decreasing pattern with increasing elevation is also frequently reported (Rahbek, 2005; Sam et al., 2015; Noriega and Realpe, 2018; Finnie et al., 2021). The richness and Shannon-Wiener diversity index of the green lacewings from Derong County are consistent with the latter pattern mentioned above, the relationship of the abundance of green lacewings and elevation fit with the right half of hump-shaped (**Figure 5A**), which contradict our hypothesis of highest α-diversity from the mid-elevation areas. This may be explained in the following two aspects. First, incomplete sampling from the areas with different elevations could result in biased patterns of elevational diversity. Second, the definition of low- or mid-elevation may vary among the different regions. Here, low-elevation (2,000–2,500 m) is frequently defined as mid-elevation (Peters et al., 2016; Supriya et al., 2019). However, Bozdoğan (2020) found the monotonic decreasing pattern of green lacewing species-richness along the increasing elevation from the East Mediterranean area of Turkey, which has a relatively lower elevation ranging from 400 to 1,400 m. Moreover, the elevational diversity is found to shift from a unimodal pattern toward a monotonic decline with increasing taxonomic coverage of the plant and animal communities (Peters et al., 2016). Thus, the patterns of elevational diversity appear to be variable rather than uniform among taxa (Rahbek, 1995). Different taxa, seasonality, climate, resource availability, and interaction between species (Beck et al., 2010; Sundqvist et al., 2013; Supriya et al., 2019) may contribute to the differentiation of diversity patterns. Concerning the presently revealed decline of α-diversity of the green lacewings along increasing elevation, the decline of temperature and the decrease of resources with higher elevation may have major effects on driving this pattern, because the lower temperature could hamper the growth and induce diapause of the green lacewings (Kostál, 2006; Wang et al., 2019), and the cannibalism among green lacewing larvae could be provoked when the preys are scarce (Mochizuki et al., 2006; Ye and Li, 2020).

TABLE 3 | The ANOVA result of α-diversity and β-diversity among the elevational gradients in Derong County.

Elevation	Shannon-Wiener index	βsor
Low-elevation	1.01 ± 0.32^{a}	0.37 ± 0.15^{ns}
Mid-elevation	0.79 ± 0.63^{a}	0.75 ± 0.26^{ns}
High-elevation	0.14 ± 0.31^{b}	0.70 ± 0.48^{ns}

Means with a different letter inside the same factor indicate significantly different (CI: 95%). ns, no significant differences (P > 0.05).

Being components of β-diversity, turnover and nestedness may better reflect the extent of changes in community composition among the sites (Baselga, 2010). We found that species turnover decreased and nestedness increased with the increasing elevation (**Figure 5**). Herein, turnover is the main component of β-diversity from low and middle elevation. It is supported by the distributions of most green lacewing species, which are not overlapped on a small scale at low and middle elevations. However, the difference of the Sørensen dissimilarity index (βsor) based on the green lacewing species richness among the elevational gradients is not significant (**Table 3**). In other words, the turnover from low- and mid-elevation has a similar

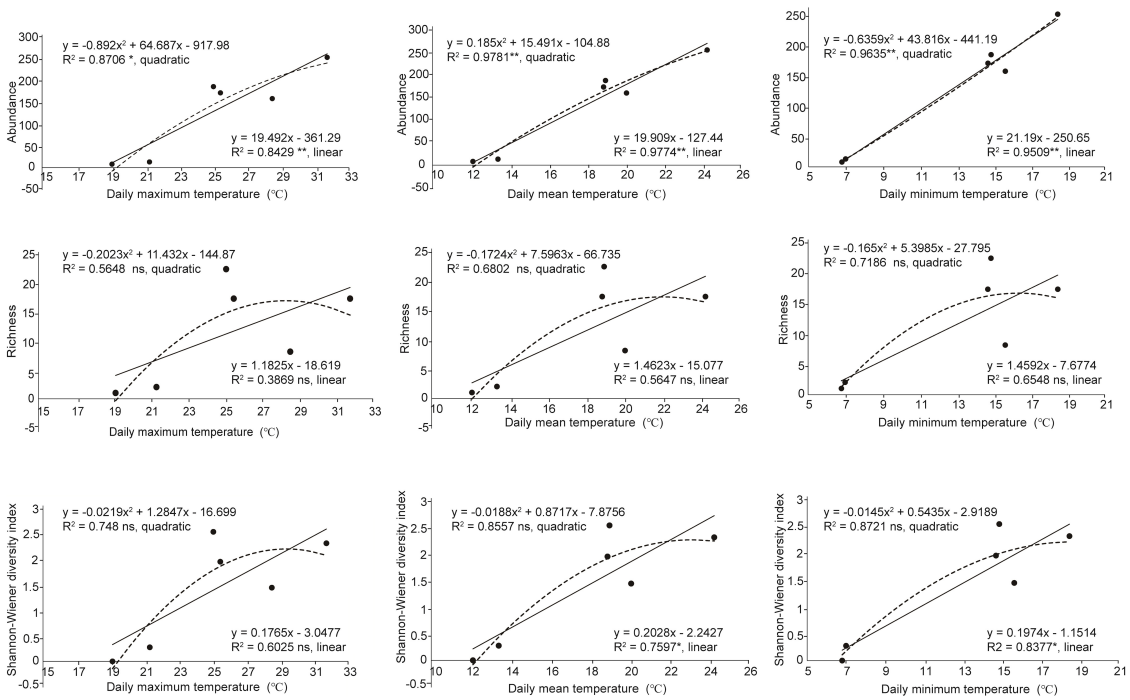


FIGURE 6 | Relationships between temperature and α-diversity of the green lacewing community from six counties (Batang County, Derong County, Litang County, Muli County, Yanyuan County, and Daocheng County). The historical temperature data of the above counties are from China Meteorological Data Service Centre. The linear and quadratic regression were used to fit the temperature pattern. The goodness of fit was assessed based on the adjusted R^2. **, *, and ns indicate significances at $P \leq 0.01$, $P \leq 0.05$, and non-significant, respectively. The solid lines indicate linear regression model and the dashed line indicates quadratic regression.

effect on the species composition as nestedness from a high elevation. Turnover is commonly due to species competitive exclusion and/or niche specialization (Ulrich et al., 2009; Baselga, 2010). Because there are a relatively high richness and abundance of the green lacewing communities at low- and middle-elevation, species competitive exclusion may be common in limited space at low- and mid-elevation for the predatory green lacewings, eventually leading to higher turnover in the two elevations. Other studies also found that the nestedness component is greater in higher latitude areas (Dobrovolski et al., 2012; Paknia and Sh, 2015), which may be relevant to physiologic limitation as it is too cold to survive for most green lacewings at high elevation.

Temperature Shaping Species Diversity

Temperature is considered to be an important factor to explain population dynamics (Geurts et al., 2012) and an important predictor of species richness along with the elevational gradients (Fu et al., 2007; Peters et al., 2016). In this study, the species diversity of the green lacewings is strongly affected by the temperature on a broad-scale elevational gradient. The relationship between the temperature and α-diversity of the green lacewings community was incompletely consistent with our hypothesis of α-diversity increases with the increasing temperature. We found that the abundance of green lacewings significantly increased with the increasing temperature (**Figure 6**). This finding further confirms that the terrestrial ectotherms may experience an initial increase in the population growth rates at mid- to high-latitudes with climate warming (Deutsch et al., 2008). Considering the richness of the green lacewing communities and temperature, there was no significant correlation. It indicates that other factors may independently or together with temperature change the richness of the green lacewings communities in different temperature blocks, such as water availability and geographic distance (Qian et al., 2005; Peters et al., 2016; Liu et al., 2020). The Shannon-Wiener diversity index of the green lacewing communities was positively correlated with daily minimum temperature and daily mean temperature, but not daily maximum temperature (**Figure 6**). The previous studies (Araújo et al., 2013; Liu et al., 2020) suggested that physiological tolerance to cold evolves more quickly than tolerance to heat in plants and animals. As such, the organisms are expected to have limited plasticity in the increase of their upper thermal limits (Bai et al., 2019), explaining more variability of the green lacewing species diversity among the cooler areas than that among the warmer areas.

The higher nestedness component of β-diversity occurred at the low-temperature blocks (**Supplementary Table 6**), which is consistent with our hypothesis. We found only three species from the low-temperature blocks, even though we made extensive collecting from 22 sampling sites within the two low-temperature blocks (**Supplementary Table 1**). The low-temperature blocks are located around Haizishan Mountain and its adjacent areas and their elevational range are above 4,000 m, where the glaciation

was particularly active during the Pleistocene and massive extinction of thermophilic organisms, such as green lacewings might have taken place (Cao and Zhao, 2007; Dobrovolski et al., 2012), as many thermophilic organisms are limited at higher elevations by their inability to withstand cold temperature (Fu et al., 2007).

Conservation Implications

As the increase in temperature above optima has negative effects on the organisms (Bozinovic et al., 2011), the species with limited thermal tolerance usually undergo uphill shifts or extinct with climate warming (Chen et al., 2009; Birkett et al., 2018; Stanbrook et al., 2021). Particularly, the tropical species are considered more sensitive to temperature than their counterparts from the temperate regions, because they currently live under nearly optimal temperature (Addo-Bediako et al., 2000; Deutsch et al., 2008; Chen et al., 2009), although our sampling region can be classified as temperate-zone, the climatic conditions vary dramatically among the different sites (Wu et al., 2013; Wen et al., 2016). The species of the community at low-elevation and from relatively high-temperature blocks with the greater turnover component are expected to shift uphill due to climate warming, which may alter the monotonic decline pattern to the unimodal pattern along increasing elevation. Besides, along with the increasing of species diversity at mid-elevation, extinction could be triggered by more severe interspecific competitions. Accordingly, the priority of green lacewing conservation should be given to the species from the regions with low- or mid-elevation, for example, protecting or artificially offering favorable habitats (forests for most species), and assessing the threatened species in these regions.

Cunctochrysa albolineatoides, *A. barkamana*, and *Chrysopa* sp. 1 are the dominant green lacewing species in Mts. Shaluli, and could be exploited as the potential biological control agents for the local region. Thus, more attention to conservation should be paid to these three species, especially for *A. barkamana* and *Chrysopa* sp. 1, because they may be more susceptible to climate warming as they mainly live in low-elevation and high-temperature regions.

CONCLUSION

We first analyzed the α-diversity and β-diversity of the green lacewings along elevation in a biodiversity hotspot of southwest China, combining DNA barcoding and the five molecular species delimitation methods. The distinct barcode gaps are present in 93.10% of species, which highlights the availability of DNA barcoding in regional species diversity assessment of the green lacewings. The elevational α-diversity of the green lacewings communities did not show a hump-shaped pattern similar to most of the previous studies but decreased with the increasing elevation. Nestedness replaced turnover as the main component of Sørensen dissimilarity index of the green lacewing communities with the increasing elevation, and greater nestedness occurred at low temperature blocks. Species diversity of the green lacewings from Mts. Shaluli is strongly affected by the temperature on a broad-scale elevational gradient. Additionally, the resource and resultant interspecific competitions may also be the potential drivers on the elevational diversity pattern of the green lacewings. Priority of conservation should be given to the species at low- and mid-elevation and relatively high-temperature regions facing the challenge of climate warming.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

XL, YhL, and YL conceived the idea for this study, designed the research, and wrote the manuscript. YL provided and analyzed the data. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Biodiversity Survey and Assessment Project of the Ministry of Ecology and Environment, China (No. 2019HJ2096001006), and the National Animal Collection Resource Center, China.

ACKNOWLEDGMENTS

We thank the following persons for helping us in the field works: Di Li, Juan Zhong, Yuezheng Tu, Liang Wang, Tao Li, Tianxiang Mao, Chongjun Huang, Xulong Chen, Bintao Du, Deyu Cui, Longbing Wang, Ning Yu, Xiaodong Cai, Gang Yao and Wenliang Li. We are grateful to authorities of Zhubalong Nature Reserve in Batang County, Xiayong Nature Reserve in Derong County, Fozhuxia Nature Reserve in Xiangcheng County, Gexigou Nature Reserve in Yajiang County, Haizishan Nature Reserve in Litang County, and Qialangduoji Nature Reserve in Muli County for permitting us to collect specimens. We are also grateful to Zihua Zhao for helping acquire historical temperature data of six temperature blocks in the Shaluli Mountains; Chengwu Zou for guiding in the analysis of biometrics; Aili Lin for helping the preparation of graphs; Lulan Jie for guiding in using ArcMap v.10.4.1; and Yunlong Ma for confirming the identification of some green lacewing species.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2021.778686/full#supplementary-material

REFERENCES

Addo-Bediako, A., Chown, S. L., and Gaston, K. J. (2000). Thermal tolerance, climatic variability and latitude. *Proc. R. Soc. B Biol. Sci.* 267, 739–745. doi: 10.1098/rspb.2000.1065

Alford, D. V. (2019). *Beneficial Insects*. Boca Raton, FL: Taylor & Francis Group.

Araújo, M. B., Ferri-Yanez, F., Bozinovic, F., Marquet, P. A., Valladares, F., and Chown, S. L. (2013). Heat freezes niche evolution. *Ecol. Lett.* 16, 1206–1219. doi: 10.1111/ele.12155

Axmacher, J. C., Holtmann, G., Scheuermann, L., Brehm, G., Müller-Hohenstein, K., and Fiedler, K. (2004). Diversity of geometrid moths (Lepidoptera: Geometridae) along an afrotropical elevational rainforest transect. *Divers. Distrib.* 10, 293–302. doi: 10.1111/j.1366-9516.2004.00101.x

Bai, C. M., Ma, G., Cai, W. Z., and Ma, C. S. (2019). Independent and combined effects of daytime heat stress and nighttime recovery determine thermal performance. *Biol. Open* 8:bio038141. doi: 10.1242/bio.038141

Barbarossa, V., Bosmans, J., Wanders, N., King, H., Bierkens, M. F. P., Huijbregts, M. A. J., et al. (2021). Threats of global warming to the world's freshwater fishes. *Nat. Commun.* 12:1701. doi: 10.1038/s41467-021-21655-w

Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. *Glob. Ecol. Biogeogr.* 19, 134–143. doi: 10.1111/j.1466-8238.2009.00490.x

Beck, J., Altermatt, F., Hagmann, R., and Lang, S. (2010). Seasonality in the altitude-diversity pattern of Alpine moths. *Basic Appl. Ecol.* 11, 714–722. doi: 10.1016/j.baae.2010.08.009

Beck, J., and Chey, V. K. (2008). Explaining the elevational diversity pattern of geometrid moths from Borneo: a test of five hypotheses. *J. Biogeogr.* 35, 1452–1464. doi: 10.1111/j.1365-2699.2008.01886.x

Birkett, A. J., Blackburn, G. A., and Menendez, R. (2018). Linking species thermal tolerance to elevational range shifts in upland dung beetles. *Ecography* 41, 1510–1519. doi: 10.1111/ecog.03458

Bishop, T. R., Robertson, M. P., van Rensburg, B. J., and Parr, C. L. (2014). Elevation-diversity patterns through space and time: ant communities of the Maloti-Drakensberg mountains of southern Africa. *J. Biogeogr.* 41, 2256–2268. doi: 10.1111/jbi.12368

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., et al. (2014). BEAST 2, a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10:e1003537. doi: 10.1371/journal.pcbi.1003537

Bozdoğan, H. (2020). Diversity of lacewing assemblages (Neuropterida: Neuroptera) in different forest habitats and agricultural areas in the East Mediterranean area of Turkey. *Entomol. Res.* 50, 163–173. doi: 10.1111/1748-5967.12426

Bozinovic, F., Bastías, D. A., Boher, F., Clavijo-Baquet, S., Estay, S. A., and Angilletta, M. J. (2011). The mean and variance of environmental temperature interact to determine physiological tolerance and fitness. *Physiol. Biochem. Zool.* 84, 543–552. doi: 10.1086/662551

Brooks, S. J., and Barnard, P. C. (1990). The green lacewings of the world: a generic review (Chrysopidae). *Bull. Br. Mus. (Nat. Hist.) Entomol.* 59, 117–286.

Cao, J., and Zhao, Y. N. (2007). The quaternary glacier in the Shaluli Mountains, West Sichuan and correlation with that in the neighborhood. *J. Sichuan Geol.* 27, 81–86.

Chen, I. C., Shiu, H. J., Benedick, S., Holloway, J. D., Chey, V. K., Barlow, H. S., et al. (2009). Elevation increases in moth assemblages over 42 years on a tropical mountain. *Proc. Natl. Acad. Sci. U.S.A.* 106, 1479–1483. doi: 10.1073/pnas.0809320106

da Silva, P. G., Lobo, J. M., Hensen, M. C., Vaz-de-Mello, F. Z., and Hernández, M. I. M. (2018). Turnover and nestedness in subtropical dung beetle assemblages along an elevational gradient. *Divers. Distrib.* 24, 1277–1290. doi: 10.1111/ddi.12763

Despland, E., Humire, R., and Martn, S. S. (2012). Species richness and phenology of butterflies along an altitude gradient in the desert of Northern Chile. *Arct. Antarct. Alp. Res.* 44, 423–431. doi: 10.1657/1938-4246-44.4.423

Deutsch, B., Paulian, M., Thierry, D., and Canard, M. (2005). Quantifying biodiversity in ecosystems with green lacewing assemblages. *Agron. Sustain. Dev.* 25, 337–343. doi: 10.1051/agro:2005036

Deutsch, C. A., Tewksbury, J. J., Huey, R. B., Sheldon, K. S., Ghalambor, C. K., Haak, D. C., et al. (2008). Impacts of climate warming on terrestrial ectotherms across latitude. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6668–6672. doi: 10.1073/pnas.0709472105

Diamond, S. E., Nichols, L. M., Pelini, S. L., Penick, C. A., Barber, G. W., Cahan, S. H., et al. (2016). Climatic warming destabilizes forest ant communities. *Sci. Adv.* 2:e1600842. doi: 10.1126/sciadv.1600842

Dobrovolski, R., Melo, A. S., Cassemiro, F. A. S., and Diniz-Filho, J. A. F. (2012). Climatic history and dispersal ability explain the relative importance of turnover and nestedness components of beta diversity. *Glob. Ecol. Biogeogr.* 21, 191–197. doi: 10.1111/j.1466-8238.2011.00671.x

Ducasse, J., Ung, V., Lecointre, G., and Miralles, A. (2020). LIMES: a tool for comparing species partition. *Bioinformatics* 36, 2282–2283. doi: 10.1093/bioinformatics/btz911

Dupont, S. T., and Strohm, C. J. (2020). Integrated pest management programmes increase natural enemies of pear psylla in Central Washington pear orchards. *J. Appl. Entomol.* 144, 109–122. doi: 10.1111/jen.12694

Finnie, S., Sam, K., Leponce, M., Basset, Y., Drew, D., Schutze, M. K., et al. (2021). Assemblages of fruit flies (Diptera: Tephritidae) along an elevational gradient in the rainforests of Papua New Guinea. *Insect Conserv. Diver.* 14, 348–355. doi: 10.1111/icad.12456

Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* 3, 294–299.

Fu, C., Hua, X., Li, J., Chang, Z., Pu, Z., and Chen, J. (2006). Elevational patterns of frog species richness and endemic richness in the Hengduan Mountains, China: geometric contraints, area and climate effects. *Ecography* 29, 919–927. doi: 10.1111/j.2006.0906-7590.04802.x

Fu, C. Z., Wang, J. X., Pu, Z. C., Zhang, S. L., Chen, H. L., Zhao, B., et al. (2007). Elevational gradients of diversity for lizards and snakes in the Hengduan Mountains, China. *Biodivers. Conserv.* 16, 707–726. doi: 10.1007/s10531-005-4382-4

Fu, H., Yuan, G. X., Jeppesen, E., Ge, D. B., Li, W., Zou, D. S., et al. (2019). Local and regional drivers of turnover and nestedness components of species and functional beta diversity in lake macrophyte communities in China. *Sci. Total Environ.* 687, 206–217. doi: 10.1016/j.scitotenv.2019.06.092

Geurts, K., Mwatawala, M., and De, M. M. (2012). Indigenous and invasive fruit fly diversity along an altitudinal transect in Eastern Central Tanzania. *J. Insect Sci.* 12:12. doi: 10.1673/031.012.1201

Gong, Z. D., Wu, H. Y., Duan, X. D., Feng, X. G., Zhang, Y. Z., and Li, Q. (2005). Species richness and vertical distribution pattern of flea fauna in Hengduan Mountains of western Yunnan, China. *Biodiver. Sci.* 13, 279–289. doi: 10.1360/biodiv.040177

Grinnell, J. (1914). Barriers to distribution as regards birds and mammals. *Am. Nat.* 48, 248–254.

Hebert, P. D., Cywinska, A., Ball, S. L., and deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proc. R. Soc. B Biol. Sci.* 270, 313–321. doi: 10.1098/rspb.2002.2218

Hendrich, L., Morinire, J., Haszprunar, G., Hebert, P. D. N., Hausmann, A., Khler, F., et al. (2015). A comprehensive DNA barcode database for Central European beetles with a focus on Germany: adding more than 3500 identified species to BOLD. *Mol. Ecol. Resour.* 15, 795–818. doi: 10.1111/1755-0998.12354

Huang, W. D., Xie, X. F., Huo, L. Z., Liang, X. Y., Wang, X. M., and Chen, X. S. (2020). An integrative DNA barcoding framework of ladybird beetles (Coleoptera: Coccinellidae). *Sci. Rep.* 10:10063. doi: 10.1038/s41598-020-66874-1

Jackson, J. M., Pimsler, M. L., Oyen, K. J., Koch-Uhuad, J. B., Herndon, J. D., Strange, J. P., et al. (2018). Distance, elevation and environment as drivers of diversity and divergence in bumble bees across latitude and altitude. *Mol. Ecol.* 27, 2926–2942. doi: 10.1111/mec.14735

Kaltsas, D., Dede, K., Giannaka, J., Nasopoulou, T., Kechagioglou, S., Grigoriadou, E., et al. (2018). Taxonomic and functional diversity of butterflies along an altitudinal gradient in two NATURA 2000 sites in Greece. *Insect Conserv. Diver.* 11, 464–478. doi: 10.1111/icad.12292

Kapli, P., Lutteropp, S., Zhang, J., Kobert, K., Pavlidis, P., Stamatakis, A., et al. (2017). Multi-rate poisson tree processes for single-locus species delimitation under maximum likelihood and markov chain Monte Carlo. *Bioinformatics* 33, 1630–1638. doi: 10.1093/bioinformatics/btx025

Kostál, V. (2006). Eco-physiological phases of insect diapause. *J. Insect Physiol.* 52, 113–127. doi: 10.1016/j.jinsphys.2005.09.008

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Lai, Y., and Liu, X. Y. (2020). The natural enemy species of Chrysopidae from China and their applications in biological control: a review. *J. Plant Protect.* 47, 1169–1187.

Li, X., Yuan, X. Z., and Deng, H. L. (2009). Vertical distribution and diversity of butterflies in Hengduan mountains, Southwest China. *Chin. J. Ecol.* 28, 1847–1852.

Liu, H., Ye, Q., and Wiens, J. J. (2020). Climatic-niche evolution follows similar rules in plants and animals. *Nat. Ecol. Evol.* 4, 753–763. doi: 10.1038/s41559-020-1158-x

Martins, C. C., Santos, R. S., Sutil, W. P., and de Oliveira, J. F. A. (2019). Diversity and abundance of green lacewings (Neuroptera: Chrysopidae) in a Conilon coffee plantation in Acre, Brazil. *Acta Amazon* 49, 173–178. doi: 10.1590/1809-4392201804470

Martoni, F., Simon, B., Andrew, P., Gary, T., and Karen, A. (2018). DNA barcoding highlights cryptic diversity in the New Zealand Psylloidea (Hemiptera: Sternorrhyncha). *Divers. Basel* 10:50. doi: 10.3390/d10030050

McCain, C. M. (2007). Could temperature and water availability drive elevational species richness patterns? A global case study for bats. *Glob. Ecol. Biogeogr.* 16, 1–13. doi: 10.1111/j.1466-8238.2006.00263.x

Michonneau, F. (2015). Cryptic and not-so-cryptic species in the complex "*Holothuria* (Thymiosycia) imaptiens" (Forsskål, 1775) (Echinodermata: Holothuroidea: Holothuriidae). *BioRxiv*[Preprint] BioRxiv 014225 doi: 10.1101/014225

Mochizuki, A., Naka, H., Hamasaki, K., and Mitsunaga, T. (2006). Larval cannibalism and intraguild predation between the introduced green lacewing, *Chrysoperla carnea*, and the indigenous trash-carrying green lacewing, *Mallada desjardinsi* (Neuroptera: Chrysopidae), as a case study of potential nontarget effect assessment. *Environ. Entomol.* 35, 1298–1303.

Monaghan, M. T., Wild, R., Elliot, M., Fujisawa, T., Balke, M., Inward, D. J. G., et al. (2009). Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Syst. Biol.* 58, 298–311. doi: 10.1093/sysbio/syp027

Morinière, J., Hendrich, L., Hausmann, A., Hebert, P., Haszprunar, G., and Gruppe, A. (2014). Barcoding fauna bavarica: 78% of the neuropterida fauna barcoded! *PLoS One* 9:e109719. doi: 10.1371/journal.pone.0109719

Negi, H. R., and Gadgil, M. (2002). Cross-taxon surrogacy of biodiversity in the Indian Garhwal Himalaya. *Biol. Conserv.* 105, 143–155. doi: 10.1016/S0006-3207(01)00158-6

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Noriega, J. A., and Realpe, E. (2018). Altitudinal turnover of species in a neotropical peripheral mountain system: a case study with dung beetles (Coleoptera: Aphodiinae and Scarabaeinae). *Environ. Entomol.* 47, 1376–1387. doi: 10.1093/ee/nvy133

Paknia, O., and Sh, H. R. (2015). Geographical patterns of species richness and beta diversity of Larentiinae moths (Lepidoptera: Geometridae) in two temperate biodiversity hotspots. *J. Insect Conserv.* 19, 729–739. doi: 10.1007/s10841-015-9795-0

Pappas, M. L., Broufas, G. D., and Koveos, D. S. (2011). Chrysopid predators and their role in biological control. *J. Entomol.* 8, 301–326. doi: 10.3923/je.2011.301.326

Park, D. S., Suh, S. J., Oh, H. W., and Hebert, P. D. N. (2010). Recovery of the mitochondrial COI barcode region in diverse Hexapoda through tRNA-based primers. *BMC Genomics* 11:423. doi: 10.1186/1471-2164-11-423

Perillo, L. N., de Castro, F. S., Solar, R., and de Siqueira Neves, F. (2021). Disentangling the effects of latitudinal and elevational gradients on bee, wasp, and ant diversity in an ancient neotropical mountain range. *J. Biogeogr.* 48, 1564–1578. doi: 10.1111/jbi.14095

Peters, M. K., Hemp, A., Appelhans, T., Behler, C., Classen, A., Detsch, F., et al. (2016). Predictors of elevational biodiversity gradients change from single taxa to the multi-taxa community level. *Nat. Commun.* 7:13736. doi: 10.1038/ncomms13736

Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., et al. (2006). Sequence based species delimitation for the DNA taxonomy of undescribed insects. *Syst. Biol.* 55, 595–609. doi: 10.1080/10635150600852011

Puillandre, N., Brouillet, S., and Achaz, G. (2021). ASAP: assemble species by automatic partitioning. *Mol. Ecol. Res.* 21, 609–620. doi: 10.1111/1755-0998.13281

Puillandre, N., Lambert, A., Brouillet, S., and Achaz, G. (2012). ABGD, automatic barcode gap discovery for primary species delimitation. *Mol. Ecol.* 21, 1864–1877. doi: 10.1111/j.1365-294X.2011.05239.x

Qian, H., Ricklefs, R. E., and White, P. S. (2005). Beta diversity of angiosperms in temperate floras of eastern Asia and eastern North America. *Ecol. Lett.* 8, 15–22. doi: 10.1111/j.1461-0248.2004.00682.x

Rahbek, C. (1995). The elevational gradient of species richness: a uniform pattern? *Ecography* 18, 200–205.

Rahbek, C. (2005). The role of spatial scale and the perception of large-scale species-richness patterns. *Ecol. Lett.* 8, 224–239. doi: 10.1111/j.1461-0248.2004.00701.x

Rambaut, A., and Drummond, A. J. (2016). *Figtree v1.4.3*. Available online at: http://tree.bio.ed.ac.uk/software/figtree/ (accessed December 2017).

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarisation in bayesian phylogenetics using tracer 1.7. *Syst. Biol.* 67, 901–904. doi: 10.1093/sysbio/syy032

Reymond, A., Purcell, J., Cherix, D., Guisan, A., and Pellissier, L. (2013). Functional diversity decreases with temperature in high elevation ant fauna. *Ecol. Entomol.* 38, 364–373. doi: 10.1111/een.12027

Ruiter, D. E., Boyle, E. E., and Zhou, X. (2013). DNA barcoding facilitates associations and diagnoses for Trichoptera larvae of the Churchill (Manitoba, Canada) area. *BMC Ecol.* 13:5. doi: 10.1186/1472-6785-13-5

Sam, K., Koane, B., and Novotny, V. (2015). Herbivore damage increases avian and ant predation of caterpillars on trees along a complete elevational forest gradient in Papua New Guinea. *Ecography* 38, 293–300. doi: 10.1111/ecog.00979

Smith, M. A., Fisher, B. L., and Hebert, P. D. N. (2005). DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philos. Trans. R. Soc. B* 360, 1825–1834. doi: 10.1098/rstb.2005.1714

Stanbrook, R., Wheater, C. P., Harris, W. E., and Jones, M. (2021). Habitat type and altitude work in tandem to drive the community structure of dung beetles in Afromontane forest. *J. Insect Conserv.* 25, 159–173. doi: 10.1007/s10841-020-00289-1

Sundqvist, M. K., Sanders, N. J., and Wardle, D. A. (2013). Community and ecosystem responses to elevational gradients: processes, mechanisms, and insights for global change. *Annu. Rev. Ecol. Evol. Syst.* 44, 261–280. doi: 10.1146/annurev-ecolsys-110512-135750

Supriya, K., Moreau, C. S., Sam, K., and Price, T. D. (2019). Analysis of tropical and temperate elevational gradients in arthropod abundance. *Front. Biogeogr.* 11:e43104. doi: 10.21425/F5FBG43104

Thierry, D., Deutsch, B., Paulian, M., Villenave, J., and Canard, M. (2005). Typifying ecosystems by using green lacewing assemblages. *Agron. Sustain. Dev.* 25, 473–479. doi: 10.1051/agro:2005047

Trifinopoulos, J., Nguyen, L. T., von Haeseler, A., and Minh, B. Q. (2016). W-IQ-TREE: a fast onlinephylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* 44, W232–W235. doi: 10.1093/nar/gkw256

Uhey, D., Haubensak, K., and Hofstetter, R. (2021). Mid-elevational peaks in diversity of ground-dwelling arthropods with high species turnover on the Colorado Plateau. *Environ. Entomol.* 50, 337–347. doi: 10.1093/ee/nvaa166

Ulrich, W., Almeida-Neto, M., and Gotelli, N. J. (2009). A consumer's guide to nestedness analysis. *Oikos* 118, 3–17. doi: 10.1111/j.1600-0706.2008.17053.x

Wang, M. Z., Li, Y. Y., Gao, F., and Zhang, L. S. (2019). Research advances in diapause of green lacewings (Neuroptera: Chrysopidae). *Chin. J. Biol. Control* 35, 474–486.

Wang, Y. H., Chen, B. G., and Su, Z. Y. (2000). Advance in species diversity research. *Ecol. Sci.* 19, 50–54.

Wen, Z. X., Quan, Q., Du, Y. B., Xia, L., Ge, D. Y., and Yang, Q. S. (2016). Dispersal, niche, and isolation processes jointly explain species turnover patterns of nonvolant small mammals in a large mountainous region of China. *Ecol. Evol.* 6, 946–960. doi: 10.1002/ece3.1962

Werenkraut, V., and Ruggiero, A. (2014). The richness and abundance of epigaeic mountain beetles in north-western Patagonia, Argentina: assessment of patterns and environmental correlates. *J. Biogeogr.* 41, 561–573. doi: 10.1111/jbi.12210

Woodcock, T. S., Boyle, E. E., Roughley, R. E., Kevan, P. G., Labbee, R. N., Smith, A. B. T., et al. (2013). The diversity and biogeography of the Coleoptera of Churchill: insights from DNA barcoding. *BMC Ecol.* 13:40. doi: 10.1186/1472-6785-13-40

Wu, Y. J., Colwell, R. K., Rahbek, C., Zhang, C. L., Quan, Q., Wang, C. K., et al. (2013). Explaining the species richness of birds along a subtropical elevational gradient in the Hengduan Mountains. *J. Biogeogr.* 40, 2310–2323. doi: 10.1111/jbi.12177

Xie, H. J., Liao, M. A., Jiang, G. L., Gao, Y., Chen, D., and Li, X. J. (2011). Speculation and developing present condition of apple industry in Sichuan Province. *North Horticul.* 35, 169–171.

Yang, X. K., Yang, J. K., and Li, W. Z. (2005). *Fauna Sinica, Insecta, Neuroptera, Chrysopidae*, Vol. 39. Beijing: Science Press.

Yang, Y. Y., Zhou, Y., Shi, Z., Viscarra Rossel, R. A., Liang, Z. Z., Wang, H. Z., et al. (2020). Interactive effects of elevation and land use on soil bacterial communities in the Tibetan Plateau. *Pedosphere* 30, 817–831. doi: 10.1016/S1002-0160(19)60836-2

Yao, Y. F., Song, X. Y., Wortley, A. H., Blackmore, S., and Li, C. S. (2015). A 22 570-year record of vegetational and climatic change from Wenhai Lake in the Hengduan Mountains biodiversity hotspot, Yunnan, Southwest China. *Biogeosciences* 12, 1525–1535. doi: 10.5194/bg-12-1525-2015

Ye, J. W., and Li, J. (2020). Factors affecting cannibalism by *Mallada basalis*. *Biocontrol Sci. Tech.* 30, 442–450. doi: 10.1080/09583157.2020.1729700

Yi, P., Yu, P., Liu, J. Y., Xu, H., and Liu, X. Y. (2018). A DNA barcode reference library of Neuroptera (Insecta, Neuropterida) from Beijing. *ZooKeys* 807, 127–147. doi: 10.3897/zookeys.807.29430

Young, M. R., Proctor, H. C., deWaard, J. R., and Hebert, P. D. N. (2019). DNA barcodes expose unexpected diversity in Canadian mites. *Mol. Ecol.* 28, 5347–5359. doi: 10.1111/mec.15292

Zhang, J. J., Kapli, P., Pavlidis, P., and Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* 29, 2869–2876. doi: 10.1093/bioinformatics/btt499

Zou, Y., Sang, W. G., Zhou, H. C., Huang, L. Y., and Axmacher, J. C. (2014). Altitudinal diversity patterns of ground beetles (Coleoptera: Carabidae) in the forests of Changbai Mountain, Northeast China. *Insect Conserv. Divers.* 7, 161–171. doi: 10.1111/icad.12039

Zou, Y., van der Werf, W., Liu, Y. H., and Axmacher, J. C. (2020). Predictability of species diversity by family diversity across global terrestrial animal taxa. *Glob. Ecol. Biogeogr.* 29, 629–644. doi: 10.1111/geb.13043

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 18 November 2021
doi: 10.3389/fevo.2021.786450

Specialization on *Ficus* Supported by Genetic Divergence and Morphometrics in Sympatric Host-Populations of the Camellia Aphid, *Aphis aurantii*

*Qiang Li[1], Cui Chen[1], Yangxue Wu[1], Junaid Ali Siddiqui[1], Congcong Lu[1], Zhentao Cheng[1], Yonghui Li[1], Qian Liu[1] and Xiaolei Huang[1,2]**

[1] State Key Laboratory of Ecological Pest Control for Fujian and Taiwan Crops, College of Plant Protection, Fujian Agriculture and Forestry University, Fuzhou, China, [2] Fujian Provincial Key Laboratory of Insect Ecology, Fujian Agriculture and Forestry University, Fuzhou, China

Adaptation to different host plants is considered to be an important driver of the divergence and speciation of herbivorous insects. The application of molecular data and integrated taxonomic practices in recent years may contribute to our understanding of population divergence and speciation, especially for herbivorous insects considered to be polyphagous. *Aphis aurantii* is an important agricultural and forestry pest with a broad range of host plants. In this study, samples of *A. aurantii* feeding on different host plants in the same geographical area were collected, and their population genetic divergence and morphological difference were analyzed. Phylogenetic analysis and haplotype network analysis based on five genes revealed that the population on *Ficus* exhibited significantly genetic divergence from populations on other host plants, which was also supported by the statistical analysis based on measurements of 38 morphological characters. Our results suggest that *A. aurantii* has undergone specialized evolution on *Ficus*, and the *Ficus* population may represent a lineage that is experiencing ongoing sympatric speciation.

Keywords: adaptation, host plant, population divergence, phylogeny, speciation

OPEN ACCESS

Edited by:
Xin Zhou,
China Agricultural University, China

Reviewed by:
Cheng-Min Shi,
Agricultural University of Hebei, China
Qiang Xie,
Sun Yat-sen University, China

****Correspondence:***
Xiaolei Huang
huangxl@fafu.edu.cn

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *30 September 2021*
Accepted: *21 October 2021*
Published: *18 November 2021*

Citation:
Li Q, Chen C, Wu Y, Siddiqui JA, Lu C, Cheng Z, Li Y, Liu Q and Huang X (2021) Specialization on Ficus Supported by Genetic Divergence and Morphometrics in Sympatric Host-Populations of the Camellia Aphid, Aphis aurantii. Front. Ecol. Evol. 9:786450. doi: 10.3389/fevo.2021.786450

INTRODUCTION

The mechanism of speciation has been a hot research topic in biology from Darwin's time to present day (Darwin, 1859; Futuyma and Mayer, 1980; McKinnon et al., 2004; Li et al., 2015; Taylor and Friesen, 2017). Whether speciation can occur without geographical barriers, i.e., sympatric speciation, is one of the core points of the debate. During much of the twentieth century, sympatric speciation was considered to be more unreliable when compared with allopatric speciation (Futuyma and Mayer, 1980). However, due to the in-depth research on biogeography and phylogeny in recent years, the concept of sympatric speciation has been accepted gradually (Via, 2001; Berlocher and Feder, 2002; Drès and Mallet, 2002; Bolnick and Fitzpatrick, 2007; Li et al., 2015). The growing acceptance of sympatric divergence and speciation has crucial implications for the interpretation of high biodiversity on Earth and the optimization of systematic theory and

practice (Berlocher and Feder, 2002). Compared with allopatric speciation, complete sympatric speciation events in nature may take a long time (Mallet, 2008), and much fewer empirical studies have been reported (Savolainen et al., 2006; Bolnick and Fitzpatrick, 2007). However, exploring divergence among sympatric populations, which may indicate ongoing sympatric speciation, can be helpful for understanding mechanisms of sympatric speciation (Drès and Mallet, 2002; Peccoud et al., 2009).

Phytophagous insects are considered as ideal candidates for the study of sympatric divergence due to intimate and specialized relationship with their host plants (Berlocher and Feder, 2002; Bolnick and Fitzpatrick, 2007; Peccoud et al., 2009; Lee et al., 2015). Differences in physical structure, nutritional composition, and chemical defense of different host plants may generate variant selection pressures on phytophagous insects that feeding on them (Egan and Ott, 2007). Moreover, the microenvironments provided by different host plants vary greatly, which may lead to different exposure probabilities to natural predators for phytophagous insects (Nosil, 2004; Nosil and Crespi, 2006; Rull et al., 2009). Therefore, for phytophagous insect populations in a sympatric area, long-term specialization on certain host plants may lead to adaptive evolution and reproductive isolation (Malausa et al., 2005; Xue et al., 2014; Lee et al., 2015). Host races of phytophagous insects are important evidence of sympatric genetic divergence driven by host plant (Peccoud et al., 2009).

Aphids exhibit varying degrees of host specialization. About half of all aphid species are specific to a single plant species, and at higher taxonomic levels, some aphid genera or families are strictly to a single plant genus or family (Eastop, 1973; Peccoud et al., 2010). There are also polyphagous aphid species in ecosystems, including many important agricultural pests such as *Aphis gossypii*, *Myzus persicae* and *Acyrthosiphon pisum*, having very high diversity of host plants (Blackman and Eastop, 2021). Host races or host-specialized populations with a relatively narrow host range are also frequently present in these polyphagous species (Via et al., 2000; Margaritopoulos et al., 2005; Carletto et al., 2009). This phenomenon indicates that these polyphagous species may have undergone population divergence or speciation events due to specialization on specific host plants (Peccoud et al., 2010).

Aphis aurantii (Hemiptera: Aphididae), known as the black citrus aphid or camellia aphid, is one of the most destructive pests of citrus and tea plants, mainly distributed in tropical and subtropical regions (Carver, 1978; Sevim et al., 2012; Blackman and Eastop, 2021). It is also a polyphagous species, which can feed on more than 120 plant species belonging to various families such as Rutaceae, Theaceae, Moraceae, Rosaceae, and Asteraceae (Blackman and Eastop, 2021). Although this aphid species can feed on phylogenetically and physiologically different host plants, at present there has been no report on host specificity or host races in it. However, previous studies discussed that the *A. aurantii* population on *Ficus* (Moraceae) exhibits some special features. Tao (1961) described *Toxoptera schlingeri* from *Ficus*, which was later considered as a synonym of *Aphis* (*Toxoptera*) *aurantii* by Raychaudhuri (1980) and Remaudière and Remaudiere (1997). Martin (1989) ever discussed that the validity of *T. schlingeri* might be supported by more in-depth study on the numbers and distribution of antennal rhinaria in more alatae samples. Qiao et al. (2008) found some morphological difference between *T. schlingeri* and *A. aurantii* specimens, but they suggested that morphological overlap would be found if more materials can be examined, and *T. schlingeri* was also regarded as a synonymy of *A. aurantii* in their paper. In subtropical and tropical areas of southern China, we observe that the morphology of *A. aurantii* varies across populations feeding on *Ficus* and other host plants, implying the *Ficus* population of *A. aurantii* may have undergone divergent evolution. Given the rapid advances in sequencing technology and new research methods such as DNA barcoding (Hebert et al., 2003; Foottit et al., 2008; Li et al., 2020) over the years, we think exploring the divergence of host-related populations of *A. aurantii* by integrating genetic evidence should be a worthwhile effort.

Considering *A. aurantii* and its host plants (especially *Ficus*) mainly distributed in subtropical and tropical areas (Volf et al., 2018; Blackman and Eastop, 2021), and with the aim to test sympatric population divergence of this species, our study took the subtropical Fujian province in southeastern China as target area. The specimens of *A. aurantii* were collected extensively to cover as many host plants as possible. Several molecular markers, including two mitochondrial genes (*COI*, cytochrome c oxidase subunit I; *Cytb*, cytochrome b), one nuclear gene (*EF-1α*, elongation factor-1α) and two genes (*gnd*, gluconate-6-phosphate dehydrogenase; *16S* rDNA) of *Buchnera*, the primary endosymbiont of aphids, were analyzed to explore the genetic structure of *A. aurantii* sympatric host-populations. We also undertook morphometrics of *A. aurantii* samples feeding on *Ficus* and other host plants to test population divergence in morphology.

MATERIALS AND METHODS

Specimen Sampling

A total of 48 *A. aurantii* specimens were collected from host plants of 11 families. The live morphology and habitats of *A. aurantii* in the field were photographed with digital cameras (Cannon EOS 7D plus Canon EF 100 mm f/2.8LMacro IS USM Lens). After recording the ecological information, aphid clones were stored in 95% ethanol and kept at −20°C for further morphological measurement and molecular experiments. All samples and voucher specimens were deposited in the Insect Systematics and Diversity Lab at Fujian Agriculture and Forestry University. Detailed information (host plant, voucher number, and GenBank accession number) of the specimens were listed in **Supplementary Table 1**.

DNA Extraction, PCR, and Sequencing

The genomic DNA of both aphids and *Buchnera* symbionts was extracted from each single specimen with the DNeasy Blood and Tissue Kit (QIAGEN, GERMANY). In order to obtain more accurate and comprehensive phylogenetic information, two mitochondrial genes (*COI*, *Cytb*), one nuclear gene (*EF-1α*) and two *Buchnera* genes (*gnd* and *16S* rDNA) were

amplified in this study. The primers for amplification of *COI* were LepF (5′-ATTCAACCAATCATAAAGATATTGG-3′) and LepR (5′-TAAACTTCTGGATGTCCAAAAAATCA-3′) (Foottit et al., 2008). *Cytb* sequences were amplified based on CP1 (5′-GATGATGAAATTGGATC-3′) and CP2 (5′-CTAATGCAATAACTCCTCC-3′) (Harry et al., 1998). The primers EF3 (5′-GAACGTGAACGTGGTATCAC-3′) and EF2 (5′-ATGTGAGCAGTGTGGCAATCCAA-3′) (von Dohlen et al., 2002) were used to amplify *EF-1α* sequences. *gnd* sequences were amplified based on *Bam*HI (5′-CGCGGATCCGGWCCWWSWATWATGCCWGGWGG-3′) and *Apa*I (5′-CGCGGGCCCGTATGWGCWCCAAAATAATCW CKTTGWGCTTG-3′) (Clark et al., 1999). The primers 16SA1 (5'-AGAGTTTGATCMTGGCTCAG-3′) and 16SB1 (5′-TACGGYTACCTTGTTACGACTT-3′) were used to amplify *16S* rDNA sequences (Weisburg et al., 1991).

PCR were performed in a final volume of 50 μl reaction mixture containing 28.5 μl dd H_2O, 8 μl dNTPs, 5 μl 10Xbuffer, 4 μl of template DNA, 2 μl of both forward and reverse primers (10 μM) and 0.5 μl of Taq DNA polymerase (5 U/μl). An initial denaturation step (95°C, 5 min) and final extension step (72°C, 10 min) were included in all polymerase chain reactions. The cycling conditions for *COI* were 35 cycles of 20 s at 94°C, 30 s at 50°C and 2 min at 72°C. The thermal setup for *Cytb* was 35 cycles of 1 min at 92°C, 1.5 min at 48°C and 1 min at 72°C. The cycling conditions of *EF-1α* included 35 cycles of denaturation at 95°C for 30 s, annealing at 51°C for 30 s and extension at 72°C for 1 min. The conditions for 35 cycles of *gnd* were 95°C for 20 s, 53°C for 30 s, and 72°C for 2 min. The PCR conditions for *16S* rDNA were according to the following procedure: 30 cycles at 94°C for 1 min; an annealing temperature of 50°C for 1 min; an extension at 72°C for 2 min. The PCR products were visualized by electrophoresis on a 1.5% agarose gel and then bidirectionally sequenced by Sangon Biotech (Shanghai). All sequences obtained in this study were uploaded to the GenBank, and the accession numbers were shown in **Supplementary Table 1**.

Based on the chromatograms, the raw sequences were corrected and assembled using BioEdit software (Hall, 1999). All sequences for each gene fragment were aligned using MAFFT (Kazutaka and Standley, 2013) and then verified manually. The introns of *EF-1α* sequences were removed based on GT-AG rule and the cDNA region of a reference sequence of *Schizaphis graminum* (GenBank accession number AF068479), and the coding regions of *EF-1α* were used in further phylogenetic analyses.

Genetic Distance and Phylogenetic Analysis

On the basis of current knowledge of the phylogenetic relationships within Aphididae, *A. gossypii* and *A. spiraecola* were chosen as outgroups for subsequent phylogenetic analyses. The *COI*, *Cytb*, *EF-1α* and *gnd* sequences of the outgroups were sequenced in our study, and the *16S* rDNA for their *Buchnera* symbionts were downloaded from the GenBank (*A. gossypii*: KC897373, KC897372; *A. spiraecola*: KC897427, KT175934). The MEGA 7.0 software (Kumar et al., 2016) was used to calculate pairwise distance among nucleotide sequences based on the Kimura 2-parameter (K2P) model (Kimura, 1980). We also downloaded the *COI* sequences of *Aphis* species from the BOLD database,[1] and then used SpeciesIdentifier software (Meier et al., 2006) to explore the distribution of intraspecific and interspecific genetic distances among *Aphis* species.

Phylogenetic analyses (Maximum likelihood, ML; Bayesian inference, BI) were performed based on three types of genes (mitochondrial: *COI* and *Cytb*; nuclear: *EF-1α*; *Buchnera*: *gnd* and *16S* rDNA), respectively. The appropriate nucleotide substitution models were selected based on Akaike Information Criterion (AIC) by using PartitionFinder 2 (Lanfear et al., 2016) and ModelFinder (Kalyaanamoorthy et al., 2017). The best-fit model for *COI* was GTR + I, for *Cytb* was GTR + I, for *EF-1α* was GTR + F, for *gnd* was GTR + G, for *16S* rDNA was GTR + I + G. RAxML (Stamatakis, 2014) was used to build the ML trees based on random starting trees with the GTRGAMMA substitution model and topological robustness was investigated using 1,000 non-parametric bootstrap replicates. Bayesian analyses were performed on all datasets using MrBayes 3.2.6 (Drummond et al., 2012). The combined dataset was divided into different gene partitions, and the best fitting models were assigned, respectively. For each dataset, two million generations MCMC (Markov Chain Monte Carlo) chains were run, with trees sampled every 100 generations. The first 5,000 trees (25%) for each dataset were discarded as burn-in to acquire posterior probability values (PP). The remaining trees were used to construct Bayesian consensus trees and viewed in iTOL (Letunic and Bork, 2016). In addition, the haplotypes based on the three gene datasets of different host-populations were analyzed using DnaSP 5.0 (Librado and Rozas, 2009). A median-joining network (MJ) was constructed using NETWORK 5.0.0.3 based on default setting (Bandelt et al., 1999).

Morphometry and Statistical Analysis

The samples used for morphometrics were collected from populations on eight genera of main host plants, including *Ficus*, *Calliandra*, *Camellia*, *Citrus*, *Loropetalum*, *Michelia*, *Morinda*, *Xylosma*. In principle, 3 clones of *A. aurantii* were selected from each genus of host plant, and then 10 adult apterous viviparous females were randomly selected from each aphid clone for morphological measurement. For some aphid clones with an insufficient number of adults, only those meeting above criteria were measured (**Supplementary Table 2**). All specimens were examined using Nikon SMZ18 stereomicroscope. A total of 20 morphological features were measured: body length (BL), body width (BW), length of 1st antennal segment (Ant1), length of 2nd antennal segment (Ant2), length of 3rd antennal segment (Ant3), hair length of 3rd antennal segment (Ant3_HL), width of 3rd antennal segment (Ant3_W), length of 4rth antennal segment (Ant4), length of 5th antennal segment (Ant5), base length of 6th antennal segment (Ant6_BL), processus terminalis of 6th antennal segment (Ant6_PT), whole antennal length (WA), hind femur (HF), siphunculi length (SIPH), basal width of siphunculi (SIPH_BW), distal width of siphunculi (SIPH_DW), cauda length (Cauda), basal width of cauda (Cauda_BW),

[1] www.boldsystems.org

length of dorsal hairs of tergite 1 (T1_DHL), hair length of tergite 8 (T8_HL).

We also calculated the ratios of WA/BL, Ant1/WA, Ant2/WA, Ant3/WA, Ant3_HL/WA, Ant3_W/WA, Ant4/WA, Ant5/WA, Ant6_BL/WA, Ant6_PT/WA, Ant3_HL/BL, Ant3_HL/Ant3, Ant3_HL/Ant3_W, Ant3_W/T1_DHL, Ant6_BL/Ant6_PT, T1_DHL/BL, SIPH/BL, SIPH/Cauda as supplementary morphological characters. The average as well as the minimum and maximum values of each morphological character for *A. aurantii* from different host plants were calculated separately (**Supplementary Table 2**). A one-way analysis of variance (ANOVA) was performed for 38 morphological characters to determine whether significant morphological difference among *A. aurantii* different host-populations can be found. In addition, *post hoc* multiple comparisons were performed based on LSD to detect the pairwise differences of each morphological feature between taxa. All statistical analyses were performed in SPSS ver. 24 (IBM, Chicago, IL, United States).

RESULTS

Sequence Features and Genetic Variation

Five gene fragments of most samples were successfully amplified. The 610 bp long *COI* alignment with 48 sequences included 595 conserved sites, 15 variable sites, and 15 parsimony-informative sites. The 45 *Cytb* sequences were trimmed to a 732 bp long alignment with 681 conserved sites, 51 variable sites, and 22 parsimony-informative sites. The 44 exon sequences of *EF-1α* were aligned to a final length of 712 bp, which included 703 conserved sites, 9 variable sites, and 5 parsimony-informative sites. A total of 45 *gnd* sequences (821 bp, conserved sites: 765; variable sites: 56; parsimony-informative sites: 54) and 46 *16S* rDNA sequences (337 bp, conserved sites: 329; variable sites: 8; parsimony-informative sites: 4) were successfully generated. The nucleotide composition of mitochondrial gene (*COI* and *Cytb*) and *gnd* fragments showed a strong bias toward A + T content (76, 77.5, and 75.3%, respectively), while *EF-1α* and *16S* have no similar bias (**Supplementary Table 3**).

The sequences of each gene were divided into two groups, group 1 feeding on nearly 20 other host plants and group 2 feeding on *Ficus*, then the genetic distances within and between groups of different genes were calculated, respectively (**Supplementary Table 3**). Using the *COI* gene as an example, we found that the genetic distance of samples between groups was much larger than that within groups. The genetic distance range of samples on *Ficus* was 0–0%, and that of samples feeding on other host plant was 0–0.8%. However, the genetic distances between samples from *Ficus* and other host plants could reach as high as 1.8%. Within group 1, the Theaceae population contributed the largest genetic distance (0.8%) with other host-populations (**Supplementary Table 3**).

The COI sequences of Aphis downloaded from BOLD database were collated and corrected, and 3,429 sequences of 99 species were obtained. The analysis of distribution of the intraspecific and interspecific genetic distances of Aphis species showed an obvious barcoding gap (**Supplementary Figure 1**). For the intraspecific distances, 97.56% of them were less than 1%, 98.04% less than 1.5, and 99.1% less than 2%. Besides, 99.18 and 98.68% of the interspecific distances were greater than 2 and 2.5% (**Supplementary Table 4**).

Phylogeny and Haplotype Network

The phylogenetic trees that inferred from mitochondrial (*COI* and *Cytb*) and *Buchnera* (*gnd* and *16S*) genes showed that sympatric host-populations of *A. aurantii* were divided into two well-supported clades (**Figures 1A,B**), corresponding to the populations feeding on *Ficus* (G2) and the other plants (G1), respectively. All the samples feeding on *Ficus* were clustered into a separate clade at the base of the phylogenetic tree. In addition, some populations of *A. aurantii* feeding on Theaceae in the G1 clade also showed relatively obvious divergence (**Figures 1A,B**). The nuclear gene (*EF-1α*), which was most conserved among the five gene markers, however, showed a less unambiguous phylogenetic pattern (**Figure 1C**).

A total of 16 haplotypes were identified among the 45 mitochondrial (*COI* and *Cytb*) sequences (**Figure 2A**). Haplotype H1 contains the most samples and host plant families. The most frequently observed host plant was Theaceae and appeared in several haplotypes (H1, H5, H6, H8, H12, H13, H14, H15). All the samples feeding on *Ficus* were assigned as haplotype H4, which showed greatest differentiation from other haplotypes. Similarly, in the haplotype network analyses of nuclear and *Buchnera* genes, there were significant genetic differences among the populations feeding on *Ficus* and other host plants (**Figures 2B,C**).

Morphology and Statistics

The results of the ANOVA showed that most morphological characters exhibited significant difference between different host plant genera ($P < 0.05$), except for the length of Ant1, Ant2, Ant4, T8_HL and ratio of WA/BL, Ant6_BL/WA, Ant6_BL/Ant6_PT (**Supplementary Table 5**). With integration of the results of *post hoc* analysis by LSD, morphological characters having significant difference between populations on *Ficus* and other host plant genera were determined according to a rule that there should have samples from more than four different plant genera showing significant difference with the *Ficus* samples. Our results indicated that there were significant difference in eighteen morphological characters between *A. aurantii* populations on *Ficus* and other plant genera, including length of Ant3_HL, Ant5, Ant6_PT, WA, SIPH_DW, Cauda, T1_DHL; ratio of WA/BL, Ant3/WA, Ant3_HL/WA, Ant3_W/WA, Ant5/WA, Ant3_HL/Ant3, Ant3_HL/Ant3_W, Ant3_HL/BL, Ant3_W/T1_DHL, SIPH/Cauda, T1_DHL/BL (**Figure 3** and **Supplementary Table 5**). Among these morphological characters, Ant3_HL, T1_DHL, Ant3_HL/WA, Ant3_HL/Ant3_W, Ant3_HL/BL, Ant3_W/T1_DHL, T1_DHL/BL of the *Ficus* population showed significant difference with samples from all other plant genera. That is to say, at least seven morphological characters can be used to distinguish the *A. aurantii* population feeding on *Ficus* from those feeding on other host plants.

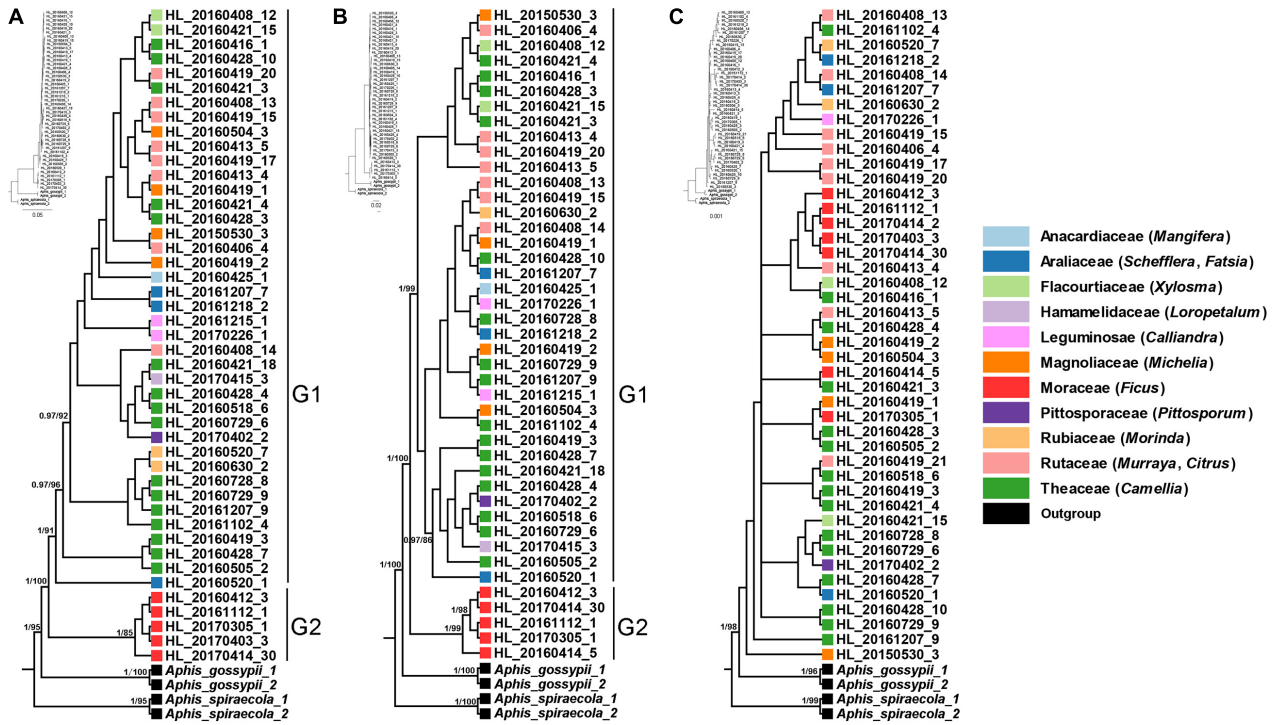


FIGURE 1 | Phylogeny of *Aphis aurantii* populations inferred from Bayesian analyses (BI) based on *COI* + *Cytb* **(A)**, *gnd* and *16S* **(B)** and *EF-1α* **(C)**. Different color represent samples from different host plant families. The bootstrap values (>80) and posterior probabilities (>0.9) are shown for main nodes.

DISCUSSION

Host plants are considerable source of selective pressure for their associated aphids, constituting their only food resource, habitat, mating and oviposition sites (Peccoud and Simon, 2010; Peccoud et al., 2010; Blackman and Eastop, 2021). In the past three decades, the pea aphid *A. pisum* has been considered as a good model to investigate host specialization and sympatric speciation (Via, 1991a,b). In Western Europe, at least eight conspecific host races were found on different host plants, which showed that adaptation to different host plants could indeed play a very important role in sympatric speciation of insects (Peccoud et al., 2009). In the present study, by using the polyphagous *A. aurantii* as a model, we tested whether host-related divergence exist among sympatric populations. Our results based on mitochondrial (*COI* and *Cytb*) and *Buchnera* (*gnd* and *16S*) genes clearly showed that all samples feeding on *Ficus* were clustered in a separate clade with relatively high divergence (**Figure 1**) and formed a unique haplotype (**Figure 2**), indicating that the *Ficus* population of *A. aurantii* have experienced a host specialization process. In addition, we also noted that the nuclear gene (*EF-1α*) did not show a clear evolutionary relationship. Considering *EF-1α* is a relatively conserved nuclear gene (Cho et al., 1995), we think the lack of accumulation of sufficient genetic variation at the early stage of species divergence may be an important factor leading to unclear phylogenetic pattern of this gene. The conflict of phylogenetic signals between different types of genes, in fact, indicates they these genes may undergo different histories of lineage sorting in the process of species divergence.

In previous DNA barcoding studies, *COI* genetic distance thresholds were used to determine different animal species (Hajibabaei et al., 2006; Zhou et al., 2010; Schmidt et al., 2015). Although the genetic distance thresholds vary slightly among different taxonomic groups, a threshold of 2–2.5% is generally accepted by aphid taxonomists (Liu et al., 2013; Zhu et al., 2017; Li et al., 2020). Our analysis of distribution of the intraspecific and interspecific genetic distances of *Aphis* species (**Supplementary Figure 1** and **Supplementary Table 4**) also confirmed this threshold. For most *Aphis* species, 2–2.5% can be used as a reasonable genetic distance threshold to distinguish different species. In the present study, the maximum *COI* genetic distance between the G1 and G2 clades (**Supplementary Table 3**) reached 1.8%, which exceeds about 99% of the intraspecific genetic difference of the *Aphis* species dataset, indicating a relatively deep divergence of the *Aphis aurantii* lineage on *Ficus*.

The ecological adaptation of phytophagous insects may first occur at the physiological and genetic levels, and is sometimes relatively unobvious in phenotype (Doolittle and Sapienza, 1980; Kearney and Porter, 2009). Since identification of species in traditional taxonomy is mainly based on morphological characters, this phenomenon of insignificant phenotypic information might cause troubles for taxonomy, especially when specimens and characters are insufficiently sampled. Morphological comparison of different host-populations of *A. aurantii* showed that among

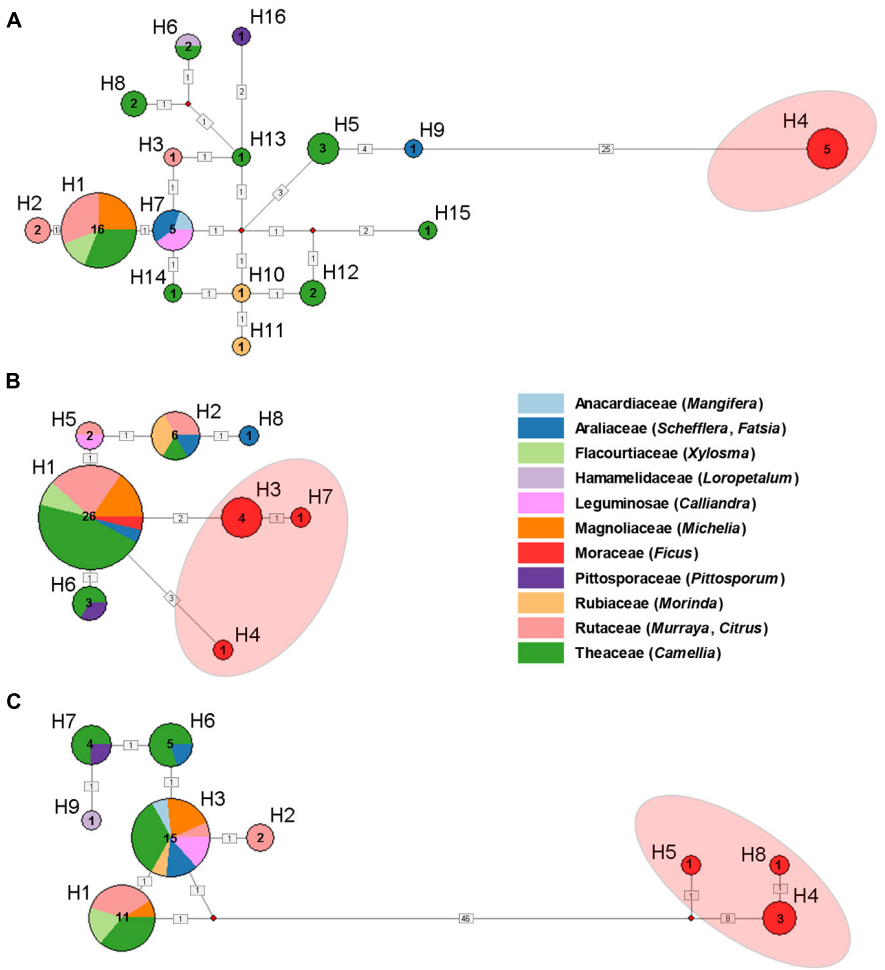


FIGURE 2 | Haplotype network of *Aphis aurantii* populations based on *COI* + *Cytb* **(A)**, EF-1α **(B)** and *gnd* and *16S* **(C)**. Numbers near the circle indicate haplotype numbers. Sample numbers and host plants (in different color) of haplotypes are annotated in the circles. The numbers in the black square indicate number of mutations. The red shading on the haplotype network represents *Ficus* samples.

the 38 characters, more characters were relatively similar. For instances, the body length, body width, hind femur and siphunculi did not show significant difference. Tao (1961) described *T. schlingeri* just based on a single sample collected from *Ficus* in Hong Kong. Qiao et al. (2008) compared the morphology of *T. schlingeri* and *A. aurantii* with more apterous viviparous female specimens, and found only the length of marginal hairs on abdominal tergite I and the widest diameter of antennal segment III were different. They suggested that further examination and thorough research were needed to determine the validity of *T. schlingeri*. Our measurement results of the length of Ant3_HL and T1_DHL were similar to those of Qiao et al. (2008), in addition, we also found that ratios of Ant3_HL/WA, Ant3_HL/Ant3_W, Ant3_HL/BL, Ant3_W/T1_DHL, T1_DHL/BL of the *Ficus* population of *A. aurantii* were significantly different from that of the populations on other host plants. By integrating the genetic divergence and morphological difference of *A. aurantii* on *Ficus* and other plants, and our field observations on this species, we think a reasonable explanation is that the specimen used to define *T. schlingeri* might be actually an individual from the population of *A. aurantii* specialized on *Ficus*. And the *Ficus* population of *A. aurantii* may represent a lineage that is experiencing ongoing sympatric speciation.

The physical and chemical characters of host plants are important factors that lead to the specialization of phytophagous insects. For example, the tobacco biotype of *M. persicae* can respond specifically to the volatile released by tobacco, whereas other biotypes do not, suggesting that difference in olfactory perception exists among different biotypes (Vargas et al., 2005). Furthermore, gustatory reception also affects the aphids' selection of host plants, as aphids probe the phloem for nutritional composition and chemical information before deciding whether to feed (Caillaud and Via, 2000; Powell et al., 2006). *Ficus* is a special plant group. It can produce a broad range of chemical and physical defenses to against phytophagous insects (Cruaud et al., 2012). These include the production of latex, polyphenols and terpenoids. *Ficus* can produce some special

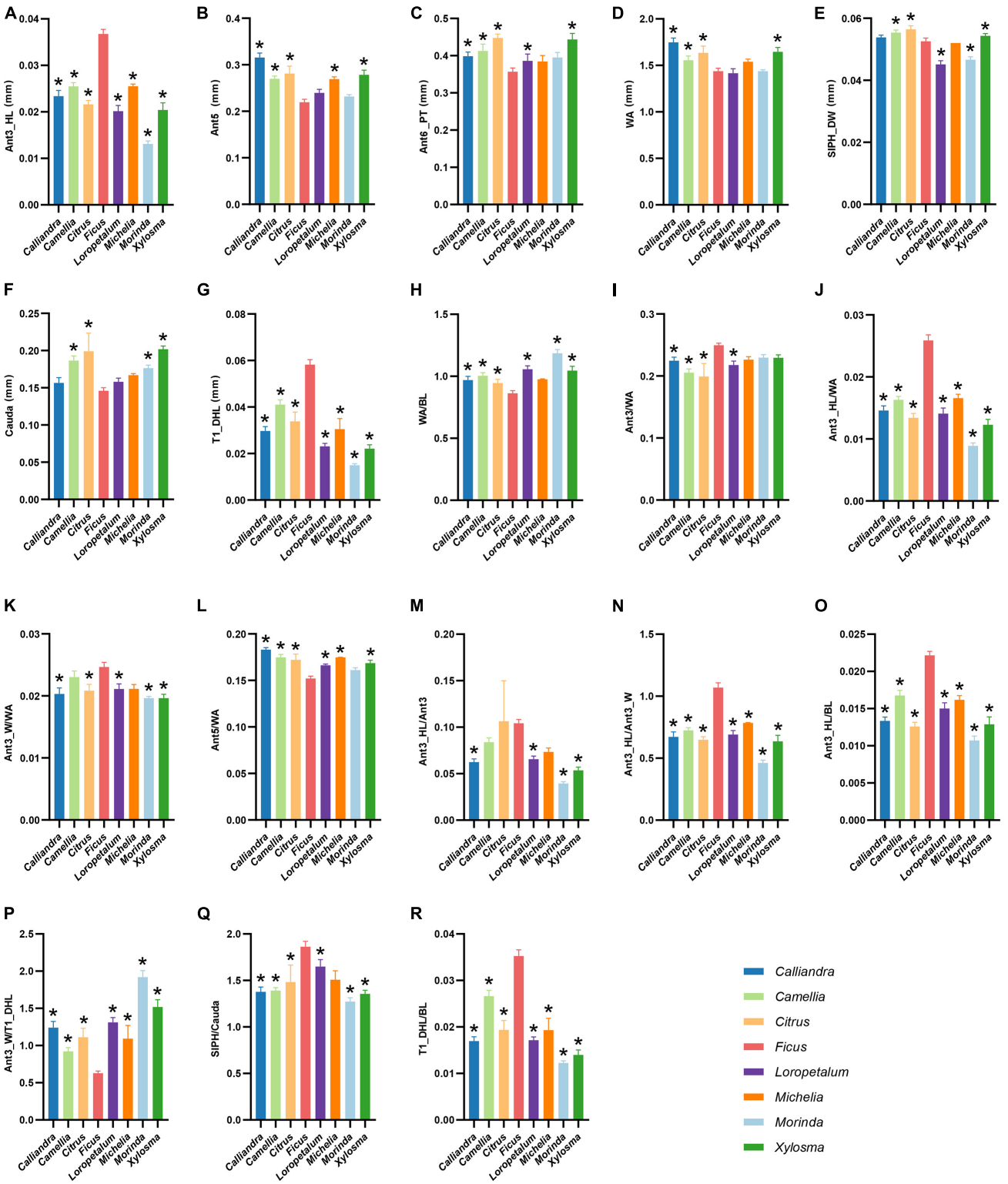


FIGURE 3 | Eighteen morphological characters with significant difference between *A. aurantii* populations feeding on *Ficus* and other plant genera, including Ant3_HL **(A)**, Ant5 **(B)**, Ant6_PT **(C)**, WA **(D)**, SIPH_DW **(E)**, Cauda **(F)**, T1_DHL **(G)**; ratios of WA/BL **(H)**, Ant3/WA **(I)**, Ant3_HL/WA **(J)**, Ant3_W/WA **(K)**, Ant5/WA **(L)**, Ant3_HL/Ant3 **(M)**, Ant3_HL/Ant3_W **(N)**, Ant3_HL/BL **(O)**, Ant3_W/T1_DHL **(P)**, SIPH/Cauda **(Q)**, T1_DHL/BL **(R)**. Different color bars represent samples from different host plant genera. *Represents host-populations significantly different with the *Ficus* population.

chemical substances, including phenanthroindolizidine alkaloids (Damu et al., 2005) and cysteine proteases, which can interfere the digestive function of insects and increase the mortality of them (Konno et al., 2004). Considering *A. aurantii* shows obvious host specialization on *Ficus*, it can be a good model for investigating the mechanism of host specialization. Further studies on comparative transcriptomics among different host populations and detailed screening of symbiotic bacteria are extremely needed.

DATA AVAILABILITY STATEMENT

The sequence data are publicly archived in the GenBank, and the accession numbers are provided in **Supplementary Table 1**.

AUTHOR CONTRIBUTIONS

XH conceived and designed the study, reviewed and edited the manuscript, and contributed resources during the study. QLi, CC, JAS, YL, and QLiu performed the experiments. QLi, CC, YW, CL, and ZC analyzed the data. QLi, CC, CL, and ZC wrote the original draft. All authors have read and agreed to the published version of the manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (Grant No. 31772504) and the Fujian Provincial Department of Science and Technology (Grant No. 2015J06005).

ACKNOWLEDGMENTS

We would like to thank Xiaolan Lin and Lingda Zeng for help in sample collection and molecular experiment, respectively.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2021.786450/full#supplementary-material

Supplementary Figure 1 | Frequency histogram of intra- and inter-specific genetic distances of *Aphis COI* sequences downloaded from the BOLD database.

Supplementary Table 1 | Sample information including voucher number, host plant family and species, and GenBank accession number.

Supplementary Table 2 | Morphometrics of *A. aurantii* samples collected on different host plant genera.

Supplementary Table 3 | The nucleotide composition and genetic distances of *A. aurantii* feeding on different host plants.

Supplementary Table 4 | The distribution of intraspecific and interspecific distances among *Aphis COI* sequences.

Supplementary Table 5 | Result of one-way ANOVA and *post hoc* LSD for morphological characters of *A. aurantii* samples collected on different host plant genera. *Calliandra* (n = 15), *Camellia* (n = 30), *Citrus* (n = 28), *Ficus* (n = 30), *Loropetalum* (n = 20), *Michelia* (n = 20), *Morinda* (n = 10), *Xylosma* (n = 10).

REFERENCES

Bandelt, H.-J., Forster, P., and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16, 37–48. doi: 10.1093/oxfordjournals.molbev.a026036

Berlocher, S. H., and Feder, J. L. (2002). Sympatric speciation in phytophagous insects: moving beyond controversy? *Annu. Rev. Entomol.* 47, 773–815. doi: 10.1146/annurev.ento.47.091201.145312

Blackman, R. L., and Eastop, V. F. (2021). *Aphids on the World's Plants: an online identification and information guide*. Available Online at: http://www.aphidsonworldsplants.info/ (accessed June 23, 2021)

Bolnick, D. I., and Fitzpatrick, B. M. (2007). Sympatric speciation: models and empirical evidence. *Annu. Rev. Ecol. Evol. Syst.* 38, 459–487. doi: 10.1146/annurev.ecolsys.38.091206.095804

Caillaud, M. C., and Via, S. (2000). Specialized feeding behavior influences both ecological specialization and assortative mating in sympatric host races of pea aphids. *Am. Nat.* 156, 606–621. doi: 10.1086/316991

Carletto, J., Lombaert, E., Chavigny, P., Brévault, T., Lapchin, L., and Vanlerberghe Masutti, F. (2009). Ecological specialization of the aphid *Aphis gossypii* Glover on cultivated host plants. *Mol. Ecol.* 18, 2198–2212. doi: 10.1111/j.1365-294X.2009.04190.x

Carver, M. (1978). The black citrus aphids, *Toxoptera citricidus* (Kirkaldy) and *T. aurantii* (Boyer de Fonscolombe) (*Homoptera*: *aphididae*). *Aust. J. Entomol.* 17, 263–270. doi: 10.1111/j.1440-6055.1978.tb00156.x

Cho, S., Mitchell, A., Regier, J. C., Mitter, C., Poole, R. W., Friedlander, T. P., et al. (1995). A highly conserved nuclear gene for low-level phylogenetics: elongation factor-1 alpha recovers morphology-based tree for heliothine moths. *Mol. Biol. Evol.* 12, 650–656. doi: 10.1093/oxfordjournals.molbev.a040244

Clark, M. A., Moran, N. A., and Baumann, P. (1999). Sequence evolution in bacterial endosymbionts having extreme base compositions. *Mol. Biol. Evol.* 16, 1586–1598. doi: 10.1093/oxfordjournals.molbev.a026071

Cruaud, A., Rønsted, N., Chantarasuwan, B., Chou, L. S., Clement, W. L., Couloux, A., et al. (2012). An extreme case of plant-insect codiversification: figs and fig-pollinating wasps. *Syst. Biol.* 61, 1029–1047. doi: 10.1093/sysbio/sys068

Damu, A. G., Kuo, P. C., Shi, L. S., Li, C. Y., Kuoh, C. S., Wu, P. L., et al. (2005). Phenanthroindolizidine alkaloids from the stems of *Ficus septica*. *J. Nat. Prod.* 68, 1071–1075. doi: 10.1021/np050095o

Darwin, C. R. (1859). *The origin of species by means of natural selection: or, the preservation of favored races in the struggle for life*. London, UK: John Murray.

Doolittle, W. F., and Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284, 601–603. doi: 10.1038/284601a0

Drès, M., and Mallet, J. (2002). Host races in plant-feeding insects and their importance in sympatric speciation. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 357, 471–492. doi: 10.1098/rstb.2002.1059

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075

Eastop, V. (1973). "Deductions from the present day host plants of aphids and related insects," in *Insect/Plant relationships*, ed. H. F. V. Emden (London: Symposia of the Royal Entomological Society of London), 157–178.

Egan, S. P., and Ott, J. R. (2007). Host plant quality and local adaptation determine the distribution of a gall-forming herbivore. *Ecology* 88, 2868–2879. doi: 10.1890/06-1303.1

Foottit, R. G., Maw, H. E., Von Dohlen, C. D., and Hebert, P. D. (2008). Species identification of aphids (*Insecta*: *hemiptera*: *aphididae*) through DNA barcodes. *Mol. Ecol. Resour.* 8, 1189–1201. doi: 10.1111/j.1755-0998.2008.02297.x

Futuyma, D. J., and Mayer, G. C. (1980). Non-allopatric speciation in animals. *Syst. Biol.* 29, 254–271. doi: 10.1093/sysbio/29.3.254

Hajibabaei, M., Janzen, D. H., Burns, J. M., Hallwachs, W., and Hebert, P. D. N. (2006). DNA barcodes distinguish species of tropical Lepidoptera. *Proc. Natl. Acad. Sci. U. S. A.* 103, 968–971. doi: 10.1073/pnas.0510466103

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.

Harry, M., Solignac, M., and Lachaise, D. (1998). Molecular evidence for parallel evolution of adaptive syndromes in fig-breeding *Lissocephala* (*Drosophilidae*). *Mol. Phylogenet. Evol.* 9, 542–551. doi: 10.1006/mpev.1998.0508

Hebert, P. D., Ratnasingham, S., and De Waard, J. R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 270, S96–S99. doi: 10.1098/rsbl.2003.0025

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285

Kazutaka, K., and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: improvements in Performance and Usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kearney, M., and Porter, W. (2009). Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecol. Lett.* 12, 334–350. doi: 10.1111/j.1461-0248.2008.01277.x

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120. doi: 10.1007/BF01731581

Konno, K., Hirayama, C., Nakamura, M., Tateishi, K., Tamura, Y., Hattori, M., et al. (2004). Papain protects papaya trees from herbivorous insects: role of cysteine proteases in latex. *Plant J.* 37, 370–378. doi: 10.1046/j.1365-313X.2003.01968.x

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., and Calcott, B. (2016). PartitionFinder 2: new methods for selecting partitioned models of evolution formolecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34, 772–773. doi: 10.1093/molbev/msw260

Lee, Y., Lee, W., Lee, S., and Kim, H. (2015). A cryptic species of *Aphis gossypii* (*Hemiptera*: *aphididae*) complex revealed by genetic divergence and different host plant association. *Bull. Entomol. Res.* 105, 40–51. doi: 10.1017/S0007485314000704

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242–W245. doi: 10.1093/nar/gkw290

Li, K., Hong, W., Jiao, H., Wang, G.-D., Rodriguez, K. A., Buffenstein, R., et al. (2015). Sympatric speciation revealed by genome-wide divergence in the blind mole rat *Spalax*. *Proc. Natl. Acad. Sci. U. S. A.* 112, 11905–11910. doi: 10.1073/pnas.1514896112

Li, Q., Deng, J., Chen, C., Zeng, L., Lin, X., Cheng, Z., et al. (2020). DNA Barcoding Subtropical Aphids and Implications for Population Differentiation. *Insects* 11:11. doi: 10.3390/insects11010011

Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187

Liu, Q. H., Jiang, L. Y., and Qiao, G. X. (2013). DNA barcoding of Greenideinae (*Hemiptera*: *Aphididae*) with resolving taxonomy problems. *Invertebr. Syst.* 27, 428–438. doi: 10.1071/IS13014

Malausa, T., Bethenod, M.-T., Bontemps, A., Bourguet, D., Cornuet, J.-M., and Ponsard, S. (2005). Assortative mating in sympatric host races of the European corn borer. *Science* 308, 258–260. doi: 10.1126/science.1107577

Mallet, J. (2008). Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 2971–2986. doi: 10.1098/rstb.2008.0081

Margaritopoulos, J., Tsourapas, C., Tzortzi, M., Kanavaki, O., and Tsitsipis, J. (2005). Host selection by winged colonisers within the *Myzus persicae* group: a contribution towards understanding host specialisation. *Ecol. Entomol.* 30, 406–418. doi: 10.1111/j.0307-6946.2005.00700.x

Martin, J. (1989). Identification, occurrence and pest status of *Toxoptera odinae* (van der Goot) (*Hemiptera*: *aphididae*) in Africa. *Bull. Entomol. Res.* 79, 607–611. doi: 10.1017/S0007485300018757

McKinnon, J. S., Mori, S., Blackman, B. K., David, L., Kingsley, D. M., Jamieson, L., et al. (2004). Evidence for ecology's role in speciation. *Nature* 429, 294–298. doi: 10.1038/nature02556

Meier, R., Shiyang, K., Vaidya, G., and Ng, P. K. (2006). DNA barcoding and taxonomy in *Diptera*: a tale of high intraspecific variability and low identification success. *Syst. Biol.* 55, 715–728. doi: 10.1080/10635150600969864

Nosil, P. (2004). Reproductive isolation caused by visual predation on migrants between divergent environments. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 271, 1521–1528. doi: 10.1098/rspb.2004.2751

Nosil, P., and Crespi, B. J. (2006). Experimental evidence that predation promotes divergence in adaptive radiation. *Proc. Natl. Acad. Sci. U. S. A.* 103, 9090–9095. doi: 10.1073/pnas.0601575103

Peccoud, J., Ollivier, A., Plantegenest, M., and Simon, J. C. (2009). A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proc. Natl. Acad. Sci. U. S. A.* 106, 7495–7500. doi: 10.1073/pnas.0811117106

Peccoud, J., and Simon, J. C. (2010). The pea aphid complex as a model of ecological speciation. *Ecol. Entomol.* 35, 119–130. doi: 10.1111/j.1365-2311.2009.01147.x

Peccoud, J., Simon, J. C., Von Dohlen, C., Coeur D'acier, A., Plantegenest, M., Vanlerberghe Masutti, F., et al. (2010). Evolutionary history of aphid-plant associations and their role in aphid diversification. *C. R. Biol.* 333, 474–487. doi: 10.1016/j.crvi.2010.03.004

Powell, G., Tosh, C. R., and Hardie, J. (2006). Host plant selection by aphids: behavioral, evolutionary, and applied perspectives. *Annu. Rev. Entomol.* 51, 309–330. doi: 10.1146/annurev.ento.51.110104.151107

Qiao, G., Wang, J., and Zhang, G. (2008). *Toxoptera* Koch (*Hemiptera*: *aphididae*), a generic account, description of a new species from China, and keys to species. *Zootaxa* 1746, 1–14. doi: 10.11646/zootaxa.1746.1.1

Raychaudhuri, D. N. (1980). *Aphids of North-East India and Bhutan*. Calcutta: The Zoological Society.

Remaudière, G., and Remaudiere, M. (1997). *Catalogue of the world's Aphididae: homoptera Aphidoidea*. Paris: Institut National de la Recherche Agronomique (INRA).

Rull, J., Wharton, R., Feder, J. L., Guillén, L., Sivinski, J., Forbes, A., et al. (2009). Latitudinal variation in parasitoid guild composition and parasitism rates of North American hawthorn infesting *Rhagoletis*. *Environ. Entomol.* 38, 588–599. doi: 10.1603/022.038.0310

Savolainen, V., Anstett, M. C., Lexer, C., Hutton, I., Clarkson, J. J., Norup, M. V., et al. (2006). Sympatric speciation in palms on an oceanic island. *Nature* 441, 210–213. doi: 10.1038/nature04566

Schmidt, S., Schmid-Egger, C., Morinière, J., Haszprunar, G., and Hebert, P. D. (2015). DNA barcoding largely supports 250 years of classical taxonomy: identifications for Central European bees (*Hymenoptera*, *Apoidea partim*). *Mol. Ecol. Resour.* 15, 985–1000. doi: 10.1111/1755-0998.12363

Sevim, E., Çelebi, Ö, and Sevim, A. (2012). Determination of the bacterial flora as a microbial control agent of *Toxoptera aurantii* (*Homoptera*: *aphididae*). *Biologia* 67, 397–404. doi: 10.2478/s11756-012-0022-0

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Tao, C. C. (1961). Revision of the genus *Toxoptera* Koch, 1856 (*Homoptera*: *aphididae*). *Q. J. Taiwan Mus.* 14, 257–260.

Taylor, R. S., and Friesen, V. L. (2017). The role of allochrony in speciation. *Mol. Ecol.* 26, 3330–3342. doi: 10.1111/mec.14126

Vargas, R. R., Troncoso, A. J., Tapia, D. H., Olivares-Donoso, R., and Niemeyer, H. M. (2005). Behavioural differences during host selection between alate virginoparae of generalist and tobacco-specialist *Myzus persicae*. *Entomol. Exp. Appl.* 116, 43–53. doi: 10.1111/j.1570-7458.2005.00311.x

Via, S. (1991a). Specialized host plant performance of pea aphid clones is not altered by experience. *Ecology* 72, 1420–1427. doi: 10.2307/1941114

Via, S. (1991b). The genetic structure of host plant adaptation in a spatial patchwork: demographic variability among reciprocally transplanted pea aphid clones. *Evolution* 45, 827–852. doi: 10.1111/j.1558-5646.1991.tb04353.x

Via, S. (2001). Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol. Evol.* 16, 381–390. doi: 10.1016/S0169-5347(01)02188-7

Via, S., Bouck, A. C., and Skillman, S. (2000). Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and

hybrids in the parental environments. *Evolution* 54, 1626–1637. doi: 10.1111/j.0014-3820.2000.tb00707.x

Volf, M., Segar, S. T., Miller, S. E., Isua, B., Sisol, M., Aubona, G., et al. (2018). Community structure of insect herbivores is driven by conservatism, escalation and divergence of defensive traits in *Ficus*. *Ecol. Lett.* 21, 83–92. doi: 10.1111/ele.12875

von Dohlen, C. D., Kurosu, U., and Aoki, S. (2002). Phylogenetics and evolution of the eastern Asian-eastern North American disjunct aphid tribe, *Hormaphidini* (*Hemiptera*: *aphididae*). *Mol. Phylogenet. Evol.* 23, 257–267. doi: 10.1016/S1055-7903(02)00025-8

Weisburg, W. G., Barns, S. M., Pelletier, D. A., and Lane, D. J. (1991). 16S ribosomal DNA amplification for phylogenetic study. *J. Bacteriol.* 173, 697–703. doi: 10.1128/jb.173.2.697-703.1991

Xue, H.-J., Li, W.-Z., and Yang, X.-K. (2014). Assortative mating between two sympatric closely-related specialists: inferred from molecular phylogenetic analysis and behavioral data. *Sci. Rep.* 4:5436. doi: 10.1038/srep05436

Zhou, X., Jacobus, L. M., Dewalt, R. E., Adamowicz, S. J., and Hebert, P. D. (2010). *Ephemeroptera*, *Plecoptera*, and *Trichoptera* fauna of Churchill (Manitoba, Canada): insights into biodiversity patterns from DNA barcoding. *J. N. Am. Benthol. Soc.* 29, 814–837. doi: 10.1899/09-121.1

Zhu, X. C., Chen, J., Chen, R., Jiang, L. Y., and Qiao, G. X. (2017). DNA barcoding and species delimitation of Chaitophorinae (*Hemiptera*, *Aphididae*). *Zookeys* 656, 25–50. doi: 10.3897/zookeys.656.11440

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 10 December 2021
doi: 10.3389/fevo.2021.771719

Flight Mill Experiments and Computer Simulations Indicate Islands Recruit More Capable Flyers of Moths

Yu-xuan Zheng[1†], Ying Wang[1†], Bo-ya Dai[1], Zheng Li[2], Qi-run Huo[2], Jian-xin Cui[3], Hao Liu[3], Xin-hai Li[4], Alice C. Hughes[5] and Ai-bing Zhang[1]*

[1] College of Life Sciences, Capital Normal University, Beijing, China, [2] College of Information Engineering, Capital Normal University, Beijing, China, [3] Breeding Center of Natural Insect Enemies for Pests, Henan Institute of Science and Technology, Xinxiang, China, [4] Institute of Zoology, Chinese Academy of Sciences, Beijing, China, [5] Landscape Ecology Group, Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, China

Understanding the traits related to species colonization and invasion, is a key question for both pest management and evolution. One of the key components is flight, which has been measured for a number of insect species through radar and tethered flight mill systems, but a general understanding of insect flight at a community level is lacking. In this study, we used flight mill experiments to quantify flight abilities of moth species, and simulation experiments to study which moths in mainland China have the potential for cross-island dispersal. We found that moths from superfamily Geometroidea (family Geometridae) have the weakest flight ability among the seven *Lepidoptera* superfamilies, which is characterized by the shortest longest single flight (LSF), the shortest time corresponding to the longest single flight (T_{LSF}) (time corresponding to the longest single flight), the lowest total distance flown (TDF), and the lowest average speed during the flight (V_{TDF}). Surprisingly, the family *Pyralidae* (superfamily *Pyraloidea*) has the highest flight endurance of all 186 species of 12 families in this study, which is unexpected, given its small size and morphological traits yet it shows the longest LSF and T_{LSF}. The comparison between species common to mainland and islands shows that flight distance (LSF) may be more important for species spread than flight speed. The results of mainland-island simulations show that when $P_{(LSF>CD)}$ (the proportion of individuals whose LSF is greater than the closest distance (CD) between mainland and island to the total number of individuals in the population) is less than 0.004, it is difficult for moth species to disperse to across islands without relying on external factors such as airflow. Over extended periods, with the immigration of species with strong flight abilities, islands are more likely to recruit species with stronger flight abilities.

Keywords: lepidoptera, flight ability, flight mill, cross-island dispersal, simulation

OPEN ACCESS

Edited by:
Gengping Zhu,
Tianjin Normal University, China

Reviewed by:
Lloyd W. Morrison,
Missouri State University, United States
Xiaolei Huang,
Fujian Agriculture and Forestry University, China

****Correspondence:***
Ai-bing Zhang
zhangab2008@cnu.edu.cn

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *07 September 2021*
Accepted: *28 October 2021*
Published: *10 December 2021*

Citation:
Zheng Y-X, Wang Y, Dai B-Y, Li Z, Huo Q-R, Cui J-X, Liu H, Li X-H, Hughes AC and Zhang A-B (2021) Flight Mill Experiments and Computer Simulations Indicate Islands Recruit More Capable Flyers of Moths. Front. Ecol. Evol. 9:771719. doi: 10.3389/fevo.2021.771719

INTRODUCTION

Insecta first evolved around 400 million years ago and was the first animal class to evolve the ability to fly over 325MYA (Panian and Wiltschko, 2004; Ross, 2017). It is also the taxa with the largest number of described species, currently estimated at around 5.5 million species (Stork, 2018). A major evolutionary advantage of flight is the ability to disperse, and flight is thought

to have enabled rapid diversification within the group (Tihelka et al., 2021). Lepidoptera and Coleoptera are the two largest orders within the class Insecta (Stork, 2018; Zhuang et al., 2018). As well as high diversity, both Lepidoptera and Coleoptera have important ecological roles, for example, larvae and adults provide prey for various predators, such as birds and bats, and adults can be important pollinators for flowering plants. However, some species have negative impacts on agricultural production, for example, every year more than 40 billion dollars are expended on chemical pesticides to control lepidopteran pests. Even single species can have huge economic implications, for instance, diamondback moth (*Plutella xylostella*) control in Brassica crops alone costs approximately US$1.4 billion annually (Furlong et al., 2013; Chattopadhyay et al., 2017). A clearer understanding of flight ability or niche characters can be used to predict where potentially invasive species may appear in the future and used to help develop and apply preventative countermeasures (Hudgins et al., 2017; Menchetti et al., 2019). Consequently, a large number of studies were conducted on the dispersal and migration of insects through radar observation methods (Chapman et al., 2008a,b) and tethered flight mill systems (Hashiyama et al., 2013; Hoddle et al., 2015; Jones et al., 2016).

Flight mills are widely used to study the flight abilities of various insects. However most studies focus on the flight ability of a single species (Ávalos et al., 2014; Hoddle and Hoddle, 2016; Fu et al., 2017a,b; Lopez et al., 2017; Jyothi et al., 2021), and there are few studies, which provide descriptions of the flight ability of multiple species (Jones et al., 2016) or a certain group. Some species of Lepidoptera have migratory behavior, which can be detected by radar-based methods. Such as the famous monarch butterfly (*Danaus plexippus*), which exhibits seasonal intergenerational migration across the continent (Gustafsson et al., 2015). Bogong moths of Australia (*Agrotis infusa*) travel 1,000 km from eastern Australia to a few caves in the Australian Alps to escape seasonally high temperatures (Common, 1954; Warrant et al., 2016). Compared to a few known species that have evolved relatively fixed long-distance migration behaviors, short-distance dispersal is likely to be more common (Nathan et al., 2008). In view of this short-distance dispersal, it is important to clarify the basic traits related to spread ability, such as the flight distance and flight speed of each species.

It remains a challenge to determine which traits are related to dispersal adaptation in island biogeography (Whittaker et al., 2017). Lepidopteran insects provide an ideal model to solve this problem, because of its high species diversity and convenience of flight ability measurement. At present, the research on the flight ability of Lepidoptera is usually limited to a few specific species and does not provide a general understanding of other Lepidopteran species, especially to understand the behavior of species dispersal and migration between mainland and islands at the community level. In this study, we aimed to: (1) investigate whether there are differences in flight capability of species between mainland and islands moth communities; and (2) explore mechanisms of island recruitment of more capable flyers of moths by dispersal simulations.

MATERIALS AND METHODS

Collecting Samples of Lepidoptera Adult

A total of 1,531 moth samples were collected in summers of 2017 (Chang, 2018; Jiang, 2018), 2018 (Jia, 2019), and 2020 from the Dongling Mountain (∼39°51′07″N, 115°29′13″ E) in Beijing, Zhoushan Island (∼30°5′22″N, 122°3′15″E) in Zhejiang Province, and Weihui city (∼35°39′18″ N, 113°59′2″E) in Henan Province, China, respectively(**Figure 1A**). In the following, samples collected from Beijing, Zhejiang, and Henan are represented by BJDLS, ZSDXG, and HNYHG, respectively. All adult moths were captured at night by setting light traps between 8 and 12 p.m. (Yang et al., 2012).

Tethered Flight Experiments

SUN-FL tethered flight mills (Beijing Taichi Pengcheng Electronics Company, China) were used to perform hanging flight experiments. This device records: total distance flown (TDF), total flight duration (T), the average speed during flight (V_{TDF}, $V_{TDF} = \frac{TDF}{T}$), longest single flight (LSF), the time corresponding to the longest single flight (T_{LSF}), and average speed corresponding to the longest single flight (V_{LSF}, $V_{LSF} = \frac{\mathrm{LSF}}{T_{LSF}}$) (Ávalos et al., 2014). The captured moths were subjected to a hanging flight experiment in a closed dark room. In the 2017 and 2018 experiments, each moth was included in a continuous 24-h hanging flight experiment without energy supplementation, while in the 2020 experiment, each experiment lasted under 10 h (moths generally became inactive before the end of the 10-h experimental period in all cases). We chose four parameters; V_{TDF}, LSF, T_{LSF}, and V_{LSF}, which may be less affected by hanging flight times (though in the subsequent analysis since we found most moths become inactive in less than 10-h hanging flight, thus experiment durations over this do not impact results).

Identification of Lepidopteran Specimens

All specimens collected from BJDLS (2017) and ZSDXG (2018) were transported at 0°C to the laboratory in Capital Normal University for DNA extraction and PCR amplification for subsequent DNA barcode identification (Hao et al., 2020). Experimental samples from HNYHG were identified based on morphology by Associate Professor Jianxin Cui of the School of Resources and Environment, Henan Institute of Science and Technology.

For all samples collected from BJDLS and ZSDXG, total genomic DNA was extracted individually from 2 to 3 legs, using QIAamp DNA Mini kit (Qiagen, Venlo, the Netherlands) following the protocol of manufacturer. PCR amplification was performed using standard DNA barcoding primers LCO1490 (GGTCA ACAAA TCATAA AGATA TTGG) and HCO2198 (TAAAC TTCAG GGTGA CCAAA AAATCA) (Folmer et al., 1994). The PCR products were bidirectionally sequenced at Beijing Zhongkexilin Biotechnology Co., Ltd., using an ABI 3730XL analyzer. All sequences of samples ∼658 bp were

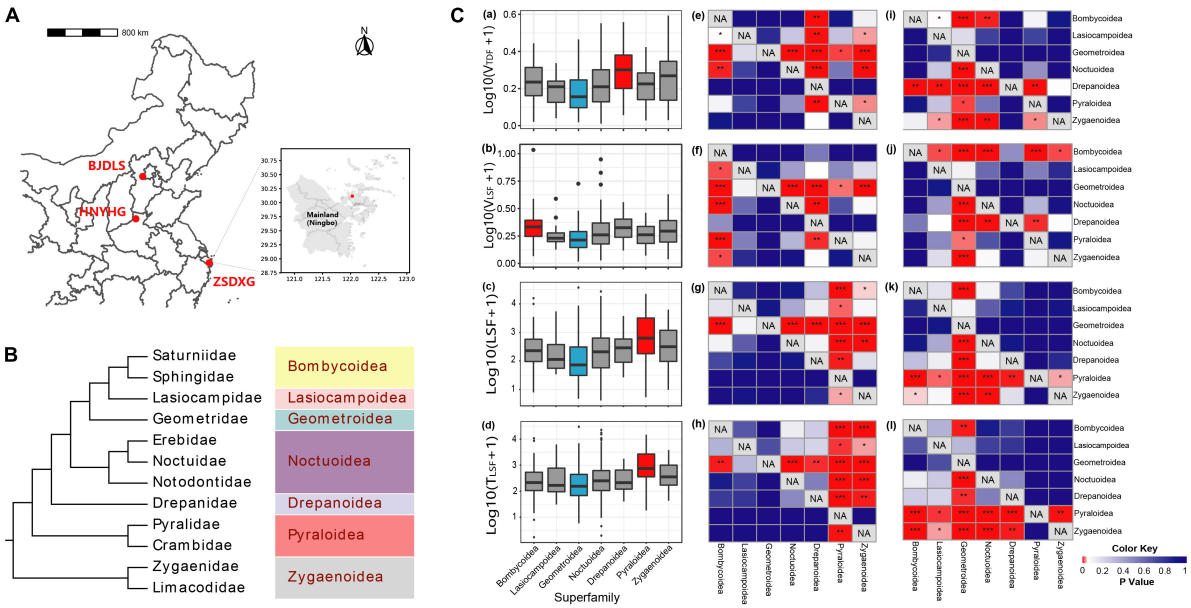


FIGURE 1 | Phylogenetic relationships above the family level and the differences of flight ability among the superfamilies were analyzed in this study. **(A)** Sample sites in this study. BJDLS, HNYHG, and ZSDXG samples were collected in 2017, 2018, and 2020, respectively; **(B)** the phylogenetic tree of all Lepidoptera samples in this study, redrawn according to Triant et al. (2018); **(C)** comparison of flight abilities among moths from different superfamilies and heat maps of *p*-value for Wilcoxon rank-sum test of the one-tailed hypothesis of the difference in flight ability between certain superfamilies (row) and other superfamilies (column). **(Ca–d)** The *X*-axis are indicators from all filtered data, and the *Y*-axis gives logarithmic conversion result of flight capabilities (V_{TDF}, V_{LSF}, LSF, T_{LSF}), respectively. The color of the box represents the strength of the flight ability, red represents the strongest, blue represents the weakest, and gray represents the medium level. **(Ce–h)** *P*-value of Wilcoxon rank-sum test of the one-tailed hypothesis (less) (*:0.01 $\leq p <$ 0.05, **: 0.001 $\leq p <$ 0.01, ***: $p <$ 0.001), respectively, the results corresponding to **(Ca–d)**. **(Ci–l)** *P*-value of Wilcoxon rank-sum test of the one-tailed hypothesis (greater) (*:0.01 $\leq p <$ 0.05, **: 0.001 $\leq p <$ 0.01, ***: $p <$ 0.001), respectively, corresponding to **(Ca–d)**.

submitted to the BOLD system[1] for taxonomic identification (Ratnasingham and Hebert, 2007).

Data Analysis

We removed the total flight distance parameter below the lower quartile. One of the reasons for these data could be that the wings of some moths are glued in the process of experiment, reducing their willingness or ability to fly. If these data were included, we would underestimate the fly ability of moths (Jantzen and Eisner, 2008; Combes et al., 2010). All calculations and statistical tests involved in this article are implemented by R (v4.0.4) (R Core Team, 2013). For the flight parameters of species level, family level, and superfamily level, the *shapiro.test* function in R was used to check for the normal distribution of the data, and the *Wilcox.test* function in the stats package to compare the differences among species, families, and superfamilies. The *lm* and *predict* functions in R were used for fitting and prediction of unary linear regression, respectively.

Simulation of Species Spread Between Mainland and Island

After obtaining the flight data of 36 ($n \geq 10$) moth species from 10 different families, we simulated the migration of different species between mainland and island (see below) in R. In the simulation, we approximate Zhoushan Island and its adjacent mainland (Ningbo) into two circular regions, and the minimum distance between the two circles (e.g., the shortest distance between Ningbo and Zhoushan Island) is about 8.3 km measured based on Map World on National Platform for Common Geospatial Information Services[2].

Basic Assumptions of the Dispersal Model

We assume that the initial moth population consists of a male and a female, and that they exist in an environment with evenly distributed resources. The female lays eggs somewhere after mating, and the larvae will only feed near the hatching place due to limited dispersal ability, so on a large spatial scale, we believe that the offspring produced by the same female moth will not spread during the larval and pupal stage, that is, the position where the female moth lays eggs is the initial position of the progeny adult's spread (Considering natural enemies, climate, and other factors, it is assumed that only r_{max} (assuming $r_{max} = 10$) individual offspring produced by each female moth can successfully emerge, and the sex ratio is 1:1; **Supplementary Figure 1**). The offspring will spontaneously spread to the surrounding environment in random directions following metamorphosis into the adults and form an extended second-generation distribution range with an area of S. The

[1]http://www.boldsystems.org/

[2]https://www.tianditu.gov.cn/

mating females in this population lay eggs nearby, to produce the second generation, and the offspring continue to circulate this process. According to the logistic population growth model (Pearl and Slobodkin, 1976), the density of species is limited by the carrying ability K (assuming K = 30, that is, 30 adults per square kilometer, though of course, it may be higher than this for many species). The number of offspring produced by each female in the population is restricted by the population density of the current generation, and all adults can mate with a sex ratio of 1:1. Supposing the number of fertile females in this generation is N_{ff} and regardless of limitation in species spawning ability, females can produce the largest number of individuals r, so the number of fertile offspring (adult offspring) produced by fertile females in each generation (r) should satisfy the following inequality:

$$r \leq \frac{K \times S}{N_{ff}}$$

Its ecological meaning is that the females are restricted by density during the larval stage, leading to a difference in the amount of food, then changing the number of eggs produced after emergence and mating (Rivero et al., 2001; Telang et al., 2006).

Key Parameters Implemented in R Program

The population dispersal simulation is analyzed using several steps in R: (a) Determination of population distribution area. Theoretically, the distribution range of a population is composed of the largest convex polygon formed by all individuals in the population; however, it is a challenge to determine the actual distribution range of a certain population under natural conditions. The R function *boxplot.stats* (package grDevices) was used to determine the outliers of all the individual abscissas and ordinates (R Core Team, 2013). According to the outliers, irregular population distribution is simplified into an area composed of two concentric rectangles. The core rectangle is the population core distribution area and is recorded as S_{core}. The rectangular area at the edge of the population distribution is recorded as S_{edge}. (b) Determination of the fertile females of the population. Supposing that the female moth mates with only one male moth in its lifetime, and the male does the same. Therefore, the number of fertile females in the population is the minimum number of males and females, and fertile females are determined without repeated sampling among all females according to this minimum. If the minimum value is 0, the population will be extinct in the next generation. (c) The production of offspring. The new individuals are generated from the locations of fertile females of the previous generation. The dispersal angle is random, and dispersal distance is the LSF of each species measured in the actual experiment (see above), and its value is determined to conform to the logarithmic normal distribution. The number of offspring produced by a female (R) is controlled by an inequality $r \leq \frac{K \times S}{N_{ff}}$. When $r > r_{max}$, $R = r_{max}$; when $1 < r \leq r_{max}$, $R = r = \frac{K \times S}{N_{ff}}$; and when $r < 1$, $R = 1$. The distribution area of the core population and the marginal population is calculated independently. If the population distribution is a point or a straight line, the number of adult offspring that a female can produce is calculated by r_{max} .

Simulation of Species Spread Across Islands

We assume that the ability of moths to cross the strait to complete the spread between the mainland (island) and the island depends on the LSF. Regardless of airflow, human factors, and more stochastic natural factors, species can complete migration only when the LSF is greater than the shortest distance between the islands (but the stepping stones that exist between the islands reduce the difficulty of crossing the islands.

According to the LSF of the wild populations of the moth species that we measured, we simulated the spread of each species from the center of the area through computer simulations and observed the spread of offspring individuals across islands within 500 generations (two generations per year (Hao et al., 2020), it will take about 250 years). In our data, there are a total of 36 species with at least 10 individuals (**Supplementary Table 1**), of which the LSF data of 34 species (94.4%) had a log-normal distribution, thus the LSF data of all moths are assumed to correspond with log-normal distribution in the simulation. We simplified Ningbo (with an area of about 9,816 km^2) and Zhoushan Island (with an area of about 502 km^2) into two circular areas with radius of 56 and 12 km, and the closest distance (CD) between the two circles is 8.3 km. Here we defined a new concept P_(LSF > CD) to explain the dispersal ability of species, and its specific definition is that the proportion of the number of individual that LSF is longer than the closest distance (CD) between mainland and island to the population size.

RESULTS

Data Filtering and Taxonomy Identification

A total of 1,142 (∼74.4%) experimental flight data samples were analyzed, i.e., including 222 samples from BJDLS (the minimum TDF is 18.84 m), 409 samples from ZSDXG (the minimum TDF is 67.23 m), and 511 samples from HNYHG (the minimum TDF is 33.91 m). Filtered data were collected from seven Lepidoptera superfamilies (Noctuoidea, Bombycoidea, Zygaenoidea, Geometroidea, Drepanoidea, Lasiocampoidea, and Pyraloidea), 12 families (Erebidae, Noctuidae, Notodontidae, Saturniidae, Sphingidae, Zygaenidae, Limacodidae, Geometridae, Drepanidae, Lasiocampidae, Crambidae, and Pyralidae), and 186 species, and among them, 81 species with at least 3 individuals and 36 species with at least 10 individuals. A total of 631 individuals from BJDLS and ZSDXG were identified by comparison with the BOLD system database. Among them, 547 individuals had best hits with similarity scores of at least 98 (part 1), and the remaining individuals had similarity scores between 90.43 and 97.99% (part 2). The remaining 511 individuals from HNYHG were identified using morphology (part 3). Part 1 contains 134 species, part 2 contains 40 species, and part 3 contains 39 species. There are 13 common species in parts 1 and

3, but only one common species, *Catocala kaki* (Noctuidae) in parts 2 and 3.

Comparison of Flight Capabilities at Different Taxon Levels

We compared the difference in flight abilities (V_{TDF}, LSF, T_{LSF}, and V_{LSF}) at different taxon levels in all filtered data. First, we checked whether the flight data at different taxon levels conformed to a normal distribution and found that only a small proportion of the data conformed (at the superfamily level, only 0–42.9% data of the above four parameters conform to a normal distribution; at the family level, only 8.3–41.7% data of the above four parameters conform to a normal distribution; only 12.5–55.6% of the flight capability data at family level captured from different sampling sites show a normal distribution. Therefore, the Wilcoxon rank-sum test of the one-tailed hypothesis is selected to compare the sample differences between different classification levels and different sampling sites.

The lowest V_{TDF}, V_{LSF}, LSF, and T_{LSF} were found in moths from superfamily Geometroidea, which were lower than those in the other five superfamilies except for no significant difference with superfamily Lasiocampoidea (**Figure 1C**). Superfamily Drepanoidea is significantly higher than the other five superfamilies ($p < 0.05$) on the flight parameter V_{TDF} except that it does not show any significant difference from superfamily Zygaenoidea. In V_{LSF}. Bombycoidea is significantly higher than the other five superfamilies ($p < 0.05$) with the exception of Drepanoidea. While superfamily Pyraloidea has the longest LSF and V_{LSF} among seven superfamilies in our study ($p < 0.05$).

Correlation analysis between flight parameters and moth morphological data showed that V_{TDF} and V_{LSF} had a significant positive correlation with the increase in moth body size [body length (BL), total wing area (TWA), forewing area (FA), and hind-wing area (HA)], while T_{LSF} had a weak negative correlation with moth TWA and HA (**Supplementary Figure 2**), highlighting the importance of individual wing-traits in determining flight performance and ability.

Comparison of Flight Capabilities of Species at Different Taxon Levels and Captured From Different Sampling Sites

Comparison of Common Families of Flight Capabilities of Mainland and Island Samples

A total of six families of species appeared in the three sampling sites (BJDLS, HNYHG, and ZSDXG). We selected samples from three families (Geometridae, Erebidae, and Notodontidae) for flight ability comparison. The data collected from BJDLS and HNYHG were combined into Mainland Dataset and that from ZSDXG was used as an Island Dataset. These three families satisfy that the number of samples collected on the island is more than 20. The three families on the mainland and island are composed of more than 10 species, but there are few overlapping species between different sites (**Figure 2E**). The speed parameters (V_{TDF} and V_{LSF}) of the mainland and the island samples did not show a consistent pattern. Our results suggested that only the Erebidae showed significant differences between mainland and island in the term of speed, and island individuals were significantly faster ($p < 0.05$) than mainland individuals. Notodontidae is only in the V_{TDF} parameter showing that the mainland is significantly stronger ($p < 0.05$) than the island samples (**Figures 2A,B**). The average values of LSF and T_{LSF} of Geometridae and Notodontidae show that the islands are higher than the mainland samples, but only Geometridae shows significant differences ($p < 0.05$) in the T_{LSF} (**Figures 2C,D**). The LSF and T_{LSF} of Erebidae are stronger than those of island samples and only one parameter, T_{LSF}, has statistical significance.

Comparison of the Flying Ability of Common Species Captured From HNYHG and ZSDXG

Among the sampling sites in the three regions, only Henan (HNYHG) and Zhejiang (ZSDXG) samples have more than three individuals shared (**Figure 3E**), and the common species data of Beijing (BJDLS) and Zhejiang (ZSDXG) do not meet the minimum requirement of at least three individuals. These three species are *Biston panterinaria* (*B. panterinaria*; Geometridae), *Thyas juno* (*T. juno*; Erebidae), and *Uropyia meticulodina* (*U. meticulodina;* Notodontidae). Our results showed that only the V_{TDF} of *T. juno* showed significant differences ($p < 0.05$) between the two sampling sites (**Figure 3A**), and the mean values of V_{LSF}, T_{LSF}, and LSF of the three species between the two sampling sites were relatively close, and no significant differences were detected (**Figures 3B–D**). The mean values LSF and T_{LSF} of species *B. panterinaria* and *T. juno* collected from ZSDXG were higher than those from HNYHG, while the opposite results were observed for *U. meticulodina* collected from the two sites (**Figures 3C,D**).

Simulation Results of Species Spread Across Islands

Cross Island Simulation Results of Species Within a Limited Generation

We selected 14 species from the above 36 species according to the LSF gradient from high to low for simulations. The results showed that as the LSF of the species decreases, the number of cross-island migrations (from the mainland to island) in the offspring will decrease significantly. *Locastra muscosalis* (*L. muscosalis*) has the strongest dispersal ability, $P_{(LSF>CD)} = 0.1516$, and an average of 166.7 individuals in each generation have successfully spread (**Figure 4A**). *Cydalima perspectalis* (*C. perspectalis*) with the second strongest dispersal ability, $P_{(LSF>CD)} = 0.0415$, and an average of 38.3 individuals in each generation have successfully spread. When the $P_{(LSF>CD)}$ is between 0.0127 and 0.0021 (from high to low), the average number of cross-island individuals per generation drops from 6.6 to 1.1. When the $P_{(LSF>CD)}$ of the species $\leq$0.0012, almost no offspring individuals can complete cross-island migration.

When the $P_{(LSF>CD)}$ is greater than 0.05, in the 500-generation expansion of the simulated population, far more than half of the generations (∼75–99%) will occur cross-island migration of individuals in the population (**Figure 4B**). When the $P_{(LSF>CD)}$ is between 0.002 and 0.05, the ratio

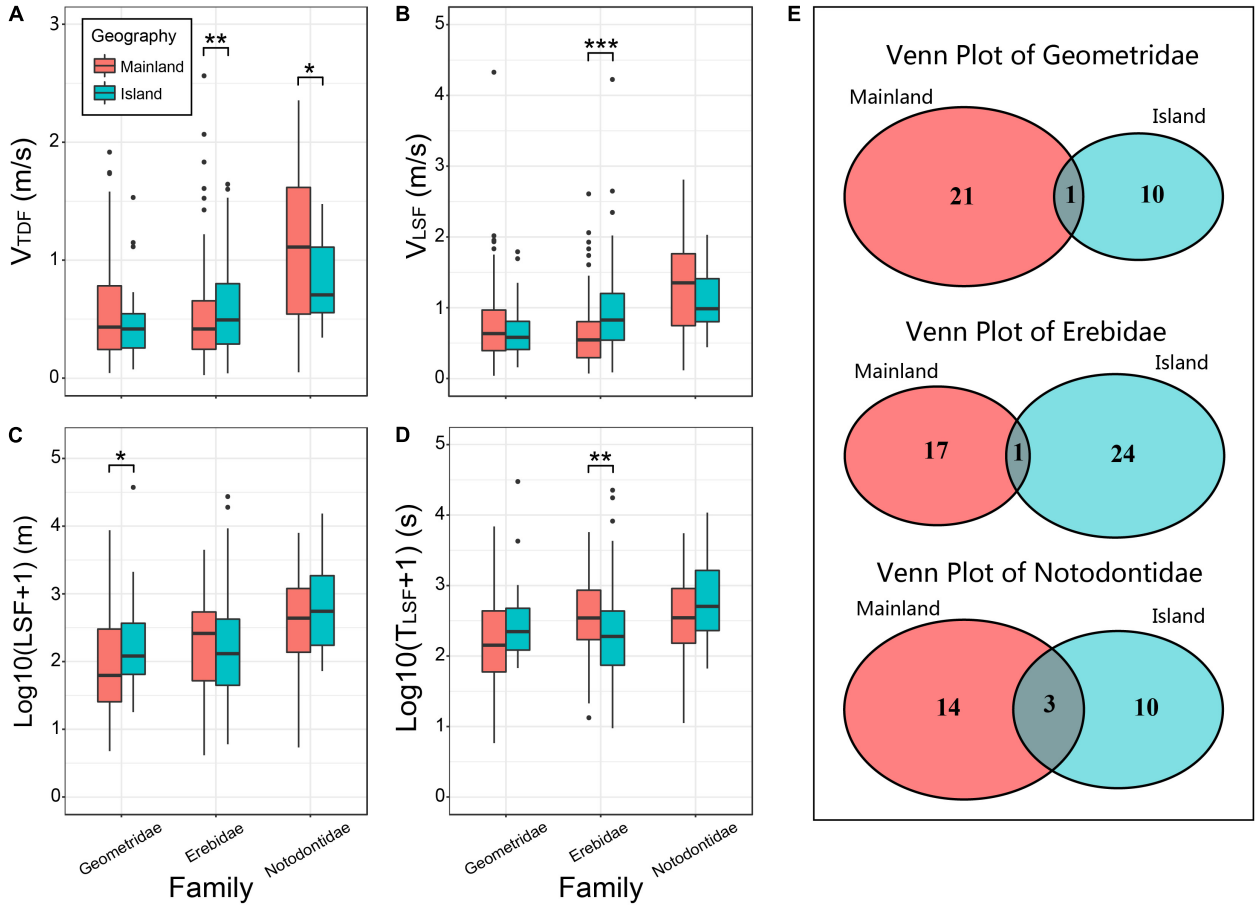


FIGURE 2 | Comparison of flight capabilities between mainland and island (Geometridae, Erebidae, and Notodontidae). **(A–D)** Box plot of comparing the flight capabilities of mainland and island samples. The difference between the samples is calculated by the Wilcoxon rank-sum test of the two-tailed hypothesis ($*$: $0.01 \leq p < 0.05$, **: $0.001 \leq p < 0.01$, ***: $p < 0.001$),only the samples with significant differences between Mainland and Island are marked with *, ** or *** in this figure. **(E)** Venn diagram of species in three families captured from mainland and island.

will increase from 25.2 to 44.8%. When the $P_{(LSF>CD)}$ is lower than about 0.001, at most only 24 generations (4.8% of all generations) have individuals which perform cross-island migration.

Comparison of the Flying Ability (LSF) of Cross-Island Individuals and Stranded Mainland Populations

We selected the 50th generation of cross-island dispersal data from eight species with strong flight abilities and compared the difference in flight ability between populations composed of cross-island individuals (the number of individuals in each species population is 160, 34, 1, 7, 2, 1, 1, and 1) and populations stranded on the mainland (random sampling is used to select data of 5,000 individuals from the mainland population; **Figure 5**). Our results show that the flying ability of the cross-island populations of each species is significantly stronger than that of the mainland populations. Among them, the LSF of the cross-island populations of *L. muscosalis*, *C. perspectalis,* and *P. changmei* in particular is significantly higher ($p < 0.001$) than that of the mainland stranded populations. The other five species also showed the same pattern with significant differences between mainland and island populations in each ($p < 0.05$).

Simulate the Relationship Between $P_{(LSF>CD)}$ and Cross-Island Individuals per Generation in the Process of Species Spreading From Mainland to Island

By establishing the direct relationship between the species ability to spread $P_{(LSF>CD)}$ and the average number of successful cross-island (spreading from mainland to island) individuals per generation (**Figure 6**), we found that when $P_{(LSF>CD)} < 0.004$, almost no cross-island individuals would appear [When $P_{(LSF>CD)} = 0.004$, the number of cross-island individuals per generation is predicted to be around 0.47, and the prediction lower and upper limit is (−16.8, 17.7); When $P_{(LSF>CD)} = 0.003$, the predicted value is −0.53, and the prediction lower and upper limit is (−17.8, 16.7)]. Considering that the area of the habitat may directly affects the size of the species population and finally affect the dispersal efficiency of the species, we also performed the same regression analysis on data which species spreading

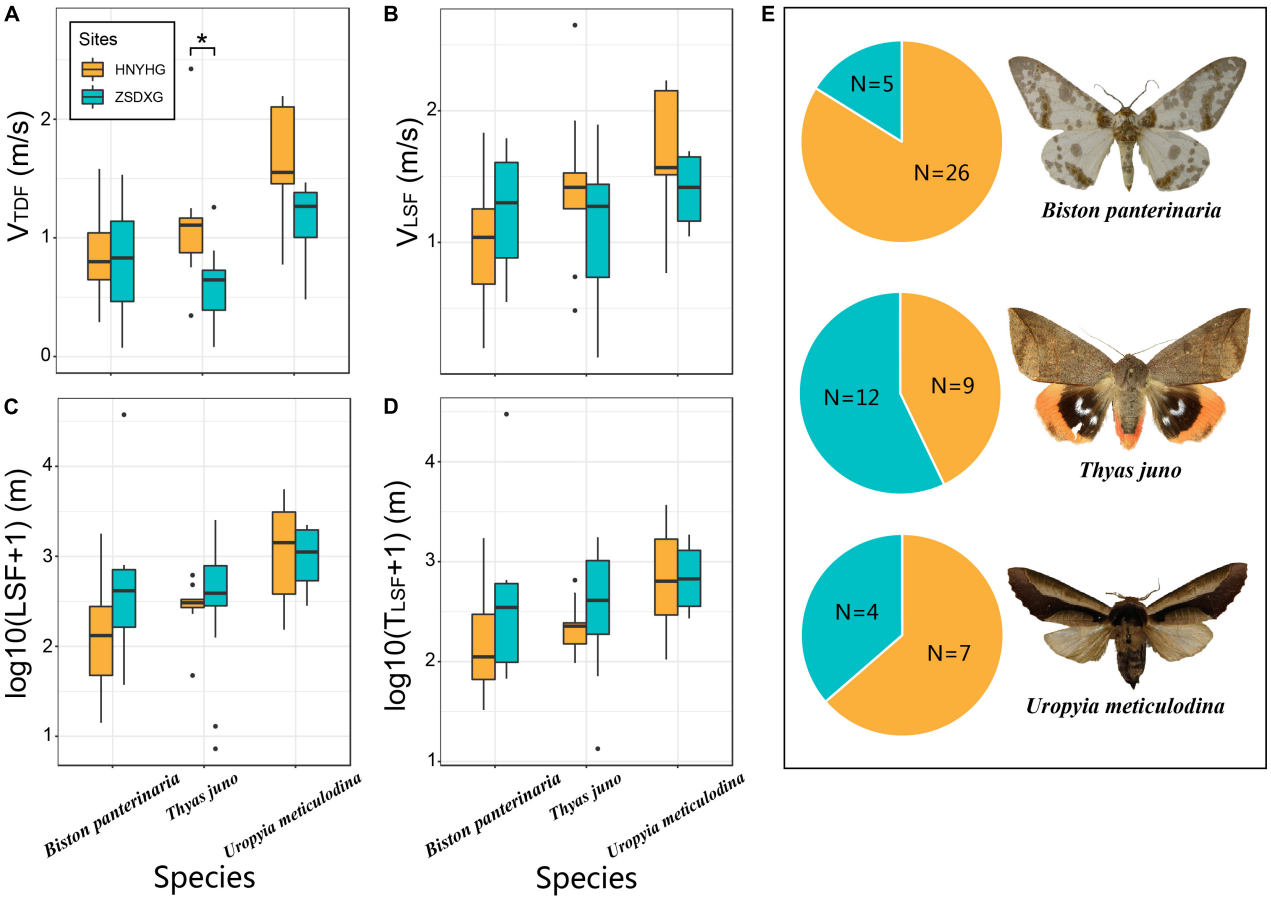


FIGURE 3 | Comparison of flight ability of common species (*Biston panterinaria*, *Thyas juno*, and *Uropyia meticulodina*) between HNYHG and ZSDXG. **(A–D)** Box plot of comparing the flight capabilities of common species captured from HNYHG and ZSDXG. The difference between the samples is calculated by the Wilcoxon rank-sum test of the two-tailed hypothesis (∗ : $0.01 \leq p < 0.05$, **: $0.001 \leq p < 0.01$, ***: $p < 0.001$), only the samples with significant differences between HNYHG and ZSDXG are marked with *, ** or *** in this figure. **(E)** Sample size of three species between two sampling sites [species photo source: *T. juno* (license holder: Unspecified, University of Oslo, Natural History Museum), *U. meticulodina* (license holder: Unspecified), and *Biston panterinaria* (license holder: Axel Hausmann, SNSB, Zoologische Staatssammlung Muenchen)].

from the island to the mainland (**Supplementary Figure 6**). The regression equation obtained is $y = -3.73 + 1330x$ ($p < 0.001$, $R^2_{adj} = 0.98$). When $P_{(LSF>CD)} < 0.003$, almost no cross-island individuals are generated. [When $P_{(LSF>CD)} = 0.003$, the predicted value of the number of cross-island individuals is −0.26, and the prediction lower and upper limit is (−15.7, 16.2). When $P_{(LSF>CD)} = 0.002$, the predicted value of the number of cross-island individuals is −1.07, and the prediction lower and upper limit is (−17.0, 14.9)].

DISCUSSION

The Differences of Flight Abilities Among Seven Superfamilies of Lepidoptera

Previous studies on the flight ability of insects using a flight mill system mainly focused on agricultural pests (Hashiyama et al., 2013; Sappington and Burks, 2014; Hoddle and Hoddle, 2016; Fu et al., 2017a,b; Lopez et al., 2017; Babu et al., 2020). This study is the first to use a unified method to study the flight ability in Lepidoptera families and superfamilies to explore traits related to possible invasion and colonization ability. In this experiment, 1,142 specimens of seven Lepidoptera superfamilies were collected. In total 186 species had their flight ability measured in the experiment. Superfamily Geometroidea showed no significant difference with superfamily Lasiocampoidea but was significantly weaker than the other five superfamilies (**Figures 1Ca–d**). A total of 197 samples of superfamily Geometroidea were used for data analysis, accounting for 17.3% of the total samples, and only family Geometridae species are included in this superfamily. Geometrid moths are frequently flattened when in a static state, and the color of body and wing is often very similar to the surrounding environment. One of the best known Geometrid moths is the peppered moth (*Biston betularia*). The initial phenotype of this moth was speckled black and white, but industrialization in Britain and other regions in the nineteenth century increased the proportion of melanic individuals (Cook et al., 2012; Cook and Saccheri, 2013).

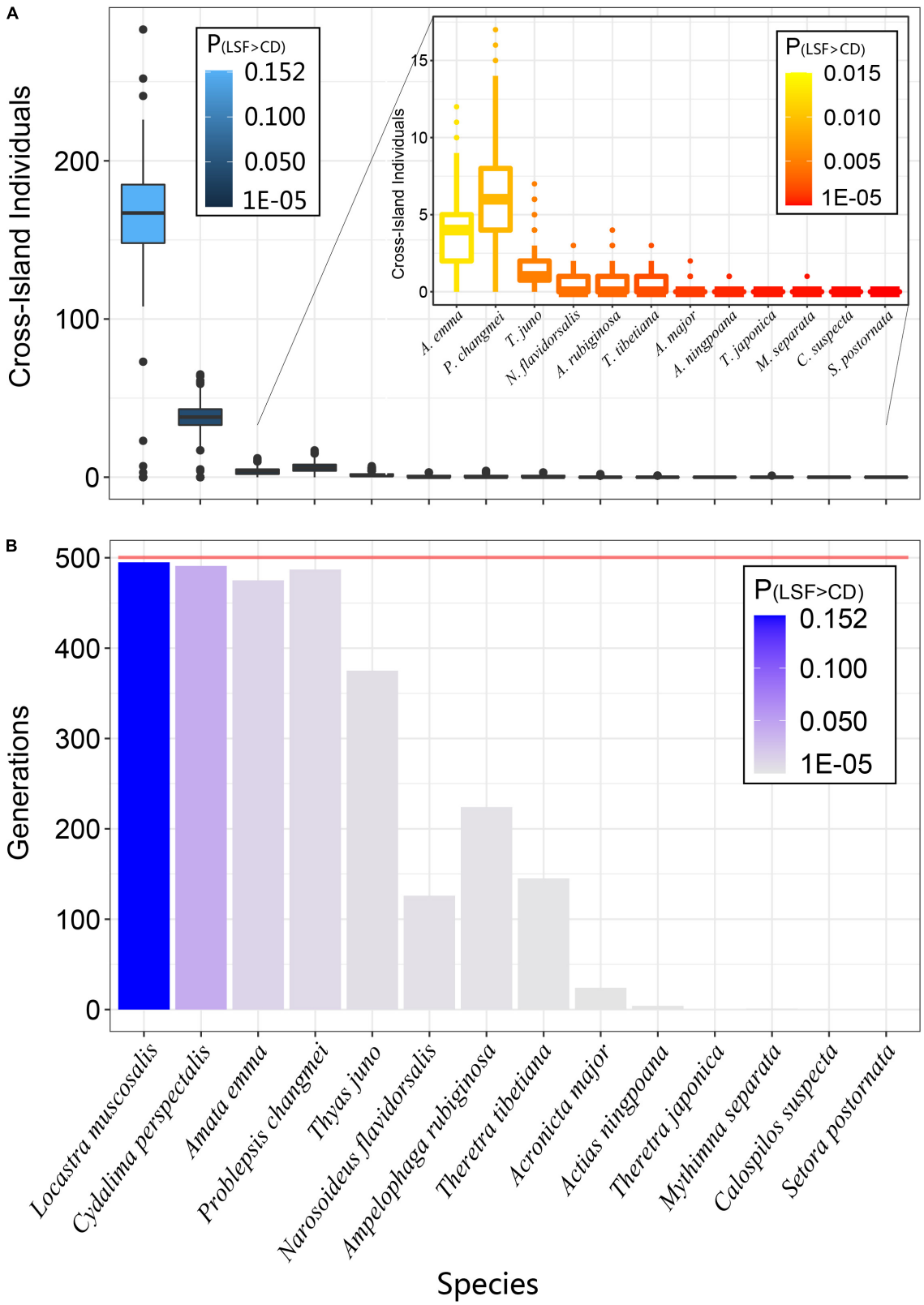


FIGURE 4 | The average number of individuals with different LSF levels of species dispersal from the mainland to island **(A)** and the number of generations **(B)** within 500 generations. LSF, longest single flight. $P_{(LSF>CD)}$ refers to the proportion of individuals with LSF greater than the CD between the two islands in the offspring of a species in the total population.

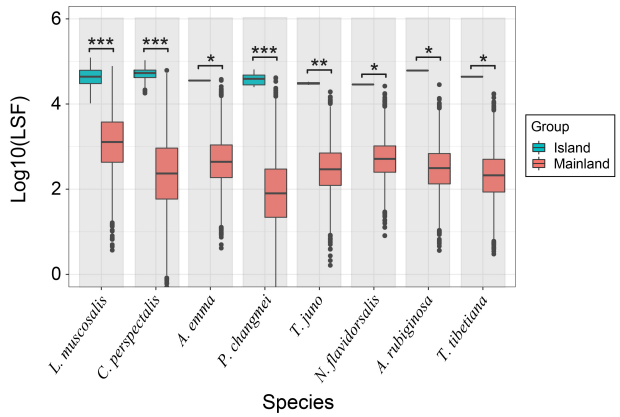


FIGURE 5 | Comparison of the flight ability of the trans-island populations of each species and the mainland stranded population (species dispersal from the mainland to island). The eight selected species are displayed on the *x*-axis, and the *y*-axis is the log10 conversion value of the LSF parameters of the different populations of each species. The difference between different populations is calculated by the Wilcoxon rank sum-test of the one-tailed hypothesis (*: $0.01 \leq p < 0.05$, **: $0.001 \leq p < 0.01$, ***$p < 0.001$). LSF, longest single flight.

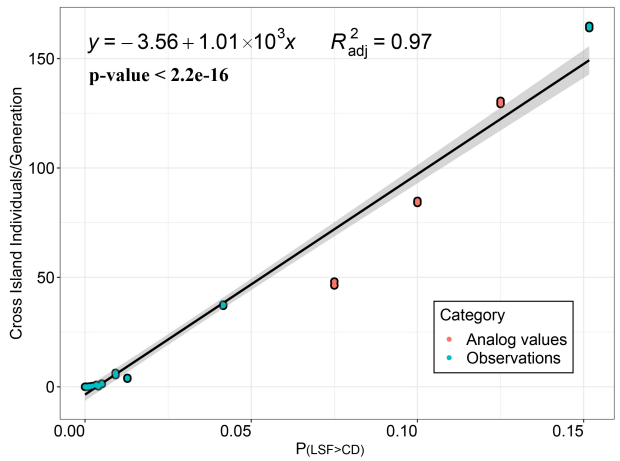


FIGURE 6 | Simulation of the relationship between $P_{(LSF>CD)}$ and cross-island individuals per generation in the process of species spreading from the mainland to island in 500 generations. Unary linear regression is shown in black and the CI is shown in gray. The blue points represent the $P_{(LSF>CD)}$ actually measured in the experiment (data come from measurements of 14 species), and the orange points represent the simulated point for supplementation (data come from simulation values of three hypothetical species). We repeated the simulation three times for each species. LSF, longest single flight; CD, closest distance.

This change in the main color-form as a consequence of selective predation is still considered as an example of the rapid evolution as a consequence of natural selection in a changing environment, which is driven by selective predation (Cook et al., 2012; Cook and Saccheri, 2013). Individuals of American peppered moth (*Biston betularia*) larvae are selected to match background color depending on their different host plants, and the slender shape makes them look more like branches, thereby using a similar approach to reduce predation (Noor et al., 2008; Eacock et al., 2017). Thus, the impact of selection on key traits is clear, but at least in this group, it does not require an ability to fly to escape predators, and conversely, the general body form of Geometrids may be expected to have a weaker flight performance.

Superfamily Drepanoidea had the highest in V_{TDF} parameters (**Figure 1Ca**), which was faster than the other five superfamilies (with the exception of Lasiocampoidea). But at present, we know very little about this superfamily, with only a few omics studies about the social caterpillar *Drepana arcuata* from this superfamily (Yack et al., 2001; Yadav et al., 2020a,b), which provide no insights into flight performance. Superfamily Pyraloidea (family Pyralidae) showed the strongest LSF and T_{LSF} in comparison to other superfamilies (**Figures 1Cc,d**), suggesting that it has high flight endurance. Superfamily Pyraloidea has the smallest body size in all samples collected in this study, such as the smallest BL, weight, and wing area. At present, there are too few comparative studies on the high taxonomic rank flight ability of Lepidoptera, but species-specific morphology is likely to be responsible for the differences found here.

Among 187 individuals belonging to superfamily Bombycoidea in this study, 171 individuals (∼91.4%) are from family Sphingidae and the remaining individuals are from family Saturniidae. The results show that the of Bombycoidea are significantly higher than those of the other five superfamilies (**Figure 1Cb**), possibly because superfamily Bombycoidea includes the family Sphingidae with the highest (**Supplementary Figure 3**). Our results prove that moths can have extensive niche differentiation, and two families with distant phylogenetic relationships (**Figure 1B**) may have similar flight abilities. Traits to escape predation are important, as Lepidoptera can account for more than 50% of the food of some insectivorous bats (Galan et al., 2018), and within this, Noctuidae can account for 50% of Lepidoptera species consumed (Zeale et al., 2011). Hawk moths (Lepidoptera, Bombycoidea, Sphingidae), Notodontidae, and Noctuidae (Lepidoptera and Noctuoidea) are more active and fly faster than most other moths. Large species with fat bodies may be particularly vulnerable to bats, such as many species in Sphingidae and Noctuidae, and need to be able to either fly fast or have other adaptations to reduce the risk of predation (Ntelezos et al., 2017; Rubin et al., 2018; Aiello et al., 2021). Hawk moths have the most aerodynamic body form of most moths, based on both wings and body morphology (Willmott and Ellington, 1997), which enables very high-speed short-distance flight.

Difference of Flight Capability of Moths Between Mainland and Island

Affectd by the environment of the habitats, different geographical populations may have different flight abilities, although LSF and T_{LSF} have no statistically significant differences, the mean value of V_{TDF} and V_{LSF} in island are much higher than mainland [ratio span ($^{island}/_{mainland}$): LSF (1.90–4.03), T_{LSF} (1.03–3.20), V_{TDF} (0.71–1.45), V_{LSF} (0.85−1.66)] (**Figures 2A-D**). We found that the same group (at family level) between mainland and island showed difference in flight ability. We speculate that the

difference may be due to different species composition. Because most of the flight data of the same family at different sites are from different species. To test this hypothesis, we compared the data of common species [*B. panterinaria* (Geometridae), *T. juno* (Erebidae), and *U. meticulodina* (Notodontidae)] from the island (ZSDXG) and the mainland (HNYHG) (**Figures 3A–D**). *U. metaculodina* showed that the mean value of island individuals is lower than that of mainland individuals (ratio of mean value are 0.62 and 0.76 in LSF and T_{LSF}, respectively). But the average values of LSF and T_{LSF} in the samples of *B. panterinaria* and *T. juno* collected from islands are larger than those from mainland samples, although the difference was not statistically significant (ratios of mean value are 23.30 and 19.84 between parameters LSF and T_{LSF} in species *B. panterinaria* and ratios are 2.00 and 2.30 between LSF and T_{LSF} in *T. juno*). Based on the above results, we infer that islands may recruit more species with strong flight capacity (in parameter LSF and T_{LSF}) than the mainland, which may expand differences in average of LSF and T_{LSF} between mainland and island; moreover, prevalent wind current speed and direction are likely to influence this. The species found on both islands and continents show that island individuals have a weaker flight speed (V_{TDF} &and V_{LSF}) than continental individuals, which indicates that flight distance (LSF) may be more important for species ability to spread than flight speed.

Zhoushan Island is only separated from the mainland about 9,000 years ago due to the rising sea level caused by climate change (Wang, 1980). Zhoushan Island is the largest island in the Zhoushan Archipelago and is close (the shortest distance is only about 8.3 km) to the mainland. In such a short period of time and extremely similar natural environment, the same species may not have time to differentiate their flight abilities or adapt to local conditions, such as differing wind speeds. However, although 9,000 years is not enough for species differentiation, the level of gene flow in addition to environmental pressures depends on wing-traits that relate to flight ability. It is generally difficult to prove this through actual observation data for these small insects, but we have measured the flight parameters of a considerable number of species, so we can verify the possibility through the simulation.

The Simulation Reveals the Possible Dispersal Threshold

Our results show that *L. muscosalis*, the species with the largest LSF, has the highest number of cross-island generations and the highest number of cross-island individuals per generation in 500 generations. With the decrease of species dispersal ability ($P_{(LSF>CD)}$), the number of cross-island generations and the number of cross-island individuals in each generation decreased rapidly. However, it is interesting to note that the decline rate of the cross-island generations is slower than that of the number of cross-island individuals per generation. When the average number of cross-island individuals decreased from 166.6 to 6.6 (corresponding to the first four species with $P_{(LSF>CD)}$ value declined from 0.151 to 0.009 in **Figure 4**), the number of cross-island generations remained at a high level (495–487). When $P_{(LSF>CD)}$ decreased from 0.005 to 0.0012 (From *T. juno* to *A. major* in **Figure 4**), the number of cross-island individuals was very low (from 1 to 2), and the number of cross-island generations decreased rapidly from 375 to 24. For the remaining species, there was almost no cross-island migration within 500 generations, and the $P_{(LSF>CD)}$ dropped to less than 0.0006.

Regression analysis also shows that when $P_{(LSF>CD)}$ drops below 0.004 (obtained from the simulation of spreading from the mainland to the island) and 0.003 (obtained from the simulation of spreading from the island to the mainland), the species lose the ability to migrate over the sea (the predicted value of the number of individuals that successfully spread is less than 0) (**Figure 6** and **Supplementary Figure 6**). Regardless of whether the species is spreading from the mainland to the island or from the island to the mainland, the requirements for $P_{(LSF>CD)}$ are relatively consistent. As long as a species can complete the spread from the mainland to the island (or from the island to the mainland), it may also be able to complete the reverse proliferation activity with the help of wind conditions. Our results showed that the flight ability of the individuals that successfully cross to or from the island is significantly stronger than that of the individuals that remain on the mainland (**Figure 5**). Similar results are also found in the simulation of species spreading from a small circle (island) to large circle (mainland) (**Supplementary Figures 4, 5**). We did not simulate whether individuals with strong flight ability would produce a larger proportion of offspring with strong flight ability (though it is likely to be heritable), but a study on *Episyrphus balteatus* (Diptera and Syrphidae) with migration behavior showed that the offspring of migrants in this species had higher flight activity than those of resident individuals (Dällenbach et al., 2018). The reality is far more complicated than the simulation. At present, we do not know how many of the species we collected from the island existed before Zhoushan Island separated from the mainland. Although the individuals successfully landing on the island from the mainland are all with stronger flying capability, the individuals from the islands that mate with them may be not strong enough, so they may could not produce offspring with strong flying capabilities. Molecular approaches would be needed to determine gene flow between mainland and island to determine how these populations interact and how many still show genetic connectivity to mainland populations. Furthermore, mainland individuals generally come from more diverse communities, so may be more competitive than species adapted to more depauperate island communities.

Factors Affecting the Spread of Lepidoptera Insects

The known factors that may affect the flight ability of Lepidoptera and other insects include sex and days post-eclosion. The effect of sex on flight ability varies between species. In *Helicoverpa armigera* (Lepidopera and Noctuidae), males have better flight abilities than females (Jyothi et al., 2021), while in *Amyelois transitella* (Lepidopera and Pyralidae) and *Rhynchophorus ferrugineus* (Coleoptera and Dryophthoridae), there is no sexual difference (Ávalos et al., 2014; Sappington and Burks, 2014). The flight capacity of newly emerged adults is weak in many

groups, reaching the peak at 3 or 4 days post-eclosion (Tanaka and Yamanaka, 2009; Liu et al., 2011; Fu et al., 2017b).

In addition, how Lepidopterans determine the direction of flight is a longstanding question for researchers. Species, such as monarch butterflies (Reppert et al., 2016) and Bogong moths (Dreyer et al., 2018) are able to navigate using geomagnetic fields, time compensated sun compass, and identifying landmarks to determine migration routes. But the processes guiding short-distance dispersal are less well understood. Lepidoptera individuals need to determine the destination if they travel between islands and other landmasses, and many sensory cues are not available over water, and loss of energy and poor weather are more likely to cause death, though studies show that some species can detect polarized light, and this may impact navigation over extended distances. Recently, MacDonald et al. (2019) partly proved that butterflies rely on vision to detect patches of suitable habitats. Their research shows that individuals with normal vision are 30.1 times more likely to navigate to the target island after being released than those with induced flash blindness.

Limitations

Although we collected a large number of Lepidopteran samples from the field, it is almost impossible to identify the physiological state of each sample, and its performance on the flight mill depends greatly on the physiological state. This may be the main reason for the high variability for some measures within species in our study. According to the activity of insects during sampling, the low TDF may be caused by the discomfort of the tested insects caused by the flight mill or the weakening of the flight ability caused by human operation error, but these factors should not be responsible for consistent differences observed between samples from different species and localities. However, whether it is reasonable to filter data in this way remains to be verified. Whilst we changed the experimental duration in 2020, the four flight parameters used in our analysis are unlikely to have been impacted. In addition, the dispersal direction of individuals in our simulation is completely random, and MacDonald et al. (2019) also pointed out that this random flight null assumption tends to overestimate the proportion of successful dispersal of butterflies toward adjacent habitat patches (MacDonald et al., 2019).

CONCLUSION

Understanding colonization and invasion patterns of species and the traits responsible is critical both to understand species biogeography and to enable proactive management. Our results show that among the common Lepidoptera groups (at the level of superfamily or family), superfamily Geometroidea and family Geometroidae have the weakest flight ability, while Pyralidae has the strongest flight endurance. Additionally, we found that island species generally have higher LSF than mainland species, although we did not detect statistically significant differences in our data overall, clear patterns were observed in some species. The simulation results show that the stronger the flight ability is [in this study, the ratio ($P_{(LSF>CD)}$) of the offspring whose LSF is greater than the shortest distance between islands is used to measure the flight ability], the species in coastal habitats are more likely to have successful cross-island dispersal, and when $P_{(LSF>CD)}$ is lower than 0.004, moths rarely spread between islands under natural conditions. If the habitat of the immigrating species is suitable for the survival of the immigrating species, it will be possible for the immigrating species to establish a population on the new island or continent. Overextended periods, with the immigration of species with strong flight abilities, islands are more likely to recruit species with stronger flight abilities.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

ABZ and YXZ conceived and designed the experiments, wrote, reviewed, and edited the manuscript. AH and YW reviewed and edited the manuscript. YXZ, JXC, and HL performed the flight mill experiments. YXZ, YW, and BYD analyzed the data. YXZ, QRH, ZL, and XHL performed the computer simulations. All authors contributed substantially to this research.

FUNDING

This work was supported by the Natural Science Foundation of China (Grant Nos. 31772501 and 32170421), the China National Funds for Distinguished Young Scientists (Grant No. 31425023), Support Project of High-level Teachers in Beijing Municipal Universities (Grant No. IDHT20180518), and the Academy for Multidisciplinary Studies, Capital Normal University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2021.771719/full#supplementary-material

REFERENCES

Aiello, B. R., Tan, M., Bin Sikandar, U., Alvey, A. J., Bhinderwala, B., Kimball, K. C., et al. (2021). Adaptive shifts underlie the divergence in wing morphology in bombycoid moths. *Proc. Biol. Sci.* 288:20210677. doi: 10.1098/rspb.2021.0677

Ávalos, J. A., Martí-Campoy, A., and Soto, A. (2014). Study of the flying ability of *Rhynchophorus ferrugineus* (Coleoptera: Dryophthoridae) adults using a computer-monitored flight mill. *Bull. Entomol. Res.* 104, 462–470. doi: 10.1017/S0007485314000121

Babu, A., Del Pozo-Valdivia, A. I., and Reisig, D. D. (2020). Baseline flight potential of *Euschistus servus* (Hemiptera: Pentatomidae) and its implications on local dispersal. *Environ. Entomol.* 49, 699–708. doi: 10.1093/ee/nvaa041

Chang, K. (2018). *Studies on the Relationship Between Flight Ability and Morphological Trait of Moths in the Field of Mount Dongling, Beijing (Insecta: Lepidoptera).* [Dissertation/Master's Thesis]. Beijing: Capital Normal University

Chapman, J. W., Reynolds, D. R., Hill, J. K., Sivell, D., Smith, A. D., and Woiwod, I. P. (2008a). A seasonal switch in compass orientation in a high-flying migrant moth. *Curr. Biol.* 18, 908–909. doi: 10.1016/j.cub.2008.08.014

Chapman, J. W., Reynolds, D. R., Mouritsen, H., Hill, J. K., Riley, J. R., Sivell, D., et al. (2008b). Wind selection and drift compensation optimize migratory pathways in a high-flying moth. *Curr. Biol.* 18, 514–518. doi: 10.1016/j.cub.2008.02.080

Chattopadhyay, P., Banerjee, G., and Mukherjee, S. (2017). Recent trends of modern bacterial insecticides for pest control practice in integrated crop management system. *Biotech* 7:60. doi: 10.1007/s13205-017-0717-6

Combes, S. A., Crall, J. D., and Mukherjee, S. (2010). Dynamics of animal movement in an ecological context: dragonfly wing damage reduces flight performance and predation success. *Biol. Lett*. 6, 426–429. doi: 10.1098/rsbl.2009.0915

Common, I. F. B. (1954). A study of the ecology of the adult bogong moth, *Agrotis infusa* (Boisd)(Lepidoptera: Noctuidae), with special reference to its behaviour during migration and aestivation. *Aust. J. Zool.* 2, 223–263. doi: 10.1071/ZO9540223

Cook, L. M., Grant, B. S., Saccheri, I. J., and Mallet, J. (2012). Selective bird predation on the peppered moth: the last experiment of Michael Majerus. *Biol. Lett*. 8, 609–612. doi: 10.1098/rsbl.2011.1136

Cook, L. M., and Saccheri, I. J. (2013). The peppered moth and industrial melanism: evolution of a natural selection case study. *Heredity* 110, 207–212. doi: 10.1038/hdy.2012.92

Dällenbach, L. J., Glauser, A., Lim, K. S., Chapman, J. W., and Menz, M. H. (2018). Higher flight activity in the offspring of migrants compared to residents in a migratory insect. *Proc. Biol. Sci.* 285, 20172829. doi: 10.5061/dryad.44hc2

Dreyer, D., Frost, B., Mouritsen, H., Günther, A., Green, K., Whitehouse, M., et al. (2018). The Earth's magnetic field and visual landmarks steer migratory flight behavior in the nocturnal Australian Bogong moth. *Curr. Biol.* 28, 2160–2166. doi: 10.1016/j.cub.2018.05.030

Eacock, A., Rowland, H. M., Edmonds, N., and Saccheri, I. J. (2017). Colour change of twig-mimicking peppered moth larvae is a continuous reaction norm that increases camouflage against avian predators. *PeerJ* 5:e3999. doi: 10.7717/peerj.3999

Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* 3, 294–299.

Fu, X. W., Chang, H., He, L. M., Zhao, S. Y., and Wu, K. M. (2017a). Flight performance of *Macdunnoughia crassisigna* (Lepidoptera: Noctuidae). *Bull. Entomol. Res.* 107, 715–723. doi: 10.1017/s0007485317000232

Fu, X. W., Zhao, S., Li, C., Wu, X., Guo, J., and Wu, K. (2017b). Flight performance of *Ctenoplusia agnata* (Lepidoptera: Noctuidae). *J. Econ. Entomol.* 110, 986–994. doi: 10.1093/jee/tox080

Furlong, M. J., Wright, D. J., and Dosdall, L. M. (2013). Diamondback moth ecology and management: problems, progress, and prospects. *Annu. Rev. Entomol.* 58, 517–541. doi: 10.1146/annurev-ento-120811-153605

Galan, M., Pons, J. B., Tournayre, O., Pierre, E., Leuchtmann, M., Pontier, D., et al. (2018). Metabarcoding for the parallel identification of several hundred predators and their prey: application to bat species diet analysis. *Mol. Ecol. Resour.* 18, 474–489. doi: 10.1111/1755-0998.12749

Gustafsson, K. M., Agrawal, A. A., Lewenstein, B. V., and Wolf, S. A. (2015). The monarch butterfly through time and space: the social construction of an icon. *BioScience* 65, 612–622. doi: 10.1093/biosci/biv045

Hao, M., Jin, Q., Meng, G., Yang, C., Yang, S., Shi, Z., et al. (2020). Regional assemblages shaped by historical and contemporary factors: Evidence from a species-rich insect group. *Mol. Ecol.* 29, 2492–2510. doi: 10.1111/mec.15412

Hashiyama, A., Nomura, M., Kurihara, J., and Toyoshima, G. (2013). Laboratory evaluation of the flight ability of female *Autographa nigrisigna* (Lepidoptera: Noctuidae), measured by actograph and flight mill. *J. Econ. Entomol.* 106, 690–694. doi: 10.1603/EC12272

Hoddle, M. S., and Hoddle, C. D. (2016). How far can the palm weevil, *Rhynchophorus vulneratus* (Coleoptera: Curculionidae), fly? *J. Econ. Entomol.* 109, 629–636. doi: 10.1093/jee/tov402

Hoddle, M. S., Hoddle, C. D., Faleiro, J. R., El-Shafie, H. A. F., Jeske, D. R., and Sallam, A. A. (2015). How far can the red palm weevil (Coleoptera: Curculionidae) fly?: computerized flight mill studies with field-captured weevils. *J. Econ. Entomol.* 108, 2599–2609. doi: 10.1093/jee/tov240

Hudgins, E. J., Liebhold, A. M., and Leung, B. (2017). Predicting the spread of all invasive forest pests in the United States. *Ecol. Lett*. 20, 426–435. doi: 10.1111/ele.12741

Jantzen, B., and Eisner, T. (2008). Hindwings are unnecessary for flight but essential for execution of normal evasive flight in Lepidoptera. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16636–16640. doi: 10.1073/pnas.0807223105

Jia, X. (2019). *A Preliminary Comparison of the Flight Abilities of Moths (Lepidoptera) in Different Habitats of Islands and Continents.* [Dissertation/Master's Thesis]. Beijing: Capital Normal University.

Jiang, C. (2018). *The Comparison of Flight Parameters of Common Lepidoptera in DongLing Mountain area of Beijing. [Dissertation/Master's Thesis]*. Beijing: Capital Normal University.

Jones, H. B., Lim, K. S., Bell, J. R., Hill, J. K., and Chapman, J. W. (2016). Quantifying interspecific variation in dispersal ability of noctuid moths using an advanced tethered flight technique. *Ecol. Evol.* 6, 181–190. doi: 10.1002/ece3.1861

Jyothi, P., Aralimarad, P., Wali, V., Dave, S., Bheemanna, M., Ashoka, J., et al. (2021). Evidence for facultative migratory flight behavior in *Helicoverpa armigera* (Noctuidae: Lepidoptera) in India. *PLoS One* 16:e0245665. doi: 10.1371/journal.pone.0245665

Liu, Z., McNeil, J. N., and Wu, K. (2011). Flight mill performance of the lacewing *Chrysoperla sinica* (Neuroptera: Chrysopidae) as a function of age, temperature, and relative humidity. *J. Econ Entomol.* 104, 94–100.

Lopez, V. M., Hoddle, M. S., Francese, J. A., Lance, D. R., and Ray, A. M. (2017). Assessing flight potential of the invasive Asian longhorned beetle (Coleoptera: Cerambycidae) with computerized flight mills. *J. Econ. Entomol.* 110, 1070–1077. doi: 10.1093/jee/tox046

MacDonald, Z. G., Acorn, J. H., Zhang, J., and Nielsen, S. E. (2019). Perceptual range, targeting ability, and visual habitat detection by greater fritillary butterflies *Speyeria cybele* (Lepidoptera: Nymphalidae) and *Speyeria atlantis*. *J. Insect. Sci.* 19:1. doi: 10.1093/jisesa/iez060

Menchetti, M., Guéguen, M., and Talavera, G. (2019). Spatio-temporal ecological niche modelling of multigenerational insect migrations. *Proc. Biol. Sci.* 286:20191583. doi: 10.1098/rspb.2019.1583

Nathan, R., Schurr, F. M., Spiegel, O., Steinitz, O., Trakhtenbrot, A., and Tsoar, A. (2008). Mechanisms of long-distance seed dispersal. *Trends Ecol. Evol.* 23, 638–647. doi: 10.1016/j.tree.2008.08.003

Noor, M. A., Parnell, R. S., and Grant, B. S. (2008). A reversible color polyphenism in American peppered moth (*Biston betularia cognataria*) caterpillars. *PLoS One* 3:e3142. doi: 10.1371/journal.pone.0003142

Ntelezos, A., Guarato, F., and Windmill, J. F. (2017). The anti-bat strategy of ultrasound absorption: the wings of nocturnal moths (Bombycoidea: Saturniidae) absorb more ultrasound than the wings of diurnal moths (Chalcosiinae: Zygaenoidea: Zygaenidae). *Biol. Open* 6, 109–117. doi: 10.1242/bio.021782

Panian, J., and Wiltschko, D. (2004). Ramp initiation in a thrust wedge. *Nat.* 427, 624–627. doi: 10.1038/nature02334

Pearl, R., and Slobodkin, L. (1976). The growth of populations. *Q. Rev. Biol.* 51, 6–24. doi: 10.1086/394288

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available online at: https://www.R-project.org

Ratnasingham, S., and Hebert, P. D. (2007). BARCODING: bold: the barcode of life data system (http://www.barcodinglife.org). *Mol. Ecol. Notes* 7, 355–364.

Reppert, S. M., Guerra, P. A., and Merlin, C. (2016). Neurobiology of monarch butterfly migration. *Annu. Rev. Entomol.* 61, 25–42. doi: 10.1146/annurev-ento-010814-020855

Rivero, A., Giron, D., and Casas, J. (2001). Lifetime allocation of juvenile and adult nutritional resources to egg production in a holometabolous insect. *Proc. R. Soc. Lond. B Biol. Sci.* 268, 1231–1237. doi: 10.1098/rspb.2001.1645

Ross, A. (2017). Insect evolution: the origin of wings. *Curr. Biol.* 27, R113–R115.
Rubin, J. J., Hamilton, C. A., McClure, C. J., Chadwell, B. A., Kawahara, A. Y., and Barber, J. R. (2018). The evolution of anti-bat sensory illusions in moths. *Sci. Adv*. 4:eaar7428. doi: 10.1126/sciadv.aar7428
Sappington, T. W., and Burks, C. S. (2014). Patterns of flight behavior and capacity of unmated navel orangeworm (Lepidoptera: Pyralidae) adults related to age, gender, and wing size. *Environ. Entomol*. 43, 696–705. doi: 10.1603/EN13279
Stork, N. E. (2018). How many species of insects and other terrestrial arthropods are there on earth? *Annu. Rev. Entomol*. 63, 31–45. doi: 10.1146/annurev-ento-020117-043348
Tanaka, K., and Yamanaka, T. (2009). Factors affecting flight activity of *Ophraella communa* (Coleoptera: Chrysomelidae), an exotic insect in Japan. *Environ. Entomol*. 38, 235–241. doi: 10.1603/022.038.0129
Telang, A., Li, Y., Noriega, F. G., and Brown, M. R. (2006). Effects of larval nutrition on the endocrinology of mosquito egg development. *J. Exp Biol*. 209, 645–655. doi: 10.1242/jeb.02026
Tihelka, E., Cai, C., Giacomelli, M., Lozano-Fernandez, J., Rota-Stabelli, O., Huang, D., et al. (2021). The evolution of insect biodiversity. *Curr. Biol.* 31, R1299–R1311.
Triant, D. A., Cinel, S. D., and Kawahara, A. Y. (2018). Lepidoptera genomes: current knowledge, gaps and future directions. *Curr. Opin. Insect. Sci*. 25, 99–105. doi: 10.1016/j.cois.2017.12.004
Wang, J. (1980). The relationship between sea level rising and climate change on east of China (in Chinese with English abstract). *Chin. J. Geogr*. 35, 299–313.
Warrant, E., Frost, B., Green, K., Mouritsen, H., Dreyer, D., Adden, A., et al. (2016). The Australian Bogong moth *Agrotis infusa*: a long-distance nocturnal navigator. *Front. Behav. Neurosci.* 10:77. doi: 10.3389/fnbeh.2016.00077
Whittaker, R. J., Fernández-Palacios, J. M., Matthews, T. J., Borregaard, M. K., and Triantis, K. A. (2017). Island biogeography: taking the long view of nature's laboratories. *Sci*ence 357:eaam8326. doi: 10.1126/science.aam8326
Willmott, A. P., and Ellington, C. P. (1997). The mechanics of flight in the hawkmoth *Manduca sexta.* II. Aerodynamic consequences of kinematic and morphological variation. *J. Exp. Biol*. 200, 2723–2745.
Yack, J. E., Smith, M. L., and Weatherhead, P. J. (2001). Caterpillar talk: acoustically mediated territoriality in larval Lepidoptera. *Proc. Natl. Acad. Sci. U.S.A*. 98, 11371–11375. doi: 10.1073/pnas.191378898
Yadav, C., Smith, M., Ogunremi, D., and Yack, J. (2020a). Draft genome assembly and annotation of the masked birch caterpillar, *Drepana arcuata* (Lepidoptera: Drepanoidea). *Data Brief*. 33:106531. doi: 10.1016/j.dib.2020.106531
Yadav, C., Smith, M. L., and Yack, J. E. (2020b). Transcriptome analysis of a social caterpillar, *Drepana arcuata*: de novo assembly, functional annotation and developmental analysis. *PLoS One* 15:e0234903. doi: 10.1371/journal.pone.0234903
Yang, C., Han, H., Chi, M., Jin, Q., Wu, C., Zhu, C., et al. (2012). Species identification of *Noctuidae moths* (Insecta: Lepidoptera) from Baihuashan, Beijing, China with DNA barcoding. *Acta Entomol. Sin*. 55, 1082–1092.
Zeale, M. R., Butlin, R. K., Barker, G. L., Lees, D. C., and Jones, G. (2011). Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. *Mol. Ecol. Resour*. 11, 236–244. doi: 10.1111/j.1755-0998.2010.02920.x
Zhuang, H., Yago, M., Settele, J., Li, X., Ueshima, R., Grishin, N. V., et al. (2018). Species richness of Eurasian *Zephyrus* hairstreaks (Lepidoptera: Lycaenidae: Theclini) with implications on historical biogeography: An NDM/VNDM approach. *PLoS One* 13:e0191049. doi: 10.1371/journal.pone.0191049

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 03 February 2022
doi: 10.3389/fevo.2021.775414

Lack of Genetic Structure Among Populations of Striped Flea Beetle *Phyllotreta striolata* (Coleoptera: Chrysomelidae) Across Southern China

Qian Li[1], Guang-Mei Li[1], Yong-Li Zheng[2] and Shu-Jun Wei[3]**

[1] College of Life Sciences, China Jiliang University, Hangzhou, China, [2] Zhejiang Agricultural Products Quality and Safety Center, Hangzhou, China, [3] Institute of Plant and Environmental Protection, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China

OPEN ACCESS

Edited by:
Charles K. Lee, University of Waikato, New Zealand

Reviewed by:
Ruth Freire Alvarez, Heinrich Heine University of Düsseldorf, Germany
Ian D. Hogg, Polar Knowledge Canada (POLAR), Canada

***Correspondence:**
Yong-Li Zheng
yonglizheng@yeah.net
Shu-Jun Wei
shujun268@163.com

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *14 September 2021*
Accepted: *29 December 2021*
Published: *03 February 2022*

Citation:
Li Q, Li G-M, Zheng Y-L and Wei S-J (2022) Lack of Genetic Structure Among Populations of Striped Flea Beetle Phyllotreta striolata (Coleoptera: Chrysomelidae) Across Southern China. Front. Ecol. Evol. 9:775414. doi: 10.3389/fevo.2021.775414

The striped flea beetle (SFB) *Phyllotreta striolata* (Fabricius) (Coleoptera: Chrysomelidae) is a major pest of cruciferous vegetables in southern China. The population diversity and genetic structure of SFB are unknown. Here, we assembled a draft genome for the SFB and characterized the distribution of microsatellites. Then, we developed 12 novel microsatellite markers across the genome. We used a segment of the *cox1* gene and newly developed microsatellite markers to genotype the genetic diversity of SFB across southern China. There were 44 mitochondrial haplotypes in the SFB populations, with haplotype 2 as the most widespread. The population genetic differentiation was very low, indicated by F_{ST}-values (<0.05 except for Guangxi population with other populations based on *cox1*), high gene flow (4.10 and 44.88 of *cox1* and microsatellite, respectively) and Principal Coordinate Analysis across all populations. Mantel test showed genetic distance in SFB was significantly associated with geographic distance based on microsatellites (R^2 = 0.2373, P = 0.014) while result based on *cox1* (R^2 = 0.0365, P = 0.155) showed no significant difference. The phylogenetic analysis did not find any geographically related clades among all haplotypes. Analyses based on microsatellites showed a lack of population genetic structure among all populations. Our study provides a foundation for the future understanding of the ecology and evolution of SFB and its management.

Keywords: *Phyllotreta striolata*, southern China, population genetic structure, microsatellite, mitochondrial gene

INTRODUCTION

The striped flea beetle (SFB) *Phyllotreta striolata* (Fabricius) (Coleoptera: Chrysomelidae) is a pest of cruciferous crops (Brassicaceae) (Soroka et al., 2018; Cao et al., 2020; Atirach et al., 2021). The adults chew on leaves and larvae feed on fiber. The field control of this pest heavily relys on insecticides (Andersen et al., 2006), leading to insecticide resistance (Feng et al., 2000; James et al., 2019).

The SFB is mainly distributed in Asia, Europe, and North America (Soroka et al., 2018; Cao et al., 2020; Atirach et al., 2021). Native to Eurasia, SFB was introduced to North America in 1663 from Europe (Rousseau and Lesage, 2016). In China, the SFB is one of the most important cruciferous crops pests in southern areas. When the adult density of SFB is 20∼50 individuals every 100 *Brassica pekinensis*, the damage rate will be 50∼76%; if the adult density more than 50, the damage rate will be 100% in Shenzhen (Zhang et al., 2000). There are six to nine generations of SFB throughout the year without overwintering in southern China (Sun et al., 2010), compared to one to two generations per year in North America (Olfert et al., 2017). Recently, damage of SFB has increased. Its geographic range expanded both in China and North America (Chai, 2010; Sun, 2010; Lee et al., 2011; Kielen, 2012). Current studies mainly focus on this pest's biology and control methods (Gao et al., 2000; Chai, 2010; Lee et al., 2011; Soroka et al., 2018; Yan et al., 2018; Atirach et al., 2021); however, no studies that we are aware of on the ecology and evolution of the SFB.

In this study, we examined the population genetic diversity and genetic structure of the SFB among populations across its main distribution range in China. Due to the lack of genetic markers, we developed a novel set of microsatellite markers from first-time assembled random genomic sequences of SFB using Illumina sequencing. We used both microsatellites and a segment of the mitochondrial cytochrome oxidase subunit I (*cox1*) gene to determine the genetic diversity and differentiation among eight representative populations across southern China. We tested the hypothesis that levels of genetic diversity and population differentiation for the SFB are low due to high levels of gene flow among populations. These results provide a basis to understand the ecology and evolution of this pest and its management.

MATERIALS AND METHODS

Sample Collection and DNA Extraction

Adults of SFB were collected across southern China from cruciferous vegetables (**Table 1** and **Figure 1**). To avoid collecting siblings, we collected specimens from about 20 sites at least 10 m apart at each sampling location. In total, eight populations were collected. The species were first identified by morphology (He et al., 2012) and then validated by molecular identification using the mitochondrial *cox1* gene (see below). All samples were kept in 100% ethanol before DNA extraction.

To construct a high-throughput sequencing library for microsatellites development, we extracted genomic DNA from 50 adults collected in Xiaoshan, Zhejiang province, using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). Then, we used 192 individuals for population-level genotyping, with 24 individuals from each population. Total genomic DNA was extracted from individual whole adult using the DNeasy Blood & Tissue Kit. The voucher specimens were stored at –80°C in the Integrated Pest Management Laboratory of the Beijing Academy of Agriculture and Forestry Sciences.

Library Construction, Genome Sequencing, and Assembly

The high-throughput sequencing library with 500-bp insert size was prepared using the Illumina TruSeq DNA PCR-Free HT Library Prep Kit (Illumina, San Diego, CA, United States). The prepared library was sequenced on an Illumina Hiseq4000 Sequencer using the Hiseq Reagent Kit v3 (Illumina, San Diego, CA, United States) by Beijing BerryGenomics Co., Ltd. The paired-end 150 bp raw data were trimmed by removing the low-quality reads using Trimmomatic 0.36 (Bolger et al., 2014), and then the sequences were evaluated by FastQC v 0.11.5 (Andrews, 2004). The genome size of *P. striolata* was estimated by JELLYFISH v2.2.6 software with a K-mer method (Kingsford, 2011). IDBA_UD v1.1.1 was used to assemble the generated genomic sequences with K-mer from 20 to 140 (Peng et al., 2010).

Development of Universal Microsatellite Markers

The Microsatellite Search and Building Database (MSDB)[1] was used to identify potential microsatellite sites from the whole genome sequences of *P. striolata*. A minimum of 12, 5, 5, 5, 5, and 5 repeats was used to distinguish mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide motifs, respectively (Lianming et al., 2013). The QDD3 program (Emese et al., 2010) was then used to extract microsatellites and their flanking sequences (300 bp each) and design the primers. The criteria and parameters for the primer design for the isolated microsatellite markers followed Cao et al. (2016). The primer design was set according to the following parameters: annealing temperature (Tm): 54°C < Tm < 60°C; the difference in Tm between paired upstream and downstream primers was < 4°C; only one pair of microsatellite primers is kept on each locus; the repeat motif of the amplified product was ≥ 3. The microsatellite loci were further filtered under stringent criteria as described in previous studies (Song et al., 2018): (i) the microsatellites had to be pure and specific, (ii) the design strategy of "A" was used, and (iii) the minimum distance between the 3′ end of a primer pair and its target region had to be no shorter than 10 bp.

We selected 70 pairs of primers for initial validation using one individual from each representative population (AH, FJ, GD, GX, HN, JS, JX, ZJ). We used three fluorescences (FAM, ROX, HEX) to label the amplified products of each locus (Schuelke, 2000; Blacket et al., 2012). The fluorescences were independently added to a primer C tail (PC-tail) (5′ CAGGACCAGGCTACCGTG 3′). The sequence of the PC-tail was added to the 5′ end of each upstream primer (Blacket et al., 2012). The reaction system was set to 10 μL, including 0.5 μL of template DNA (30–160 ng/μL), 5 μL of Master Mix (Promega, Madison, WI, United States), 0.08 μL of the PC tail modified forward primer (10 mM), 0.16 μL of reverse primer (10 mM), 0.32 μL of fluorescence-labeled PC-tail (10 mM) and 3.94 μL of ddH_2O. The following PCR amplification program was used to amplify the microsatellites: 3 min at 94°C; 35 cycles of 15 s at 94°C, 30 s at 56°C and 1 min at

[1] http://msdb.biosv.com/

TABLE 1 | Population information of *Phyllotreta striolata* used in this study.

Population	Collection location	Longitude (°E), Latitude (°N)	Collection date (month/year)	Number of haplotypes
AH	Xuanzhou, Xuancheng of Anhui province	118.89, 30.90	10/2020	8
FJ	Yunxiao, Zhangzhou Fujian province	117.58, 23.95	10/2020	4
GD	Gaoyao, Zhaoqing of Guangdong province	112.58, 23.05	10/2020	6
GX	Yizhou, Hechi of Guangxi province	108.83, 24.49	10/2020	6
HN	Xiuying, Haikou of Hainan province	110.25, 19.98	10/2020	7
JS	Huishan, Wuxi of Jiangsu province	120.24, 31.66	10/2020	6
JX	Shanggao, Yichun of Anhui province	114.90, 28.23	10/2020	4
ZJ	Xiaoshan, Hangzhou of Zhejiang province	120.63, 30.21	04/2016	3

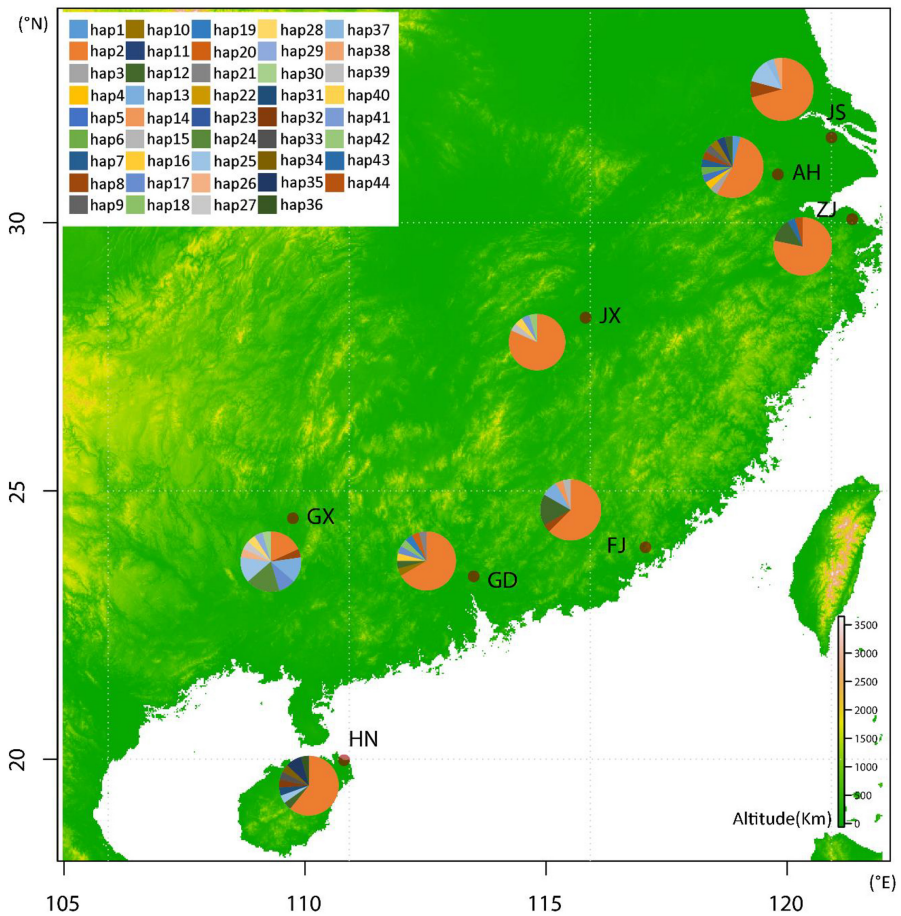


FIGURE 1 | Collection sites and geographical distribution of the mitochondrial *cox1* haplotypes of *Phyllotreta striolata*. The haplotype frequency for each population is represented by colors in the pie charts. See **Table 1** for population codes.

72°C, followed by a final 10 min extension at 72°C. Primer pairs fitting the following criteria were retained in subsequent analysis: (i) primers with PCR amplification rate equal or higher than 75%; (ii) at least two alleles present in tested individuals; (iii) without non-specific amplification.

Length of PCR products was analyzed on an ABI 3730xl DNA Analyzer (Applied Biosystems, United States) using the GeneScan^{TM} 500 LIZ^{TM} dye Size Standard (Applied Biosystems, United States) for sizing DNA fragments. Genotypes were determined using GENEMAPPER v4.0 (Applied Biosystems, United States) (Chatterji and Pachter, 2006). The remaining primer pairs were used for population-level genotyping.

Mitochondrial Gene Amplification, Sequencing, and Homologous Gene Downloading

To characterize the mitochondrial variation and validate correct identification of the specimens, a fragment of the *cox1* gene involved the DNA barcoding region of insects

was sequenced using the universal primers LCO1490 (5′ GGTCAACAAATCATAAAGATATTGG 3′) and HCO2198 (5′ TAAACTTCAGGGTGACCAAAAAATCA 3′) (Folmer et al., 1994). Polymerase chain reactions (PCR) were conducted using the Mastercycler Pro system (Eppendorf, Germany) in a 15 μL volume consisting of 1 μL of template DNA (30–160 ng/μL), 7.5 μL of Master Mix (Promega, Madison, WI, United States), 0.6 μL of forward primer (10 mM), 0.6 μL of reverse primer (10 mM) and 5.3 μL of ddH_2O. The thermal profiles for DNA amplification were as follows: 3 min at 94°C; 35 cycles of 15 s at 94°C, 20 s at 54°C, and 40 s at 72°C, followed by a final 10-min extension at 72°C. Amplified products were purified and sequenced on an ABI 3730xl DNA Analyzer by Tsingke Biotechnology Co. Ltd. (Beijing, China).

Apart from sequences obtained in this study, we searched orthologous *cox1* gene sequences from the NCBI nucleotide database and included those with sampling locations in our phylogenetic analysis (Pentinsaari et al., 2014; Hendrich et al., 2015; Nie et al., 2017; Coral şahin et al., 2018; Lalrinfeli et al., 2019) (**Supplementary Table 1**). Sequences with obvious errors were deleted, and others were used for follow-up analysis.

Population Genetic Diversity Analysis

For mitochondrial DNA, sequencing results from both strands were assembled using SeqMan in the LASERGENE version 7.1.2 (DNASTAR, Inc., United States). Sequences of the *cox1* were aligned with MUSCLE (Edgar, 2004) implemented in MEGA version X (Kumar et al., 2018). The number of polymorphic sites (S), number of haplotypes (H), haplotype diversity (H_d), nucleotide diversity (Pi), Tajima's *D*, Fu's F_s and gene flow (N_m) were analyzed in DnaSP version 6 (Rozas et al., 2017). Pairwise mean population differentiation (F_{ST}) was estimated using Arlequin 3.5 (Excoffier and Lischer, 2010).

For microsatellites, the number of alleles and the observed heterozygosity (H_o) were analyzed using the macros Microsatellite Tools (Park, 2001). The null allele frequencies and F_{ST} with excluding of null alleles (ENA) were estimated using the software FreeNA (Chapuis and Estoup, 2007). Deviation from Hardy-Weinberg equilibrium (HWE) for each locus/population combination, F_{ST} and inbreeding coefficients (F_{IS}) were estimated in GENEPOP v4.0.11 (Rousset, 2008). Principal Coordinate Analysis (PCoA) and genetic differentiation coefficient (G_{ST}) were performed with the GenAlex program ver. 6.502 (Peakall and Smouse, 2012). The gene flow was calculated using the formula $N_m = 0.5 (1- G_{ST})/G_{ST}$. HP-RARE v1.1 (Kalinowski, 2010) was used to test allelic richness (A_R) and allelic richness of private alleles (P_{AR}) of each site. GENCLONE v2.0 (Arnaud-Haond and Belkhir, 2006) was used to estimate the total number of alleles (A_T) and the unbiased expected heterozygosity (H_E). We compared the number of alleles (A_S) among samples with different sample sizes in GENCLONE. Mantel test (Mantel, 1967), used to estimate correlation between genetic and geographic distances of populations to test isolation by distance, was performed with 1,000 permutations based on Genepop.[2]

Phylogenetic and Population Genetic Structure Analysis

A Bayesian-inference phylogenetic tree was constructed for mitochondrial DNA using Mrbayes v3.2.2 (Ronquist et al., 2012) to examine phylogenetic relationships among *cox1* haplotypes of SFB. Four independent Markov chains for 50 million MCMC generations were run with tree sampling every 5,000 generations and a burn-in of 2,500 trees.

For microsatellite data, the population genetic structure was first investigated using the STRUCTURE v2.3.4 program (Pritchard et al., 2000). We used 30 replicates of each *K*-value from 1 to 8, with 200,000 Markov chain Monte Carlo iterations and a burn-in of 100,000 iterations. The results were uploaded to the online software Structure Harvester v0.6.94 (Earl and Vonholdt, 2012) to determine the optimal *K*-value by a Delta K method (last accessed on 28, Dec. 2020).[3] The *K*-value with the highest ΔK represents the number of potential genetic clusters in the population. Membership coefficient matrices (Q-matrices) associated with the optimal K were processed using CLUMPP v1.12 (Jakobsson and Rosenberg, 2007) and then visualized using the DISTRUCT v1.1 (Taubert et al., 2019). Finally, we used discriminant analysis of principal component (DAPC) to analyze population genetic structure under default settings to complement the STRUCTURE analysis. This analysis was run using an R-package adegenet v1.4-2 (Jombart et al., 2008).

RESULTS

Genome Assembly and Characterization of Microsatellites Across the Genome

A total of 47.88 Gb paired-end (PE) sequences (159,598,840 reads each with a length of 150 bp) were obtained. Trimmed reads were assembled into 595,192 scaffolds with a total length of 342.147 MB ranging from 200 bp to 118.198 KB with an N50 of 600 bp. These contigs were used for microsatellite discovery.

In total, 28,512 microsatellites were isolated from the randomly sequenced genome sequences of SFB with 8,635 mononucleotide repeat sites (30.29%), 6,735 (23.62%) dinucleotide repeat (DNR) sites, 8,115 (28.46%) trinucleotide repeat (TNR) sites, 3,550 (12.4%) tetranucleotide repeat (TTNR) sites, 905 (3.17%) pentanucleotide repeats (PNR) sites and 572 (2.00%) hexanucleotide repeat (HNR) sites. In decreased order, A, AC, AT, and AAT were the most frequent repeat motifs. The microsatellites with the three most frequently repeated motifs occupied 51.9% of all microsatellites. The average length of the repeat region ranged from 14.4 to 49.44 bp.

Development of Microsatellite Markers

The QDD3 program designed 2,585 primer pairs for 1,994 din-, 359 tri-, 70 tetra-, 29 penta-, and 21 hexa-nucleotide

[2]https://genepop.curtin.edu.au/

[3]http://taylor0.biology.ucla.edu/structureHarvester/

microsatellites. Following our stringent filter, 70 primer pairs flanking perfect microsatellites were retained. The number of primer pairs flanking tri-, tetra-and penta-nucleotide microsatellites was 67, 1, and 2, respectively. In our initial test of these 70 primer pairs, 16 pairs generated polymorphic genotypes, 12 pairs failed to amplify in any samples, and 46 pairs failed to amplify in more than two individuals. In the final test, 12 microsatellite markers were retained for population-level genotyping (**Table 2**).

Population Genetic Diversity

Based on the mitochondrial *cox1* genes which were 647 bp long, we found 44 haplotypes from the collected SFB populations. Haplotype #2 was the most common (61.17% of individuals). Thirty-four haplotypes appeared in only one population. The GX population had the most haplotypes with thirteen, while the ZJ population had the least number of haplotypes (n = 4). The haplotype diversity ranged from 0.338 to 0.931, and the nucleotide diversity ranged from 0.00059 to 0.00497. There were 11 polymorphic sites in the HN population which represented the highest number across the most in all populations (**Table 3**). The data of Tajima's D and Fu's F_s of each population were all negative except for F_S of GX (F_S = 0.57 without statistical significance). For all populations Tajima's D and Fu's F_s were –2.53 and –6.01, respectively, with statistical significance, which showed the SFB populations followed the neutral evolution model and there was a population expansion in recent history. The Nei's N_m of haplotypes and all sequences was 4.10 and 3.37, respectively.

Using all *cox1* sequences from the public database, haplotypes were reanalyzed and all populations were divided into thirty-four haplotypes. Haplotypes #30–34 were unique to populations outside China. Haplotype #2 was shared by populations in and out of China. The other twenty-eight haplotypes were unique to the populations of China. The pairwise distance among all haplotypes ranged from 0.0018 to 0.0143 (**Table 4**). We also calculated the pairwise population genetic differentiation among eight populations and F_{ST}-values ranged from –0.0187 (FJ-ZJ) to 0.1374 (JX-GX). A moderate genetic differentiation was observed between GX and other populations based on the F_{ST}-values ranging from 0.0367 to 0.1374 (**Table 5**).

TABLE 2 | Twelve microsatellite markers developed for *Phyllotreta striolata* in this study.

Locus	Motif	Forward primer	Reverse primer	Size (bp)	FL
PS-03	$(CCG)_9$	AGGTAGGCACTAGATTGGCC	GTGAGCGGCGAAGACAAATG	123	FAM
PS-04	$(AAC)_7$	CAACGCAAAGTTCGCCTGAA	ACCAACACCTTCCCTTCAGTG	124	FAM
PS-06	$(AAG)_7$	AGCAGGATAAGGATTAAGAGTGC	TGGAAGATGCAGTAAGTGTTAACA	138	FAM
PS-07	$(CCG)_7$	TCCAACACCATTAGCGCTCT	GTGAGTGATGCTCCGTCAGT	138	FAM
PS-08	$(ATC)_7$	CGGACGATGAGGACTTTCATC	CCCTCATCCTCTTCTTGTTCTTG	139	HEX
PS-09	$(ATC)_7$	TTTCCAATTTAGCGAGCTCGT	TCGAATGCGATCGTTTAGGAG	140	HEX
PS-13	$(ATC)_7$	AGTGTACCTTGGCTGTCAATAAA	TCTTCATTTCGAATGAGGCCTC	145	HEX
PS-18	$(ATC)_7$	AATTGGTCGTCGACGTCTCC	CACCAAGAGGAGCAACCGG	157	HEX
PS-20	$(ACC)_7$	CGCTCGTCAGTCCGACTTAT	ATGCCGGTGTTGAGCAAGTT	163	ROX
PS-21	$(AAC)_7$	GGTGACCTCTGGCAGAAACA	TGGTTGTTGCGATGTAGCTCT	164	ROX
PS-22	$(CCG)_{10}$	CCCGGTAGTCAACTCGAGTG	CCCAGCGGCCGGTATATTC	168	ROX
PS-33	$(AAC)_8$	AATTGACCGGCAGCGATTTG	GTGGCGGTACAAGCTCAAAC	193	ROX

FL, fluorescent label.

TABLE 3 | Population genetic diversity of *Phyllotreta striolata* collected from southern China.

Population	Mitochondrial DNA							Microsatellite loci						
	N	H	H_d	Pi	S	Tajima's D	Fu's F_S	A_R	P_{AR}	A_T	A_S	Ho	H_E	F_{IS}
AH	24	11	0.717	0.00239	7	−2.10*	−3.05*	4.25	0.08	50	39.49	0.443	0.577	0.236
FJ	24	8	0.594	0.00211	6	−1.65	−1.36	3.83	0.25	43	34.26	0.450	0.533	0.159
GD	24	9	0.565	0.00439	7	−2.26**	−3.68**	4.58	0.42	47	37.66	0.507	0.576	0.125
GX	24	13	0.931	0.00497	8	−0.10	0.57	3.83	0.08	30	30.00	0.389	0.486	0.203
HN	23	9	0.636	0.00260	11	−2.07*	−2.47	4.17	0	44	34.22	0.387	0.512	0.273
JS	24	5	0.493	0.00103	4	−1.52	−1.61	4.58	0.25	48	37.22	0.426	0.560	0.237
JX	22	5	0.338	0.00059	3	−1.88*	−2.86*	4.00	0.08	43	35.83	0.412	0.513	0.201
ZJ	23	4	0.383	0.00118	3	−1.88*	−2.90*	4.42	0.25	48	37.11	0.437	0.555	0.217

*N, number of individuals used in the analysis; H, number of haplotypes; H_d, haplotype diversity; Pi, nucleotide diversity; S, number of polymorphic sites; A_R, allelic richness for 24 individuals per population; P_{AR}, private allelic richness for 24 individuals per population; AT, total number of alleles; As, standardized total number of alleles (for 7 individuals); Ho, observed heterozygosity; H_E, expected heterozygosity; F_{IS}, inbreeding coefficient. *Means the difference is significant at the 0.05 level. **Means the difference is significant at the 0.01 level.*

TABLE 4 | Pairwise genetic distance among 34 mitochondrial *cox1* haplotypes.

	hap1	hap2	hap3	hap4	hap5	hap6	hap7	hap8	hap9	hap10	hap11	hap12	hap13	hap14	hap15	hap16	hap17	hap18	hap19	hap20	hap21	hap22	hap23	hap24	hap25	hap26	hap27	hap28	hap29	hap30	hap31	hap32	hap33	hap34
hap1		0.036	0.057	0.062	0.056	0.052	0.061	0.056	0.111	0.057	0.057	0.059	0.057	0.090	0.061	0.061	0.082	0.062	0.086	0.061	0.124	0.116	0.101	0.062	0.061	0.058	0.059	0.083	0.059	0.064	0.081	0.114	0.058	0.059
hap2	0.032		0.035	0.037	0.035	0.036	0.035	0.036	0.086	0.035	0.035	0.035	0.034	0.062	0.037	0.036	0.058	0.037	0.061	0.035	0.087	0.087	0.073	0.037	0.037	0.038	0.034	0.060	0.038	0.087	0.058	0.156	0.037	0.084
hap3	0.069	0.032		0.060	0.056	0.051	0.057	0.060	0.110	0.055	0.056	0.058	0.055	0.090	0.060	0.057	0.080	0.057	0.086	0.055	0.117	0.115	0.096	0.058	0.059	0.058	0.056	0.084	0.059	0.119	0.084	0.202	0.059	0.110
hap4	0.069	0.032	0.069		0.059	0.055	0.059	0.061	0.117	0.058	0.057	0.061	0.059	0.089	0.062	0.060	0.088	0.059	0.087	0.060	0.122	0.121	0.101	0.062	0.061	0.062	0.063	0.084	0.060	0.124	0.082	0.246	0.059	0.119
hap5	0.069	0.032	0.069	0.069		0.051	0.060	0.056	0.116	0.058	0.057	0.059	0.059	0.087	0.059	0.054	0.036	0.056	0.086	0.058	0.118	0.116	0.053	0.059	0.056	0.058	0.054	0.037	0.060	0.115	0.081	0.121	0.058	0.062
hap6	0.066	0.032	0.066	0.066	0.066		0.052	0.055	0.075	0.053	0.054	0.054	0.052	0.076	0.053	0.054	0.073	0.055	0.076	0.051	0.105	0.075	0.089	0.054	0.055	0.058	0.053	0.074	0.054	0.099	0.072	0.178	0.053	0.097
hap7	0.069	0.032	0.069	0.069	0.069	0.066		0.058	0.114	0.055	0.057	0.057	0.056	0.083	0.057	0.059	0.085	0.058	0.083	0.060	0.118	0.114	0.051	0.057	0.060	0.058	0.054	0.090	0.058	0.118	0.083	0.238	0.058	0.121
hap8	0.069	0.032	0.069	0.069	0.069	0.066	0.069		0.122	0.057	0.059	0.056	0.055	0.093	0.061	0.060	0.080	0.059	0.088	0.056	0.116	0.117	0.098	0.063	0.060	0.060	0.057	0.084	0.061	0.113	0.083	0.201	0.058	0.107
hap9	0.160	0.111	0.160	0.160	0.160	0.105	0.160	0.160		0.119	0.120	0.115	0.112	0.159	0.114	0.118	0.165	0.118	0.169	0.114	0.271	0.114	0.191	0.118	0.122	0.126	0.118	0.166	0.125	0.229	0.166	0.657	0.123	0.257
hap10	0.069	0.032	0.069	0.069	0.069	0.066	0.069	0.069	0.160		0.057	0.058	0.059	0.085	0.058	0.058	0.082	0.058	0.089	0.057	0.118	0.122	0.098	0.059	0.059	0.063	0.055	0.083	0.063	0.115	0.083	0.216	0.060	0.116
hap11	0.069	0.032	0.069	0.069	0.069	0.066	0.069	0.069	0.160	0.069		0.057	0.056	0.087	0.062	0.057	0.083	0.058	0.090	0.056	0.112	0.120	0.099	0.061	0.059	0.060	0.055	0.082	0.060	0.120	0.081	0.207	0.057	0.115
hap12	0.069	0.032	0.069	0.069	0.069	0.066	0.069	0.069	0.160	0.069	0.069		0.057	0.086	0.057	0.056	0.084	0.061	0.082	0.058	0.114	0.059	0.100	0.058	0.059	0.060	0.058	0.087	0.058	0.115	0.081	0.212	0.057	0.119
hap13	0.069	0.032	0.069	0.069	0.069	0.066	0.069	0.069	0.160	0.069	0.069	0.069		0.088	0.057	0.055	0.086	0.057	0.082	0.055	0.115	0.119	0.100	0.060	0.059	0.057	0.054	0.088	0.057	0.111	0.081	0.205	0.057	0.114
hap14	0.111	0.069	0.111	0.111	0.111	0.105	0.111	0.111	0.217	0.111	0.111	0.111	0.111		0.086	0.085	0.115	0.088	0.119	0.091	0.174	0.186	0.136	0.036	0.089	0.082	0.086	0.128	0.089	0.184	0.120	0.382	0.089	0.170
hap15	0.069	0.032	0.069	0.069	0.069	0.066	0.069	0.069	0.160	0.069	0.069	0.069	0.069	0.111		0.058	0.083	0.059	0.038	0.060	0.119	0.120	0.098	0.060	0.058	0.059	0.059	0.087	0.065	0.058	0.086	0.232	0.064	0.126
hap16	0.069	0.032	0.069	0.069	0.069	0.066	0.069	0.069	0.160	0.069	0.069	0.069	0.069	0.111	0.069		0.081	0.057	0.086	0.057	0.119	0.118	0.096	0.058	0.060	0.064	0.056	0.079	0.060	0.120	0.085	0.118	0.060	0.110
hap17	0.111	0.069	0.111	0.111	0.032	0.105	0.111	0.111	0.217	0.111	0.111	0.111	0.111	0.160	0.111	0.111		0.080	0.120	0.084	0.164	0.166	0.075	0.081	0.083	0.082	0.079	0.058	0.083	0.155	0.108	0.172	0.082	0.086
hap18	0.069	0.032	0.069	0.069	0.069	0.066	0.069	0.069	0.160	0.069	0.069	0.069	0.069	0.111	0.069	0.069	0.111		0.089	0.059	0.124	0.130	0.096	0.059	0.062	0.061	0.058	0.082	0.062	0.118	0.085	0.222	0.060	0.116
hap19	0.111	0.069	0.111	0.111	0.111	0.105	0.111	0.111	0.217	0.111	0.111	0.111	0.111	0.160	0.032	0.111	0.160	0.111		0.089	0.168	0.164	0.131	0.086	0.088	0.088	0.085	0.122	0.095	0.085	0.118	0.346	0.093	0.215
hap20	0.069	0.032	0.069	0.069	0.069	0.066	0.069	0.069	0.160	0.069	0.069	0.069	0.069	0.111	0.069	0.069	0.111	0.069	0.111		0.124	0.122	0.105	0.063	0.059	0.060	0.055	0.085	0.058	0.122	0.084	0.222	0.058	0.119
hap21	0.160	0.111	0.160	0.160	0.160	0.151	0.160	0.160	0.286	0.160	0.160	0.160	0.160	0.217	0.160	0.160	0.217	0.160	0.217	0.160		0.285	0.137	0.116	0.116	0.119	0.115	0.223	0.121	0.232	0.152	0.245	0.114	0.237
hap22	0.160	0.111	0.160	0.160	0.160	0.105	0.160	0.160	0.160	0.160	0.160	0.069	0.160	0.217	0.160	0.160	0.217	0.160	0.217	0.160	0.286		0.175	0.126	0.118	0.129	0.117	0.166	0.119	0.235	0.146	0.487	0.112	0.242
hap23	0.151	0.105	0.151	0.151	0.066	0.146	0.066	0.151	0.266	0.151	0.151	0.151	0.151	0.204	0.151	0.151	0.105	0.151	0.204	0.151	0.204	0.266		0.096	0.096	0.099	0.094	0.077	0.098	0.186	0.128	0.206	0.098	0.108
hap24	0.069	0.032	0.069	0.069	0.069	0.066	0.069	0.069	0.160	0.069	0.069	0.069	0.069	0.032	0.069	0.069	0.111	0.069	0.111	0.069	0.160	0.160	0.151		0.061	0.057	0.058	0.090	0.060	0.123	0.087	0.221	0.060	0.118
hap25	0.069	0.032	0.069	0.069	0.069	0.066	0.069	0.069	0.160	0.069	0.069	0.069	0.069	0.111	0.069	0.069	0.111	0.069	0.111	0.069	0.160	0.160	0.151	0.069		0.062	0.061	0.079	0.060	0.114	0.085	0.213	0.060	0.113
hap26	0.069	0.032	0.069	0.069	0.069	0.066	0.069	0.069	0.160	0.069	0.069	0.069	0.069	0.111	0.069	0.069	0.111	0.069	0.111	0.069	0.160	0.160	0.151	0.069	0.069		0.059	0.083	0.062	0.111	0.084	0.225	0.059	0.115
hap27	0.069	0.032	0.069	0.069	0.069	0.066	0.069	0.069	0.160	0.069	0.069	0.069	0.069	0.111	0.069	0.069	0.111	0.069	0.111	0.069	0.160	0.160	0.151	0.069	0.069	0.069		0.081	0.058	0.117	0.082	0.212	0.057	0.115
hap28	0.111	0.069	0.111	0.111	0.032	0.105	0.111	0.111	0.217	0.111	0.111	0.111	0.111	0.160	0.111	0.111	0.069	0.111	0.160	0.111	0.217	0.217	0.105	0.111	0.111	0.111	0.111		0.086	0.167	0.117	0.167	0.086	0.086
hap29	0.069	0.032	0.069	0.069	0.069	0.066	0.069	0.069	0.160	0.069	0.069	0.069	0.069	0.111	0.069	0.069	0.111	0.069	0.111	0.069	0.160	0.160	0.151	0.069	0.069	0.069	0.069	0.111		0.128	0.087	0.224	0.063	0.119
hap30	0.069	0.111	0.160	0.160	0.160	0.151	0.160	0.160	0.286	0.160	0.160	0.160	0.160	0.217	0.069	0.160	0.217	0.160	0.111	0.160	0.286	0.286	0.266	0.160	0.160	0.160	0.160	0.217	0.160		0.086	0.228	0.061	0.123
hap31	0.111	0.069	0.111	0.111	0.111	0.105	0.111	0.111	0.217	0.111	0.111	0.111	0.111	0.160	0.111	0.111	0.160	0.111	0.160	0.111	0.217	0.217	0.204	0.111	0.111	0.111	0.111	0.160	0.111	0.111		0.376	0.036	0.154
hap32	0.160	0.217	0.286	0.286	0.160	0.266	0.286	0.286	0.471	0.286	0.286	0.286	0.286	0.368	0.286	0.160	0.217	0.286	0.368	0.286	0.286	0.471	0.266	0.286	0.286	0.286	0.286	0.217	0.286	0.286	0.368		0.227	0.061
hap33	0.069	0.032	0.069	0.069	0.069	0.066	0.069	0.069	0.160	0.069	0.069	0.069	0.069	0.111	0.069	0.069	0.111	0.069	0.111	0.069	0.160	0.160	0.151	0.069	0.069	0.069	0.069	0.111	0.069	0.069	0.032	0.286		0.114
hap34	0.069	0.111	0.160	0.160	0.069	0.151	0.160	0.160	0.286	0.160	0.160	0.160	0.160	0.217	0.160	0.160	0.111	0.160	0.217	0.160	0.286	0.286	0.151	0.160	0.160	0.160	0.160	0.111	0.160	0.160	0.217	0.069	0.160	

Data of from the upper triangle refer to the standard error.

TABLE 5 | Pairwise population genetic differentiation (F_{ST}, lower triangle) and *P*-values (upper triangle) among eight populations of *Phyllotreta striolata* based on haplotypes of the *cox1* gene.

Population	AH	FJ	GD	GX	HN	JS	JX	ZJ
AH		0.6937	0.8739	0.0180	0.1171	0.0270	0.0811	0.7478
FJ	–0.0119		0.4775	0.0811	0.1261	0.0090	0.0360	0.6757
GD	–0.0129	–0.0012		0.0360	0.0991	0.0090	0.1081	0.4865
GX	**0.0520**	0.0367	**0.0458**		0.0000	0.0000	0.0000	0.0000
HN	0.0160	0.0171	0.0157	**0.0678**		0.3514	0.1261	0.2523
JS	**0.0462**	**0.0598**	**0.0365**	**0.1105**	0.0038		0.1441	0.0270
JX	0.0487	**0.0577**	0.0302	**0.1374**	0.0162	0.0197		0.1351
ZJ	–0.0187	–0.0062	–0.0062	**0.0807**	0.0126	**0.0433**	0.0405	

F_{ST}-values with significant differences indicated in bold.

TABLE 6 | Pairwise population genetic differentiation (F_{ST}, lower triangle) and *P*-values (upper triangle) among eight populations of *Phyllotreta striolata* based on microsatellites.

Population	AH	FJ	GD	GX	HN	JS	JX	ZJ
AH		0.0155	0.0072	<0.0010	<0.0010	0.3549	0.0227	0.0947
FJ	**0.0092 (0.0071)**		0.0159	0.0103	0.0371	<0.0010	0.7549	0.0021
GD	**0.0147 (0.0135)**	**0.0045 (0.0033)**		0.2558	0.0827	<0.0010	0.0690	0.0550
GX	**0.0392 (0.0340)**	**0.0291 (0.0246)**	0.0203 (0.0187)		0.7473	0.0203	0.0767	0.4836
HN	**0.0298 (0.0301)**	**0.0049 (0.0066)**	0.0003 (0.0034)	0.0032 (0.0027)		0.0091	0.6325	0.3365
JS	0.0081 (0.0092)	**0.0406 (0.0291)**	**0.0419 (0.0341)**	**0.0458 (0.0324)**	**0.0379 (0.0250)**		0.0030	0.1428
JX	**0.0136 (0.01147)**	–0.0125 (–0.0092)	**–0.0011 (–0.0007)**	0.0256 (0.0219)	–0.0042 (–0.0012)	**0.0345 (0.0245)**		0.3211
ZJ	0.0049 (0.0055)	**0.0119 (0.0098)**	0.0038 (0.0041)	0.0091 (0.0069)	0.0006 (0.0029)	0.0057 (0.00413)	0.0048 (0.0049)	

F_{ST}-values were also calculated with excluding of null alleles (data in parentheses). F_{ST}-values with significant difference were in bold.

Based on the 12 developed microsatellite markers, we evaluated the genetic diversity in eight populations of SFB. Significant departures from HWE ($p < 0.05$) were detected in 53 of the 96 population-locus combinations after sequential Bonferroni correction. The population-locus pairs colonized by primer PS-22 all showed significant departures from HWE. Thirty-five of the 528 locus-locus pairs showed linkage disequilibrium in at least one population ($p < 0.05$), whereas five (S08-S21, S20-S13, S09-S22, S22-S33, and S07-S33 pair) of 66 locus pairs showed linkage disequilibrium across all populations.

Genetic diversity parameters varied among populations. The observed (H_o) and expected (H_E) heterozygosity values ranged from 0.387 to 0.507 and from 0.486 to 0.577, respectively. The total number of alleles was the highest in population AH, with a value of 50. In calculating the standardized total number of alleles, we removed locus S06 because of the failure of amplification in the population JX. The inbreeding coefficient of all populations ranged from 0.125 to 0.273 (**Table 3**). F_{ST}-values ranged from –0.0125 (JX-FJ) to 0.0458 (JS-GX). F_{ST}-values of populations pairs JX-FJ, JX-GD, and JX-GX are below 0 (**Table 6**). F_{ST}-values had significant differences among 15 populations (AH-FJ, AH-GD, AH-GX, AH-HN, AH-JX, FJ-GD, FJ-GX, FJ-HN, FJ-JS, FJ-ZJ, GD-JS, GD-JX, GX-JS, HN-JS, and JS-JX). The G_{ST} of total alleles is 0.011 with a N_m of 44.88, indicating high geneflow among SFB populations. We also calculated the F_{ST}-values with excluding of ENA but the results are similar to that without ENA (**Table 6**). Populations GX and JS have the highest geographical distance and genetic distance. However, the JX and FJ have the nearest geographical distance but not the least genetic distance. The Principal Coordinate Analysis (PCoA) showed no obvious geographic structure (**Figure 2**). The first and second principal coordinates explained 21.78 and 16.65% of the total variation, respectively.

To determine whether the genetic distance in SFB is significantly associated with geographic distance, we conducted Mantel tests between pairwise genetic differentiation $[F_{ST}/(1 - F_{ST})]$ and geographical distance matrix based on haplotypes and microsatellites (**Figure 3**). The relationships of genetic distance and geographic distance were not identical based on microsatellites ($R^2 = 0.2373$, $P = 0.014$) and mtDNA ($R^2 = 0.0365$, $P = 0.155$).

Population Genetic Structure and Phylogenetic Relationship Analysis

We reconstructed the phylogenetic tree based on thirty-two haplotypes (**Figure 4**). Haplotypes from different geographical regions were mixed in the phylogenetic tree, which was consistent with PCoA analysis. We did not find any geographically related clades with low supporting data, except for one clade composed of two haplotypes (30 and 32) unique to populations outside of China, and another clade composed of three haplotypes (28, 29, and 31) unique to populations outside of China.

The STRUCTURE analysis indicated that the optimal *K*-value was two. There was a lack of population structure when K increased from 2 to 3 (**Figure 5A**). DAPC analysis revealed similar patterns to those obtained from the

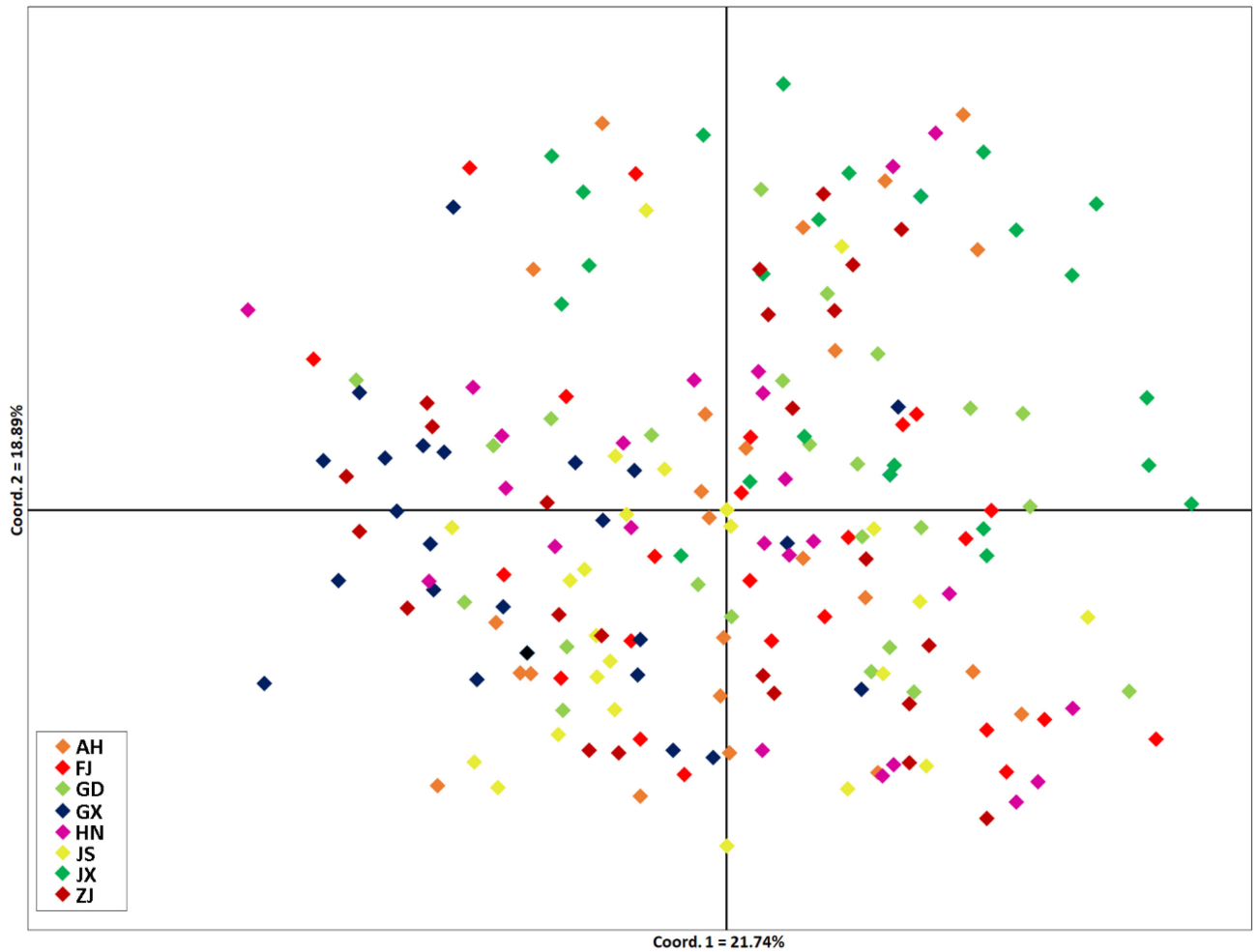


FIGURE 2 | PCoA based on pairwise Nei's genetic distance estimates for the populations.

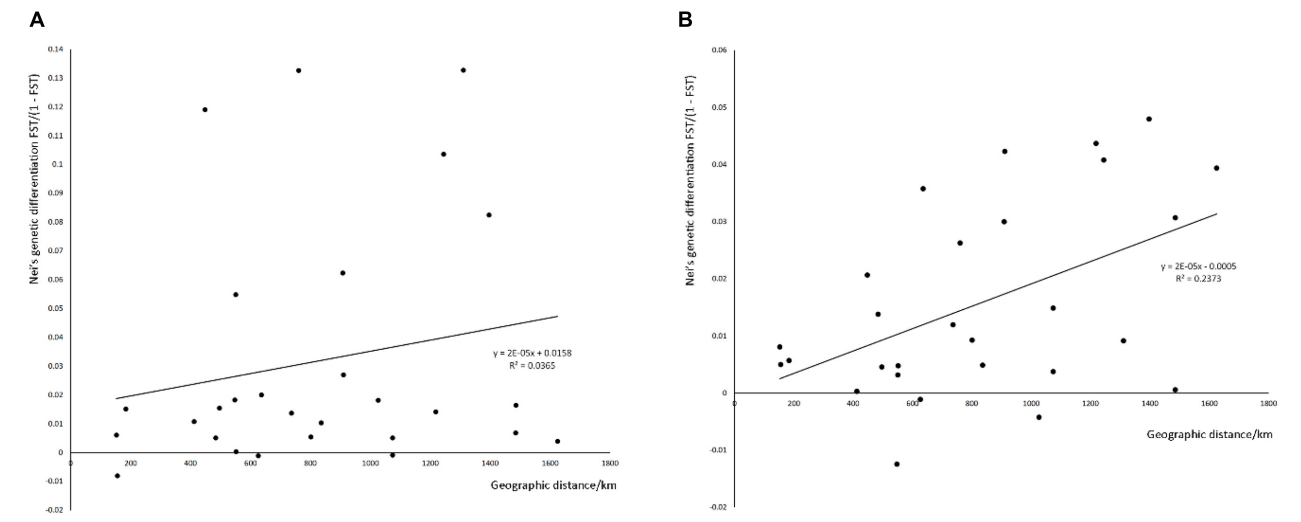


FIGURE 3 | Mantel test of the geographic distance and Nei's genetic distance based on haplotypes of *cox1* gene **(A)** and microsatellites **(B)**.

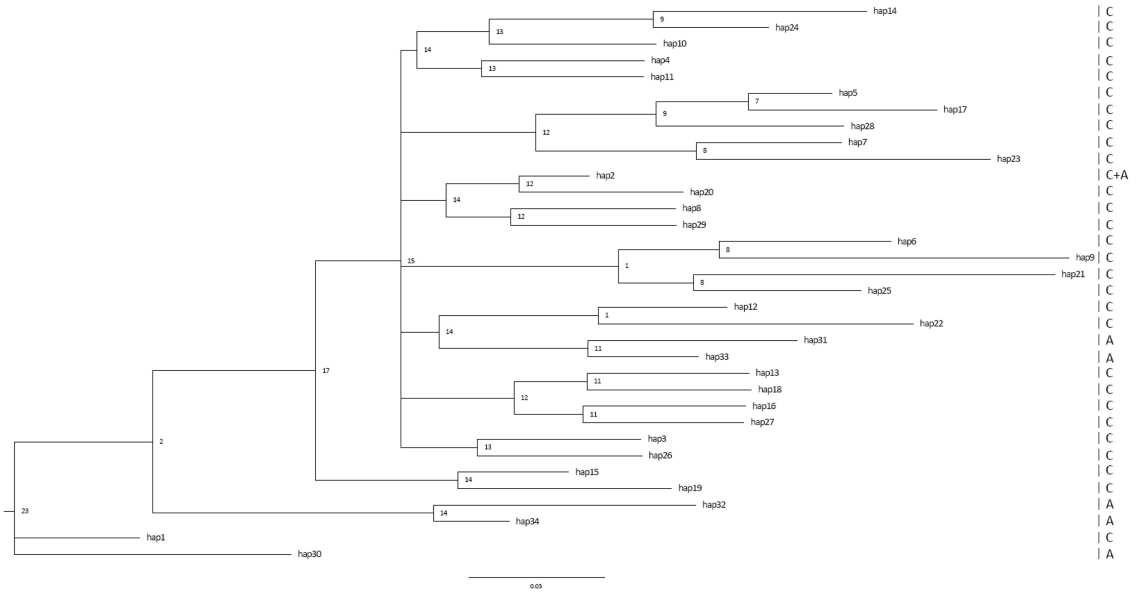


FIGURE 4 | Bayesian-inference phylogenetic relationships among mitochondrial *cox1* haplotypes of *Phyllotreta striolata* worldwide. C, haplotypes present in China. A, haplotypes present outside China.

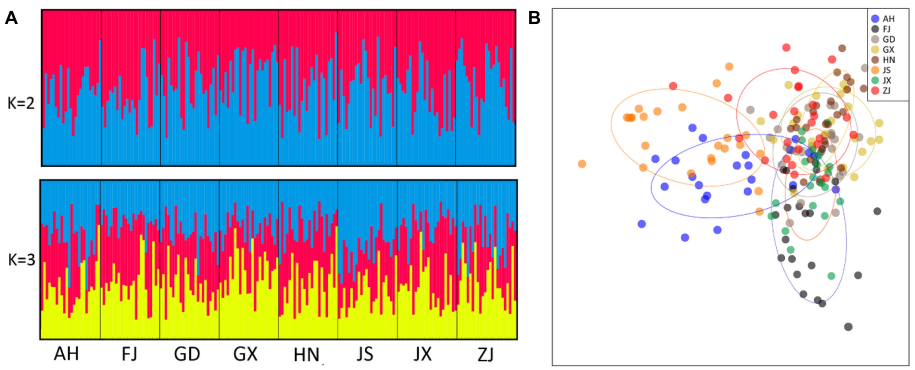


FIGURE 5 | Genetic structure of *Phyllotreta striolata* populations from southern China inferred from STRUCTURE **(A)** and DAPC **(B)** analyses based on 12 microsatellite loci. The optimal number of clusters for the STRUCTURE analysis was three; *K* = 2 and 3 are presented **(A)** arranged by population. DAPC analysis shows all populations are clumping together **(B)**.

STRUCTURE clustering analyses with all populations clumping together (**Figure 5B**).

DISCUSSION

This study developed universal microsatellite markers of SFB based on a *de novo* assembled draft genome. Both mitochondrial DNA and microsatellite markers point to the low population genetic diversity of SFB in southern China. The haplotype analyses using partial mitochondrial *cox1* genes also confirms the low genetic diversity in Europe and Asia since only two haplotypes were identified from the public databases. A lack of population structure was supported by population genetic structure analysis results using STRUCTURE and DAPC.

The mononucleotide repeat sites included four types as $(A)_n$, $(T)_n$, $(C)_n$, and $(G)_n$. The number of $(A)_n$ and $(T)_n$ far exceeded that of $(C)_n$ and $(G)_n$ in SFB, which is similar to other eukaryotic species (Katti et al., 2001). $(A)_n$, $(AT)_n$, and $(AAT)_n$ were the most frequent motifs in SFB, as in other beetle species (Song et al., 2020). The frequency and density of microsatellite markers varied among species (Selkoe and Toonen, 2010; Xu et al., 2017; Liu et al., 2019). The dominant repeats in SFB are dinucleotides except for an occasional mononucleotide. Interestingly, this differs from other beetle species that have been sequenced in the Family Chrysomelidae, in which the dominant repeats are trinucleotide or tetranucleotide repeats in *Leptinotarsa decemlineata* (Coleoptera, Chrysomelidae) and *Diabrotica virgifera* (Coleoptera, Chrysomelidae), respectively (Kim et al., 2008; Liu et al., 2018).

In general, the results based on microsatellites and haplotypes of the *cox1* gene support the notion that the sampled populations of SFB have low genetic variation and lack population genetic structure in southern China, as suggested by multiple lines of analyses such as pairwise F-statistics, PCoA, DAPC, STRUCTURE and phylogenetic reconstruction. The lack of population genetic structure was found in some species, such as recently introduced species *Obolodiplosis robiniae* and *Frankliniella occidentalis* (Shang et al., 2015; Cao et al., 2017) or migratory species like *Plutella xylostella*, whose Sichuan populations are sources populations for northern immigrants, and southern China and Yunnan populations are sources populations for central-eastern populations (Chen et al., 2020). In order to escape suboptimal environmental or weather conditions, some insects have annual migrations (Zhan et al., 2011). In Coleoptera, some species, like *Leptinotarsa decemlineata* and *Pachysternum capense* can migrate (Ferro et al., 1991; Greń and Górz, 2020). We did not find a records of migration events for the of SFB and studied have shown that it has been established in Eurasia for several centuries (Rousseau and Lesage, 2016). Our hypothesis, that the low levels of genetic diversity and population differentiation for the SLB are due to high levels of gene flow among populations, needs to be further tested.

Consistently low variation among groups is indicative of high levels of gene flow across the geographic range as seen microsatellite data. Pairwise comparisons of F_{ST}-values from mtDNA data are also consistent with high gene flow across the range, with generally low F_{ST}-values between populations. Haplotypes (n = 2) widely exist in all populations also suggest a high-level of gene flow among 8 populations. Cruciferous vegetables are a kind of common vegetable in China and is the preferential host of SFB. The high gene flow maybe due to the transportation of fresh vegetable from south to north in China, such as from Anhui Province to Zhejiang Province. The passive transportation accelerates gene flow among different populations (Martinez-Hernandez et al., 2021). The high geneflow, however, may be the evidence of its migration (Seymour et al., 2016; Yang et al., 2020).

But there are also some divergences between these two types of molecular markers. A moderate genetic differentiation was observed between GX and other populations based on genetic distances of mtDNA, which were not observed based on microsatellites. Because mtDNA is sensitive to founder effects and small population size, the probable loss or gain of a mtDNA haplotype will be greater for small populations (Roderick, 1996). Furthermore, the Mantel test of microsatellites suggests evidence of a positive relationship in genetic distances (F_{ST}) between genetic and geographic distances, while the result based on mtDNA had no significant difference. This may be caused by the use of only a single marker of mtDNA vs. 12 markers of microsatellite. Different parts of mtDNA evolve at different rates, thus different parts of mitochondrial DNA should be considered for future studies to find higher-level population differentiation (Avise et al., 1992). The significant Mantel test suggests that while genetic differences between populations are rather small, they tend to accumulate as geographic distances increase, supporting the non-negligible population structure.

We found that the population collected from Guangxi province of southwest China had the highest number of unique haplotypes. Many species in East Asia were found to have originated from southwestern China and expanded their distribution range northward during the interglacial periods in the Quaternary (Dong et al., 2013; Wei et al., 2015; Yang et al., 2020). Our study did not include samples collected from southwestern areas such as Yunnan and Sichuan. The inclusion of Guangxi populations and the relatively higher level of genetic diversity in this population may indicate that the SFB may also have originated from southwestern China and subsequently dispersed to other areas.

CONCLUSION

In conclusion, our study provides a draft genome and a set of microsatellites developed from the genome. Based on these microsatellites and a segment of mitochondrial gene, we investigated the population genetic structure of the SFB in southern China, where it is a common pest. We found a lack of population differentiation of SFB among populations in southern China, which might be caused by high frequencies of gene flow among populations. Either passive transportation or migration events will accelerate the gene flow among different populations. Species collected from Guangxi Province showed more genetic divergence than others based on mitochondrial gene which indicates that the SFB may originate from southwestern China. In consideration of the limited collection range, the hypothesized movements need further investigation.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, PRJNA761897. Data are also available on the Barcode of Life Datasystems (BOLD) database under Project–SFB "Barcoding of *Phyllotreta striolata* in southern China" (doi: 10.5883/DS-188PSLQ).

AUTHOR CONTRIBUTIONS

QL analyzed the data and wrote the article. QL and G-ML collected the data. Y-LZ and S-JW revised the article. All authors contributed significantly to the drafts and gave final approval for publication.

FUNDING

This research was funded by the China Agriculture Research System of MOF and MARA (CARS-23-C05), and "Three rural issues and six participants" of Scientific and Technological Cooperation Projects of Zhejiang Province-9#. The integration, demonstration and promotion of green technology of prevention and control of *Phyllotreta striolata* (CTZB-F180706LWZ-SNY1).

ACKNOWLEDGMENTS

We thank Lina Sun and Lijun Cao for their help in experiments and data analyses. We are also very grateful to all collectors for collecting specimens.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2021.775414/full#supplementary-material

REFERENCES

Andersen, C. L., Hazzard, R., VanDriesche, R., and Mangan, F. X. (2006). Alternative management tactics for control of *Phyllotreta cruciferae* and *Phyllotreta striolata* (Coleoptera: Chrysomelidae) on *Brassica rapa* in Massachusetts. *J. Econ. Entomol.* 99, 803–810. doi: 10.1603/0022-0493-99.3.803

Andrews, S. (2004). FastQC: a quality control tool for high throughput sequence data. *Babraham Bioinformatics*. Available online at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arnaud-Haond, S., and Belkhir, K. (2006). GENCLONE: a computer program to analyse genotypic data, test for clonality and describe spatial clonal organization. *Mol. Ecol. Notes* 7, 15–17.

Atirach, N., Sirirut, M., and Lewis, E. E. (2021). Biological control potential of entomopathogenic nematodes against the striped flea beetle, *Phyllotreta sinuata* Stephens (Coleoptera: Chrysomelidae). *Crop Prot.* 141:105448.

Avise, J. C., Alisauskas, R. T., Nelson, W. S., and Ankney, C. D. (1992). Matriarchal population genetic structure in an avian species with female natal philopatry. *Evolution* 46, 1084–1096. doi: 10.1111/j.1558-5646.1992.tb00621.x

Blacket, M. J., Robin, C., Good, R. T., Lee, S. F., and Miller, A. D. (2012). Universal primers for fluorescent labelling of PCR fragments–an efficient and cost-effective approach to genotyping by fluorescence. *Mol. Ecol. Resour.* 12, 456–463. doi: 10.1111/j.1755-0998.2011.03104.x

Bolger, A. M., Marc, L., and Bjoern, U. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Cao, C., Huang, D., Yao, J., Zhu, Z., Zheng, J., Zhou, R., et al. (2020). Field application techniques for control of *Phyllotreta striolata* with microbial insecticides on radish. *Chin. J. Biol. Control.* 36, 987–991.

Cao, L. J., Li, Z. M., Wang, Z. H., Zhu, L., Gong, Y. J., Chen, M., et al. (2016). Bulk development and stringent selection of microsatellite markers in the western flower thrips *Frankliniella occidentalis*. *Sci. Rep.* 6:26512. doi: 10.1038/srep26512

Cao, L. J., Wang, Z. H., Gong, Y. J., Zhu, L., and Wei, S. J. (2017). Low genetic diversity but strong population structure reflects multiple introductions of western flower thrips(Thysanoptera: Thripidae) into China followed by human-mediated spread. *Evol. Appl.* 10, 391–401. doi: 10.1111/eva.12461

Chai, W. (2010). The habit and control method of *Phyllotreta striolata* (Fabricius) on *Brassica napus* in Hexizoulang in Gan su Provence, China. *China Plant Prot.* 30, 23–24.

Chapuis, M. P., and Estoup, A. (2007). Microsatellite null alleles and estimation of population differentiation. *Mol. Biol. Evol.* 24, 621–631. doi: 10.1093/molbev/msl191

Chatterji, S., and Pachter, L. (2006). Reference based annotation with GeneMapper. *Genome Biol.* 7:R29. doi: 10.1186/gb-2006-7-4-r29

Chen, M. Z., Cao, L., Li, B., Chen, J., and Wei, S. (2020). Migration trajectories of the diamondback moth *Plutella xylostella* in China inferred from population genomic variation. *Pest Manag. Sci.* 77, 1683–1693. doi: 10.1002/ps.6188

Coral şahin, D., Magoga, G., Özdikmen, H., and Montagna, M. (2018). DNA barcoding as useful tool to identify crop pest flea beetles of Turkey. *J. Appl. Entomol.* 143, 105–117. doi: 10.1111/jen.12566

Dong, L., Heckel, G., Liang, W., and Zhang, Y. (2013). Phylogeography of *Silver pheasant* (*Lophura nycthemera* L.) across China: aggregate effects of refugia, introgression and riverine barriers. *Mol. Ecol.* 22, 3376–3390. doi: 10.1111/mec.12315

Earl, D. A., and Vonholdt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Emese, M., Caroline, C., Vincent, D., André, G., and Thibaut, M. (2010). QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics* 26, 403–404. doi: 10.1093/bioinformatics/btp670

Excoffier, L., and Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10, 564–567. doi: 10.1111/j.1755-0998.2010.02847.x

Feng, H. T., Huang, Y. J., and Hsu, J. C. (2000). Insecticide susceptibility of cabbage flea beetle (*Phyllotreta striolata* (Fab.)) in Taiwan. *Plant Prot. Bull.* 42, 67–72.

Ferro, D. N., Tuttle, A. F., and Weber, D. C. (1991). Ovipositional and flight behavior of overwintered Colorado potato beetle (Coleoptera: Chrysomelidae). *Environ. Entomol.* 20, 1309–1314. doi: 10.1093/ee/20.5.1309

Folmer, O., Black, M., Wr, H., Lutz, R., and Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome C oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* 3, 294–299.

Gao, Z. Z., Wu, W. J., and Cui, Z. X. (2000). Study the host range of *Phyllotreta striolata* (Fabricius). *Ecol. Sci.* 19, 70–72.

Greń, C., and Górz, A. (2020). Coprophagous hydrophilid beetles (Coleoptera, Hydrophilidae, Sphaeridiinae) distribution in the Polish Carpathians. *Insects* 11:355. doi: 10.3390/insects11060355

He, H. L., Bin, S. Y., and Lin, J. T. (2012). The research progress of biology, ecology characters and occurrence reasons of *Phyllotreta striolata*. *J. Anhui Agric. Sci.* 40, 10683–10686.

Hendrich, L., Morinière, J., Haszprunar, G., Hebert, P. D. N., Hausmann, A., Köhler, F., et al. (2015). A comprehensive DNA barcode database for Central European beetles with a focus on Germany: adding more than 3500 identified species to BOLD. *Mol. Ecol. Resour.* 15, 795–818. doi: 10.1111/1755-0998.12354

Jakobsson, M., and Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23, 1801–1806. doi: 10.1093/bioinformatics/btm233

James, M., Adam, M. A., and Thomas, P. K. (2019). Flea beetle (Coleoptera: Chrysomelidae) populations, effects of feeding injury, and efficacy of insecticide treatments on eggplant and cabbage in southwest Virginia. *J. Econ. Entomol.* 113, 887–895. doi: 10.1093/jee/toz355

Jombart, T., Devillard, S., Dufour, A. B., and Pontier, D. (2008). Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity* 101, 92–103. doi: 10.1038/hdy.2008.34

Kalinowski, S. T. (2010). Hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol. Ecol. Notes* 5, 187–189. doi: 10.1111/j.1471-8286.2004.00845.x

Katti, M. V., Ranjekar, P. K., and Gupta, V. S. (2001). Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Mol. Biol. Evol.* 18, 1161–1167. doi: 10.1093/oxfordjournals.molbev.a003903

Kielen, A. (2012). *Flea Beetles: Population are Shifting*. Winnipeg, MB: Alberta Farmer.

Kim, K. S., Ratcliffe, S. T., French, B. W., Liu, L., and Sappington, T. W. (2008). Utility of EST-derived SSRs as population genetics markers in a beetle. *J. Heredity* 99, 112–124. doi: 10.1093/jhered/esm104

Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27, 764–770. doi: 10.1093/bioinformatics/btr011

Kumar, S., Steched, G., Li, M., Knyaz, C., Tamura, K., and Notes, A. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Lalrinfeli, R., Behere, G., Firake, D., Sharma, B., Banerjee, A., and Rajesh, T. (2019). Development of DNA barcodes for major insect pest of cole crops in Mid-Hills of Meghalaya. *Int. J. Curr. Microbiol. Appl. Sci.* 8, 789–799.

Lee, C. F., Chang, H. Y., Wang, C. L., and Chen, W. S. (2011). A review of *Phyllotreta* Chevrolat in Taiwan (Coleoptera: Chrysomelidae: Galerucinae: Alticini). *Zool. Stud.* 50, 525–533.

Lianming, D., Yuzhi, L., Xiuyue, Z., and Bisong, Y. (2013). MSDB: a user-friendly program for reporting distribution and building databases of microsatellites from genome sequences. *J. Heredity* 104, 154–157.

Liu, M., Wang, X., Ma, L., Cao, L., Liu, H., Pu, D., et al. (2019). Genome-wide developed microsatellites reveal a weak population differentiation in the hoverfly *Eupeodes corollae* (Diptera: Syrphidae) across China. *PLoS One* 14:e0215888. doi: 10.1371/journal.pone.0215888

Liu, Y., Fu, K., He, J., and Guo, W. (2018). Verfication SSR primers by datamining genome SSR loci in *Leptinotarsa decemlineata*. *J. Environ. Entomol.* 40, 633–644.

Martinez-Hernandez, F., Villalobos, G., and Martínez-Ibarra, J. A. (2021). Population structure and genetic diversity of *Triatoma longipennis* (Usinger, 1939) (Heteroptera: Reduviidae: Triatominae) in Mexico. *Infect. Genet. Evol.* 89:104718.

Mantel, N. A. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27, 209–220.

Nie, R. E., Breeschoten, T., Timmermans, M. J. T. N., Nadein, K., Xue, H. J., Bai, M., et al. (2017). The phylogeny of Galerucinae (Coleoptera: Chrysomelidae) and the performance of mitochondrial genomes in phylogenetic inference compared to nuclear rRNA genes. *Cladistics* 33, 1–18. doi: 10.1111/cla.12196

Olfert, O., Weiss, R. M., Soroka, J. J., and Elliott, R. H. (2017). Bioclimatic approach to assessing factors influencing shifts in geographic distribution and relative abundance of two flea beetle species (Coleoptera: Chrysomelidae) in North America. *Can. Entomol.* 150, 1–14.

Park, S. D. E. (2001). *Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection*. Ph.D. thesis. Dublin: University of Dublin.

Peakall, R., and Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics* 28, 2537–2539. doi: 10.1093/bioinformatics/bts460

Peng, Y., Leung, H. C. M., Yiu, S. M., and Chin, F. Y. L. (2010). "IDBA – a practical iterative de Bruijn graph de novo assembler," in *Proceedings of the 14th Annual international conference on Research in Computational Molecular Biology* (Berlin: Springer).

Pentinsaari, M., Hebert, P. D. N., Mutanen, M., and Fontaneto, D. (2014). Barcoding beetles: a regional survey of 1872 species reveals high identification success and unusually deep interspecific divergences. *PLoS One* 9:e108651. doi: 10.1371/journal.pone.0108651

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1093/genetics/155.2.945

Roderick, G. K. (1996). Geographic structure of insect populations: gene flow, phylogeography, and their uses. *Annu. Rev. Entomol.* 41, 325–352. doi: 10.1146/annurev.en.41.010196.001545

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Hohna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

Rousseau, M., and Lesage, L. (2016). Earliest North American occurrence of *Phyllotreta striolata* (Coleoptera: Chrysomelidae) from Québec, Canada. *Can. Entomol.* 148, 476–478.

Rousset, F. (2008). Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* 8, 103–106. doi: 10.1111/j.1471-8286.2007.01931.x

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J., Guirao-Rico, S., Librado, P., Ramos-Onsins, S., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248

Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* 18, 233–234. doi: 10.1038/72708

Selkoe, K. A., and Toonen, R. J. (2010). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecol. Lett.* 9, 615–629. doi: 10.1111/j.1461-0248.2006.00889.x

Seymour, M., Perera, O. P., Fescemyer, H. W., Jackson, R. E., Fleischer, S. J., and Abel, C. A. (2016). Peripheral genetic structure of *Helicoverpa zea* indicates asymmetrical panmixia. *Ecol. Evol.* 6, 3198–3207. doi: 10.1002/ece3.2106

Shang, X., Yao, Y., Huai, W., and Zhao, W. (2015). Population genetic differentiation of the black locust gall midge *Obolodiplosis robiniae* (Haldeman) (Diptera: Cecidomyiidae): a North American pest invading Asia. *Bull. Entomol. Res.* 105, 736–742. doi: 10.1017/S000748531500070X

Song, W., Cao, L., Li, B., Gong, Y., Hoffmann, A. A., and Wei, S. (2018). Multiple refugia from penultimate glaciations in East Asia demonstrated by phylogeography and ecological modelling of an insect pest. *BMC Evol. Biol.* 18:152. doi: 10.1186/s12862-018-1269-z

Song, X., Yang, T., Yan, X., Zheng, F., Xu, X., and Zhou, C. (2020). Comparison of microsatellite distribution patterns in twenty-nine beetle genomes. *Gene* 757:144919. doi: 10.1016/j.gene.2020.144919

Soroka, J., Grenkow, L., Otani, J., Gavloski, J., and Olfert, O. (2018). Flea beetle (Coleoptera: Chrysomelidae) species in canola (Brassicaceae) on the northern Great Plains of North America. *Can. Entomol.* 150, 100–115. doi: 10.1603/0022-0493(2008)101[159:dipcap]2.0.co;2

Sun, J., Xian, Z., Liu, X., and Liu, Y. (2010). The cause of serious occurrence of *Phyllotreta striolata* in Guangxi Nanning and its control measures. *Guangxi Agric. Sci.* 41, 21–23.

Sun, Z. (2010). Occurrence regularity and control measures of *Phyllotreta striolata* in Chaidamu area. *Qinghai Agro Technol. Ext.* 4, 38–39.

Taubert, O., Reinartz, I., Meyerhenke, H., and Schug, A. (2019). diSTruct v1.0: generating biomolecular structures from distance constraints. *Bioinformatics* 35, 5337–5338. doi: 10.1093/bioinformatics/btz578

Wei, S., Cao, L., Gong, Y., Shi, B., Wang, S., Zhang, F., et al. (2015). Population genetic structure and approximate Bayesian computation analyses reveal the southern origin and northward dispersal of the oriental fruit moth *Grapholita molesta* (Lepidoptera: Tortricidae) in its native range. *Mol. Ecol.* 24, 4094–4111. doi: 10.1111/mec.13300

Xu, Y., Li, W., Hu, Z., Zeng, T., Shen, Y., Liu, S., et al. (2017). Genome-wide mining of perfect microsatellites and tetranucleotide orthologous microsatellites estimates in six primate species. *Gene* 643, 124–132. doi: 10.1016/j.gene.2017.12.008

Yan, X., Lin, Y., Huang, Z., and Han, R. (2018). Characterisation of biological and biocontrol traits of entomopathogenic nematodes promising for control of striped flea beetle (*Phyllotreta striolata*). *Nematology* 20, 503–518.

Yang, N., Dong, Z., Chen, A. D., Yin, Y. Q., Li, X. Y., and Chu, D. (2020). Migration of *Sogatella furcifera* between the Greater Mekong Subregion and northern China revealed by mtDNA and SNP. *BMC Ecol. Evol.* 20:154. doi: 10.1186/s12862-020-01722-4

Zhan, S., Merlin, C., Boore, J. L., and Reppert, S. M. (2011). The monarch butterfly genome yields insights into long-distance migration. *Cell* 147, 1171–1185. doi: 10.1016/j.cell.2011.09.052

Zhang, M. X., Ling, B., and Liang, G. W. (2000). Investigation and analysis of the population dynamics of striped flea beetle on crucifer vegetables. *Plant Prot.* 26, 1–3.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 14 March 2022
doi: 10.3389/fevo.2022.846499

Differentiation of the Chestnut Tiger Butterfly *Parantica sita* (Lepidoptera: Nymphalidae: Danainae) in China

Ping Hu[1,2†], Liangzhi Lu[1†], Shaoji Hu[3,4], Wa Da[5], Chia-Lung Huang[6], Huihong Zhang[7], Di Wang[1], Yifan Zhang[1], Yongqiang Xu[5] and Rongjiang Wang[1,8]*

[1] School of Life Sciences, Peking University, Beijing, China, [2] Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China, [3] Yunnan Key Laboratory of International Rivers and Transboundary Eco-Security, Yunnan University, Kunming, China, [4] Institute of International Rivers and Eco-Security, Yunnan University, Kunming, China, [5] Tibet Plateau Institute of Biology, Lhasa, China, [6] Institute of Oceanography, Minjiang University, Fuzhou, China, [7] School of Agriculture, Yunnan University, Kunming, China, [8] Institute of Ecology, Peking University, Beijing, China

OPEN ACCESS

Edited by:
Xin Zhou,
China Agricultural University, China

Reviewed by:
Zhenyong Du,
China Agricultural University, China
Zhaofu Yang,
Northwest A&F University, China

***Correspondence:**
Rongjiang Wang
rjwang@pku.edu.cn

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *31 December 2021*
Accepted: *08 February 2022*
Published: *14 March 2022*

Citation:
Hu P, Lu L, Hu S, Da W, Huang C-L, Zhang H, Wang D, Zhang Y, Xu Y and Wang R (2022) Differentiation of the Chestnut Tiger Butterfly Parantica sita (Lepidoptera: Nymphalidae: Danainae) in China. Front. Ecol. Evol. 10:846499. doi: 10.3389/fevo.2022.846499

The chestnut tiger butterfly, *Parantica sita* (Kollar) (Lepidoptera: Nymphalidae: Danainae), occurs in Asia, along the Himalayas, and into the Malayan region. Previous studies found three types of mitogenomes with substantial genetic divergence in samples from China. To clarify the level of differentiation within *P. sita*, we investigated both molecular data and morphological features in 429 individuals from China. Upon examination, mitochondrial cytochrome oxidase subunit I (COI) sequences showed three substantially diverged haplotype groups. Based on microsatellite genotypes, the samples divided into three clusters that were consistent with the COI haplotype groups. With that genetic data, we named three distinguishable *P. sita* lineages: PS-A, PS-B, and PS-C. We also found obvious morphological differences in wing color, male sex brand, and genitalia structures among the three lineages. According to the published structure of male genitalia, that of PS-A is identical to that of *P. s. sita*, and that of PS-B is identical to that of *P. pedonga*. Based on all the results, we tentatively propose dividing *P. sita* into three species: PS-A (the former *P. s. sita*) is the typical *Parantica sita* [Kollar, (1844)], mainly distributed in southwestern China; PS-C (the former *P. s. niphonica*) is elevated to full species as *Parantica niphonica* (Moore, 1883), distributed in Taiwan Island and Japan; and PS-B will be *Parantica pedonga* Fujioka, 1970, mainly distributed in Tibet and western Sichuan. Divergence time estimates showed that PS-A separated from the PS-B + PS-C clade about 8.79 million years ago (Ma), when the Hengduan Mountains underwent an appreciable elevation increase, isolating the Tibet population from the others. PS-B and PS-C diverged about 4.87 Ma, in accord with the formation of Taiwan Island mountains. The founder effect may explain why PS-C's genetic diversity is lower than that of the other clades.

Keywords: male genitalia, speciation, biogeography, COI sequence, microsatellite, genetic differentiation

INTRODUCTION

Fundamental to understanding the evolutionary sciences is the understanding of the mechanisms underlying biodiversity (Purvis and Hector, 2000), and many studies have attempted to infer how the genetic differentiation of species has been influenced through time and space (e.g., Howes et al., 2006; Cheng et al., 2016; Liu et al., 2019, 2021). Genetic differentiation is influenced by both abiotic and biotic factors, possibly working jointly, so that various evolutionary histories are found among different taxa (e.g., Nielson et al., 2001; Kozak et al., 2006; Cheng et al., 2019).

Geological events that change topography and climate may critically affect the mechanisms influencing biodiversity (Hewitt, 2000). For instance, the uplift of the Qinghai-Tibet Plateau (QTP), including the Himalaya and Hengduan Mountains, in southwestern China created a complex, heterogeneous topography [e.g., large altitudinal differences that often exceed 2,000 m between valleys and mountain ridges in a series of parallel, north-south–oriented mountains (Yao et al., 2010)]. High mountains create biogeographic barriers that can impede gene flow between populations and can drive genetic differentiation and even speciation (Long et al., 2006; Lei et al., 2014; Niu et al., 2018; Cheng et al., 2019). Additionally, the QTP uplift greatly influenced climate, such as the development of the Asian monsoon system (Tang et al., 2013) and the aridification of Central Asia (Miao et al., 2012). Subsequent habitat diversification led to increased biodiversity through ecological adaptations (Ledevin et al., 2018; Liu et al., 2021), leaving this area a major global biodiversity hotspot with high levels of endemism and species richness (Myers et al., 2000).

Taiwan Island, another biodiversity hotspot (Myers et al., 2000), was also affected by the Cenozoic Himalayan orogeny (Zhang et al., 2015), and populations distributed on both sides of the Taiwan Strait may have experienced genetic differentiation. During the interglacial-glacial cycles in the Pleistocene, a land bridge would periodically form in the Strait, leading to repeated connections and separations between Mainland China and Taiwan Island (Teng, 1990; Huang et al., 1997). Those occurrences greatly influenced Taiwan's contemporary fauna and flora. As global biodiversity hotspots, the QTP and Taiwan Island are popular areas for the exploration of biodiversity mechanisms (e.g., Lei et al., 2014; Niu et al., 2018; Cheng et al., 2019; Liu et al., 2021), and even more investigations are needed to clarify those regions' species' complex evolutionary histories.

In this study, we focus on the chestnut tiger butterfly, *Parantica sita* (Kollar) (Lepidoptera: Nymphalidae: Danainae). Found across Asia, *P. sita* has six subspecies: *P. s. sita*, *P. s. niphonica* (Moore), *P. s. tytia* (Gray), *P. s. ethologa* (Swinhoe), *P. s. oblit*a (Tsukada and Nishiyama), and *P. s. melanosticta* (Morishita)[1]. Among them, *P. s. sita* is distributed mainly in southwestern China, and *P. s. niphonica* occurs in Taiwan Island, South Korea, and Japan (Chou, 1994). In a previous study, we found substantial genetic divergence between the mitogenomes of those two subspecies—a 4.1% divergence in the complete mitogenome and a fragment insertion/deletion in the A + T rich region (Hu and Wang, 2019). Additionally, the phylogenetic analysis showed *P. s. niphonica* more closely related to *P. luzonensis* than to *P. s. sita*, so we proposed that *P. s. sita* and *P. s. niphonica* are independent species (Hu and Wang, 2019). In a subsequent study, we found a novel mitogenome in *P. s. sita* samples from Leshan, Sichuan, in southwestern China. Its sequence diverged significantly from the known mitogenomes of those two subspecies, and there was a fragment insertion/deletion in the A + T rich region (Zhang et al., 2022). Then, phylogenetic analysis showed that those three mitogenome types were not monophyletic because *P. luzonensis* was mixed in (Zhang et al., 2022). This implied that the Leshan individuals were possibly from an independent species and, consequently, that *P. sita* in China could be a species complex. To clarify *P. sita*'s genetic differentiation, we collected a large number of individuals from throughout the species' entire distribution area in China for a genotyping investigation, as well as for morphological comparisons. Meanwhile, as the distribution area covers both QTP and Taiwan Island, an understanding of *P. sita*'s genetic divergence would be valuable for explaining how geological events contribute to species evolution.

[1]https://www.nic.funet.fi/pub/sci/bio/life/

MATERIALS AND METHODS

Sampling of Butterflies and DNA Extraction

We collected 403 *P. s. sita* individuals from its distribution range (i.e., Sichuan, Chongqing, Yunnan, Guizhou, and Tibet) (**Figure 1**). Because samples collected in different months were considered different populations in some sites, 18 populations were sampled from 13 sampling sites (**Supplementary Table 1**). Besides, 26 individuals of *P. s. niphonica* were collected from Japan and Taiwan Island. Because *P. s. niphonica* maintains an annual migration route between Taiwan Island and Japan (Kanazawa et al., 2012; Cheng et al., 2015), those 26 individuals were considered to be from one population (**Supplementary Table 1**). For the outgroup, we collected 10 *P. melaneu* and *P. swinhoei* individuals. All butterflies were taken to the laboratory as soon as possible after capture, and their thoraxes were then preserved in ethanol at −20°C.

Before DNA extraction, the thoraxes were soaked in TE buffer (10 mmol/L Tris–HCl, 1 mmol/L EDTA, pH 8.0) to remove ethanol. Total genomic DNA was extracted with TIANamp Genomic DNA Kit (Tiangen, China) following the manufacturer's instructions.

Cytochrome Oxidase Subunit I Sequencing and Analysis

Cytochrome oxidase subunit I (COI) fragments were amplified using universal primers LCO1490 (5′-GGTCAACAAATCATAAAGATATTGG-3′) and HCO2198 (5′-TAAACTTCAGGGTGACCAAAAAATCA-3′) (Folmer et al., 1994; Hebert et al., 2003). We performed the PCR in a 25 μL reaction volume using the EASYTaq PCR SuperMix Kit

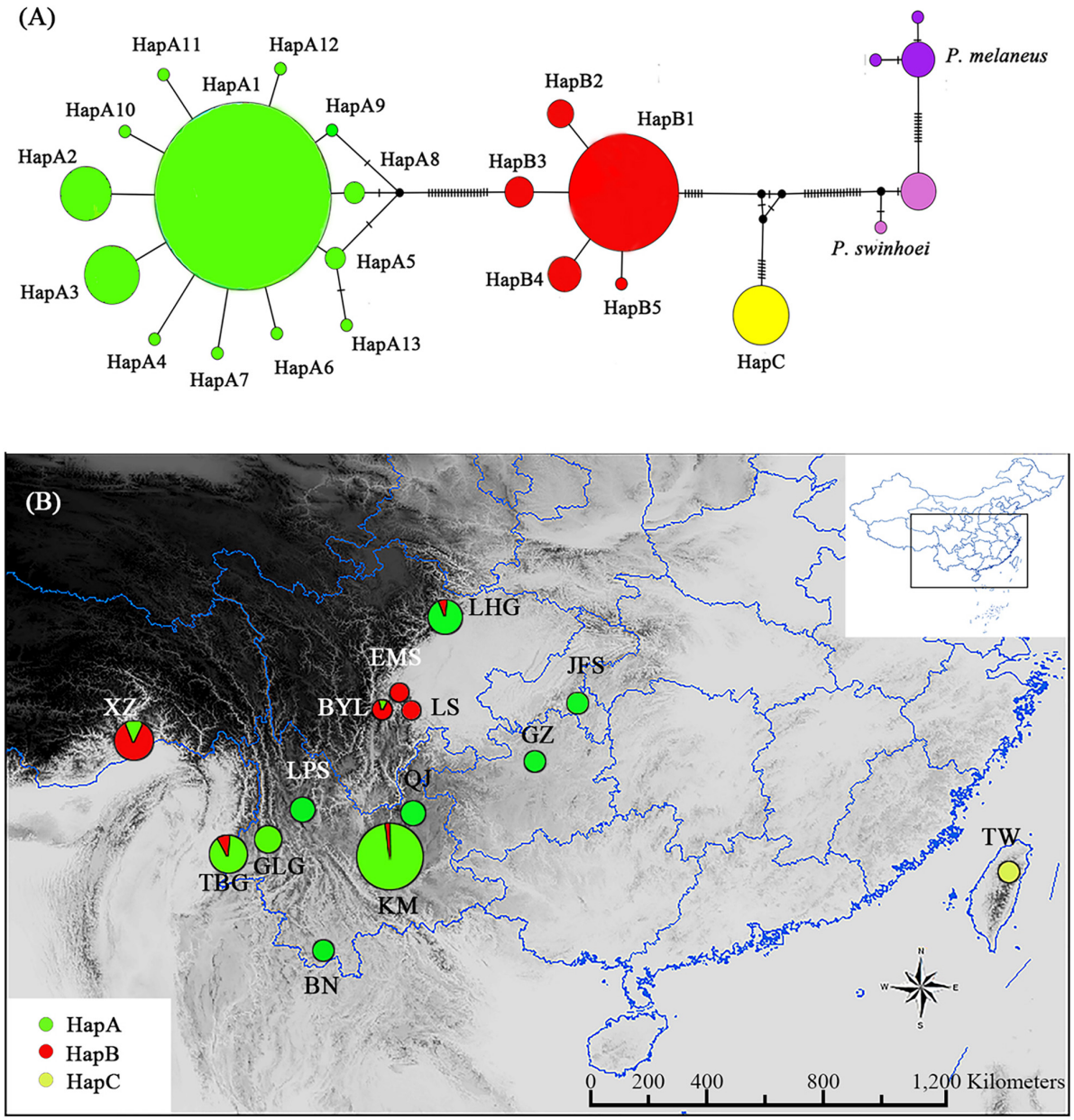


FIGURE 1 | *Parantica sita* haplotypes based on cytochrome oxidase subunit I (COI) sequencing. **(A)** Haplotype network diagram of the 19 sampled populations. Circle size is proportional to the numbers of samples and each vertical line between haplotypes indicates a base difference. **(B)** The geographical distributions of the three haplotype groups. Circle size is proportional to the number of samples. XZ: Motuo, Tibet; TBG: Tongbiguan, Yunnan; GLG: Mt. Gaoligong, Yunnan; LPS: Mt. Luoping, Yunnan; BN: Xishuangbanna, Yunnan; KM: Kunming, Yunnan; QJ: Qujing, Yunnan; BYL: Bayuelin Nature Reserve, Sichuan; EMS: Mt. E'mei, Sichuan; LS: Leshan, Sichuan; LHG: Laohegou Nature Reserve, Sichuan; JFS: Mt. Jinfo, Chongqing; GZ: Guizhou; TW: Mt. Yangming, Taiwan. The inset shows the distribution area within China.

(Transgen, China) under the following conditions: denaturing at 94°C for 4 min, then 35 cycles of denaturing at 94°C for 30 s, annealing at 54°C for 30 s, elongation at 72°C for 50 s, and a final elongation at 72°C for 10 min. The products were sequenced on both forward and reverse strands by RuiBiotech Ltd. (Beijing, China).

Sequences were edited and aligned using the MUSCLE algorithm in MEGA v6.0 (Tamura et al., 2013). Genealogical relationships among haplotypes were further assessed using a TCS network algorithm constructed in PopArt v1.7 (Leigh and Bryant, 2015) with *P. melaneu* and *P. swinhoei* as the outgroups.

Microsatellite Genotyping and Analysis

The microsatellite fragments of 14 polymorphic loci were amplified (**Supplementary Table 2**) using the protocol described in our previous study (Hu et al., 2020). For genetic diversity

TABLE 1 | Results of AMOVA of mtDNA cytochrome oxidase subunit I (COI) and microsatellite data from 19 *Parantica sita* populations.

Source of variation	Degrees of freedom	Variance components	Percentage variation (%)	*F*-statistic	*p*
COI data					
Among groups	2	10.72 Va	90.4	F_{CT} = 0.90	<0.001
Among populations within groups	16	0.06 Vb	0.5	F_{SC} = 0.06	0.002
Within populations	407	1.07 Vc	9.1	F_{ST} = 0.91	<0.001
Total	425	11.85			
Microsatellite data					
Among clusters		0.72 Va	17.3	F_{CT} = 0.173	<0.001
Among populations within clusters		0.03 Vb	0.7	F_{SC} = 0.008	0.002
Within populations		3.39 Vc	82.0	F_{ST} = 0.180	<0.001
Total		4.14			

Vc, Variance within populations; Vb, Variance among populations within groups; Va, Variance among groups/clusters.

analysis, we calculated polymorphism information content, expected heterozygosity, and observed heterozygosity with CERVUS v3.0.7 (Kalinowski et al., 2007); the average numbers of alleles and numbers of effective alleles with PopGen v1.32 (Yeh and Boyle, 1997); and allelic richness and private allelic richness of each population with HP-RARE v1.0 (Kalinowski, 2005). Population structure was evaluated using three methods: (i) Bayesian approximation conducted in STRUCTURE v2.3.4 (Pritchard et al., 2000), (ii) a spatial Bayesian model-based approach in TESS v2.3 (Chen et al., 2007), and (iii) discriminant analysis of principal components using the R package adegenet v2.0 (Jombart et al., 2010).

Genetic Variation and Spatial Analysis

To evaluate the genetic variation both among and within populations based on COI and microsatellite data, we calculated hierarchical analysis of molecular variance (AMOVA) by computing conventional *F*-statistics from haplotypes with 10,000 permutations in Arlequin v3.5 (Excoffier and Lischer, 2010). We also used Arlequin to calculate the pairwise F_{ST} for each pair of the 19 populations. Then, to assess whether there was isolation by distance, we performed Mantel correlation tests between geographical distances and the pairwise genetic distances, $F_{ST}/(1\text{-}F_{ST})$. We also investigated the relationships between genetic and geographical distances by using spatial autocorrelation analysis. Both those analyses and the Mantel tests were performed in GenALEx v6.5 (Peakall and Smouse, 2012). We used MEGA v6.0 (Tamura et al., 2013) to calculate the p-distances for each pair of the 19 populations.

Gene Flow Analysis and Divergence Time Estimation

Assuming that *P. sita* in China is actually three species (see Results), we estimated the gene flow among those species by using MIGRATE-N (Beerli and Felsenstein, 2001) to calculate Bayesian inference from the microsatellite data. The estimated divergence times of the three lineages, based on COI sequences, were performed in BEAST v1.8.4 (Drummond and Rambaut, 2007). Based on whole mitochondrial genes, Chazot et al. (2019) estimated that the divergence time between *P. s. niphonica* and *Ideopsis similis* was about 20 million years ago (Ma) with a 95% highest posterior density (HPD) of 25–16 Ma. We used this divergence time as the secondary calibration in our analysis. The results of three runs were combined using LogCombiner in BEAST (Drummond and Rambaut, 2007) and the initial 10% were discarded as burn-in. The tree with divergence times was visualized using FIGTREE v1.3.1[2].

Morphological Observations

Specimens of both sexes were spread for examination and digitally photographed against a medium-gray background. Photos were adjusted using Adobe Photoshop CS (Adobe, United States), and the ground color near the hindwing anal angle was extracted for comparison across taxa. To compare the male sex brands, specimens were observed under an SMZ1500 stereoscope (Nikon, Japan), and magnified photographs of the textured patches on the upper side of the hindwings and the modified scales on the hindwing undersides were taken with a DMX1200 digital camera (Nikon, Japan) mounted on the stereoscope.

Mainly followed Hu et al. (2018), we observed male and female genitalia as follows. For each specimen, the abdomen was removed and placed into a 1.5 mL microcentrifuge tube. Then 1 mL water was added to the abdomen to allow the tissue to rehydrate at 50°C for 30 min, after which 1 mL 10% NaOH was added to digest the soft tissue at 70°C for 1 h. The treated abdomen was neutralized with 2% acetic acid and then dissected in a water-filled Petri dish under a stereoscope, first removing residual tissues, scales, and hair. The genitalia were then soaked in 80% glycerol for 12 h to render them transparent, and then photographed with a DMX1200 digital camera (Nikon, Japan) mounted on an SMZ1500 stereoscope (Nikon, Japan). The photographs were automatically stacked using Helicon Focus v7.5.8 (Helicon Soft, United States). After observation and photography, all parts of the genitalia were fixed on a glue card using water soluble polyvinyl acetate and pinned with the specimen.

[2]http://tree.bio.ed.ac.uk/software/figtree/

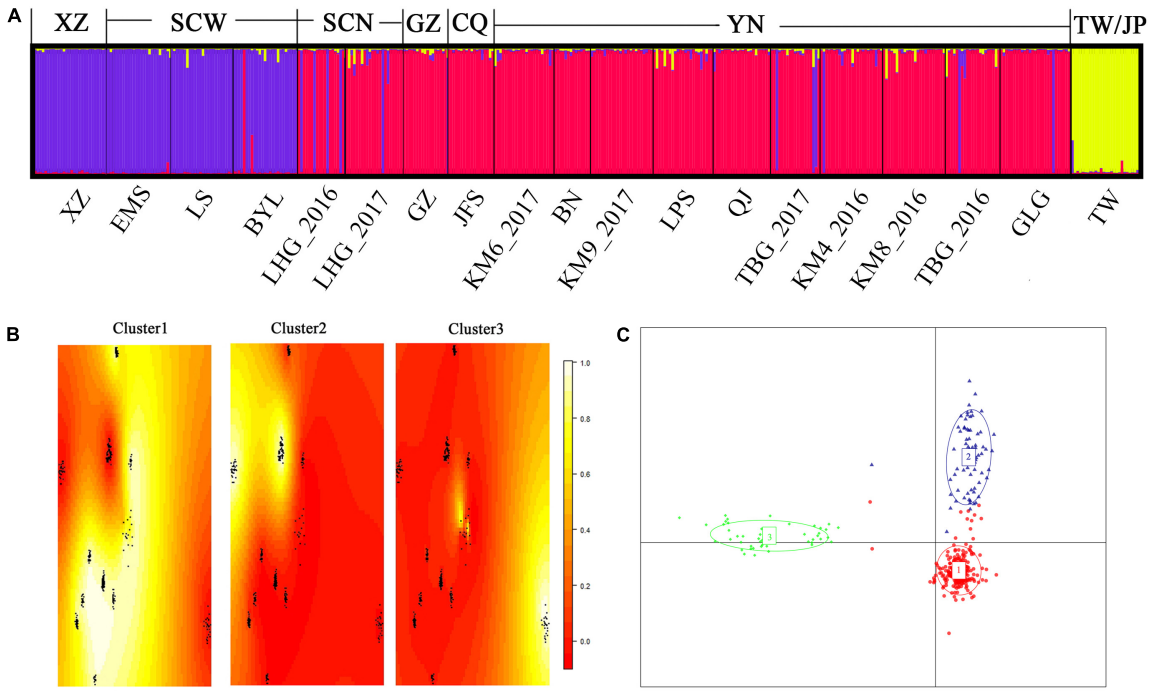


FIGURE 2 | Results of using microsatellite markers with three clustering methods to analyze the 19 *Parantica sita* populations. **(A)** A Bayesian approximation conducted in STRUCTURE. The letters below the bar plot are the geographic populations (see **Supplementary Table 1**) and the letters above are the main geographical distribution areas corresponding to the cluster. XZ: Tibet; SCW: western Sichuan; SCN: northern Sichuan; GZ: Guizhou; CQ: Chongqing; YN: Yunnan; TW/JP: Taiwan Island/Japanese archipelago. Each vertical bar represents an individual and different colors represent speculative clusters. **(B)** A spatial Bayesian model-based approach in TESS. In each heat map, yellow represents individual assignment probability close to 1 while red indicates individual assignment probability close to 0. The 19 *P. sita* populations gathered into three clusters: Cluster1 was distributed mainly in Yunnan, Guizhou, Chongqing, and northern Sichuan; Cluster2 mainly in Tibet and western Sichuan; and Cluster3 exclusively on Taiwan Island and the Japanese archipelago. **(C)** Discriminant analysis of principal components. Each dot represents an individual and each color represents a cluster. All populations gathered into three groups: Cluster1, mainly distributed in Yunnan, Guizhou, Chongqing, and northern Sichuan; Cluster2, mainly distributed in Tibet and western Sichuan; and Cluster3, distributed on Taiwan Island and the Japanese archipelago.

RESULTS

Cytochrome Oxidase Subunit I Sequences

We obtained 658-bp COI fragments from 426 *P. sita* individuals. After alignment and primer sequence removal, the sequences, now 591 bp in length, encoded 197 amino acids. There were 44 polymorphic sites, among which 36 occurred in the third codon site, two in the second codon site, and six in the first site (**Supplementary Table 3**).

We found 19 COI haplotypes in 426 individuals (**Supplementary Table 3**). In the haplotype network with *P. melaneus* and *P. swinhoei* as outgroups, the *P. sita* individuals clustered into three significantly diverged groups (**Figure 1A**). The HapA group consisted of 13 haplotypes with 1–6-bp differences, the HapB group consisted of five haplotypes with 1–2-bp differences, and the HapC group had only one haplotype (**Figure 1A**). There were also great differences among the groups: 25 bp between haplotype A1 and B1, 31 bp between A1 and C, and 15 bp between B1 and C.

The HapA group was distributed mainly in southwestern China, including Yunnan, Guizhou, Chongqing, and northern Sichuan, and the HapB group was distributed mainly in western Sichuan and Tibet (**Figure 1B**). Some individuals in the HapA group were distributed in the HapB group's range and vice versa (**Figure 1B**). The HapC group was distributed exclusively in Japan and Taiwan Island (**Figure 1B**). So, with the 19 *P. sita* populations divided into three haplotype groups, AMOVA indicated that the genetic variation came mainly from among groups (90.4%; **Table 1**). The pairwise F_{ST} values among those three groups were distinctly higher than those among populations within each group (**Supplementary Table 4**). The p-distances among populations within each group ranged from 0.000 to 0.014, and those among three groups ranged from 0.025 to 0.052 (**Supplementary Table 4**).

Microsatellite Genotypes

We analyzed the microsatellite markers at 14 polymorphic loci in 421 of the 429 individuals from 19 populations of *P. sita* (eight individuals failed amplification). The results showed high levels of polymorphism with polymorphism information content ranging from 0.376 to 0.480 and expected heterozygosity from 0.432 to 0.534 (**Supplementary Table 5**).

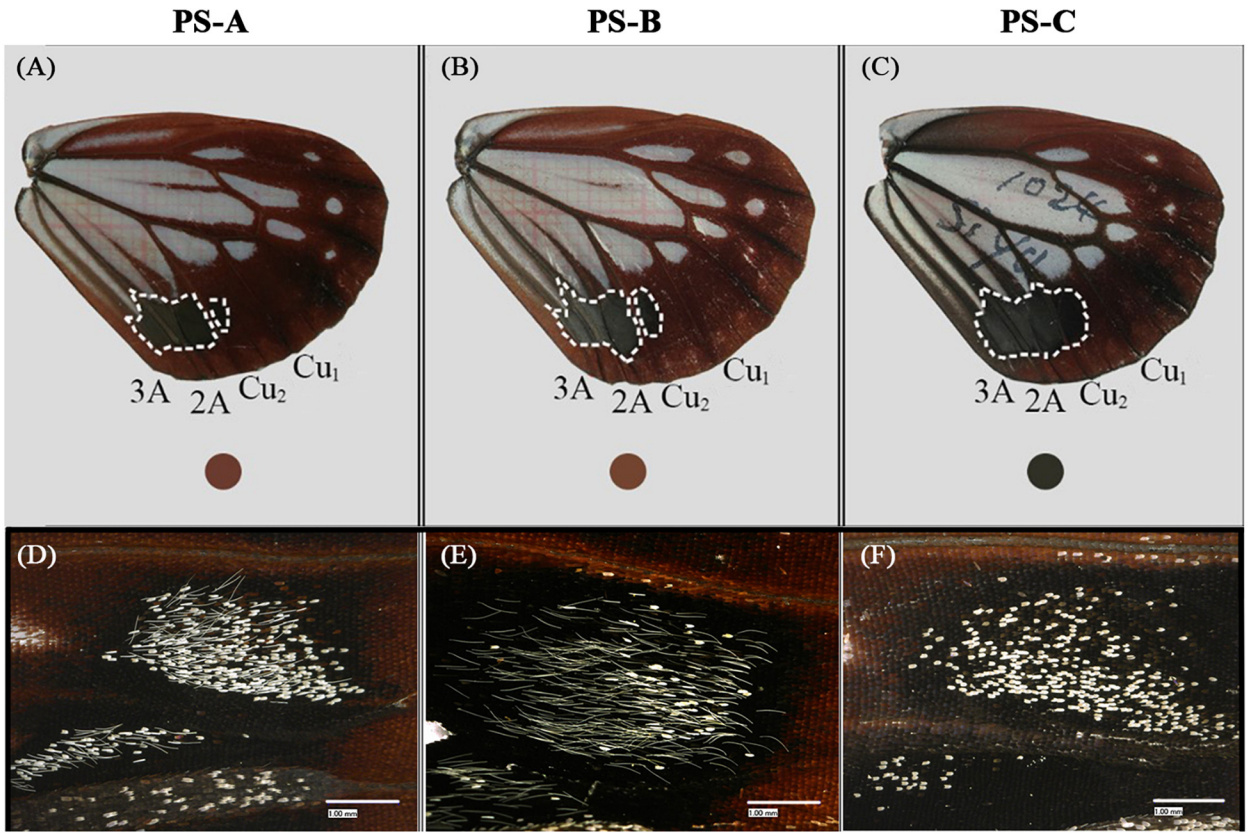


FIGURE 3 | The uppersides **(A–C)** and undersides **(D–F)** of representative hindwings of the three *Parantica sita* lineages: PS-A, PS-B, and PS-C. **(A–C)** The dotted line surrounds the range of the black patch of the male sex brand, and the color within the colored circle indicates the color near the anal angle of the hindwing. **(D–F)** Male sex brand variability as shown by modified white scales on the undersides of the sex brands. Scale bar is 1 mm.

Clustering analysis in STRUCTURE revealed the 19 populations clearly divided into three clusters (**Figure 2A**). The populations in northern Sichuan, Guizhou, Chongqing, and Yunnan gathered into Cluster1 and those in Tibet and western Sichuan into Cluster2, while the individuals from Taiwan Island and Japan were in Cluster3 (**Figure 2A**). Based on their distribution ranges, these three clusters coincided with haplotype groups, corresponding to HapA, HapB, and HapC, respectively. A subsequent analysis using TESS obtained same result—the 19 populations were divided into three clusters (**Figure 2B**). In discriminant analysis of principle components, all individuals formed the same three clusters as those found by STRUCTURE and TESS analysis (**Figure 2C**).

Analysis of molecular variance of the microsatellite data showed that genetic variation came mainly from within populations (82.0%) and among clusters (17.3%, **Table 1**).

Integration of Cytochrome Oxidase Subunit I and Microsatellite Data

After integrating the COI haplotypes and microsatellite genotypes, all individuals matched perfectly. Those belonging to HapA were in Cluster1, those belonging to HapB were in Cluster2, and all of HapC were in Cluster3. Notably, the genetic properties of 12 individuals were "mismatched" to their corresponding geographic ranges: Eleven individuals from the Cluster1/HapA distribution range were identified as Cluster2/HapB and one individual from western Sichuan was identified as Cluster1/HapA (**Figures 1A**, **2**). That implies that Cluster1/HapA and Cluster2/HapB are distinguishable lineages with substantial genetic divergence and, although they are sympatric, some factors may hamper hybridization. Although there were no "mismatched" individuals between Cluster3/HapC and the other two clusters, it is safe to propose that, based on genetic divergence, Cluster3/HapC is an independent lineage. For simplification, we designated the three lineages PS-A, PS-B, and PS-C, respectively, in the following analyses.

Morphological Differences

Overall, the wings of these three lineages appeared similar, but upon closer examination we found differences. First, the sub-hyaline patch in the PS-C discal cell was whitish, but it was almost transparent with a grayish blue hue in PS-A and PS-B (**Figures 3A–C**). Second, the hindwing's upperside background color was blackish brown in PS-C, reddish brown in PS-A, and brick red in PS-B (**Figures 3A–C**). Due to either individual differences or worn wings, some individuals

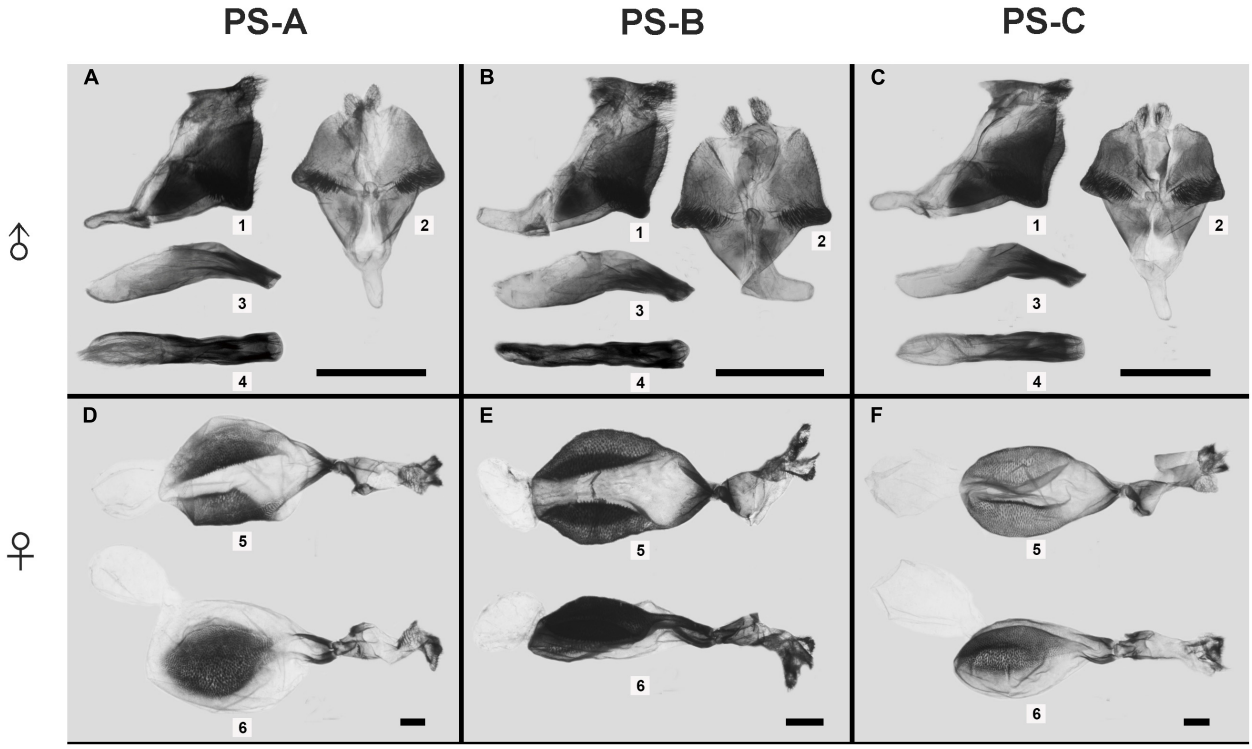


FIGURE 4 | The male and female genitalia of the three *Parantica sita* lineages: PS-A, PS-B, and PS-C. **(A–C)** Male external genitalia (1, genitalia in lateral view; 2, genitalia in dorsal view; 3, aedeagus in lateral view; 4, aedeagus in dorsal view); **(D–F)** Female genitalia (5, dorsal view; 6, lateral view). Scale bar is 1 mm.

could not be accurately distinguished based on wing color. Finally, the male sex brand differed among the three lineages. There are two types of modified white scales, long hairy scales and short round scales, on the undersides of sex brands. PS-B had mainly long hairy scales, PS-C had mainly short round scales, and the ratio of the two scale types in PS-A was approximately 1:1 (**Figures 3D–F**). The proportion of the two types of scales in a couple of individuals were not distinguishable.

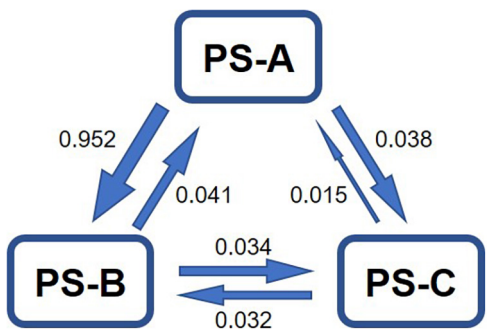


FIGURE 5 | Estimated asymmetric gene flow, based on microsatellite data, among the three lineages of *Parantica sita* (PS-A, PS-B, and PS-C). Numbers indicate the number of migrants per generation.

The male and female genitalia were obviously different among the three lineages. For male genitalia, the distance between the socii of PS-B was obviously broader than those of PS-A and PS-C; the valve of PS-B was narrower than those of PS-A and PS-C; and the saccus of PS-B bent upward (**Figure 4**). The most promising differences between PS-A and PS-C were a blunt projection at the middle of the valve margin and the distance between the socii of PS-C was wider than that of PS-A (**Figure 4**). ANOVA comparing the distances between the socii showed significant differences among the three lineages ($F = 7.694$, $p < 0.01$). For female genitalia, PS-B's papillae anales were longer and the signum granules larger than those of PS-A and PS-C (**Figure 4**). The major difference between PS-A and PS-C was a longer ductus bursae in PS-A (**Figure 4**).

According to the published structure of male genitalia (Lang, 2012), PS-A's overall structure was identical to that of *P. s. sita* and that of PS-B was identical to that of *P. pedonga*.

Gene Flow Among the Three Lineages

The estimated asymmetric gene flow, based on the microsatellite data, among the three lineages showed that the magnitudes of gene flow from PS-A to PS-B and from PS-A to PS-C were much higher than those in the opposite direction (**Figure 5**). The number of migrants per generation in all cases was lower than 1 (**Figure 5**), indicating no obvious gene flow among the three lineages.

TABLE 2 | AMOVA results based on mtDNA cytochrome oxidase subunit I (COI) and microsatellite data from two *Parantica sita* lineages: PS-A and PS-B.

	Source of variation	Degree of freedom	Variance components	Percentage variation (%)	*F*-statistic	*p*
PS-A	**COI data**					
	Among populations	13	0.013 Va	3.6	0.036	0.007
	Within populations	268	0.345 Vb	96.4		
	Total	281	0.358			
	Microsatellite data					
	Among populations		0.010 Va	0.3	0.003	<0.001
	Within populations		3.306 Vb	99.7		
	Total		3.317			
PS-B	**COI data**					
	Among populations	3	0.003 Va	2.1	0.021	0.108
	Within populations	94	0.174 Vb	97.9		
	Total	97	0.177			
	Microsatellite data					
	Among populations		0.064 Va	1.8	0.018	<0.001
	Within populations		3.523 Vb	98.2		
	Total		3.587			

Vb, Variance within populations; Va, Variance among populations.

Estimations of Divergence Times

Divergence time dating results showed that the PS-A clade separated from the PS-B + PS-C clade at about 8.79 Ma (95% HPD: 13.05–4.85 Ma), and the PS-B and PS-C clades separated about 4.87 Ma (95% HPD: 7.95–2.18 Ma; **Figure 6**).

Genetic Differentiations Within Lineages

For PS-A, the AMOVA result based on COI data indicated that the main genetic variation was 96.4% within populations (**Table 2**). A similar result was obtained based with microsatellite data (99.7% within populations, **Table 2**). Mantel test results detected no isolation by distance in the PS-A populations (**Figure 7**). Also, spatial autocorrelation traces were always within the 95% confidence limits, indicating no spatial autocorrelation among all PS-A populations (**Figure 8**).

For PS-B, the AMOVA result based on COI data showed that the genetic variation occurred mainly within populations (97.9%, **Table 2**) and the AMOVA based on microsatellite data showed a similar result (98.2% within populations, **Table 2**). The Mantel test revealed correlated genetic and geographical distances in PS-B populations (**Figure 7**). When the geographical distance between populations was within 0–120 km and greater than 800 km, spatial autocorrelation traces were outside the 95% confidence limits (**Figure 8**), thus indicating that there was spatial autocorrelation between PS-B populations in western Sichuan and Tibet.

DISCUSSION

Differentiation of *P. sita*

In previous studies, we found substantial genetic divergence among three types of mitogenomes in *P. sita* (Wu et al., 2014; Hu and Wang, 2019; Zhang et al., 2022). In this study, the large number of samples we collected across the *P. sita* distribution range in China allowed us to investigate mitochondrial COI haplotypes, and the three resulting haplotype groups we detected in those samples (**Figure 1A**) were consistent with the findings from those previous studies. The genetic divergences among the three groups are even greater than that between two sibling species, *P. melaneus* and *P. swinhoei* (**Figure 1A**). Because great genetic divergences in mitochondrial sequences can be found within insect species (Borchers and Marcus, 2014), we also analyzed the microsatellite genotypes of our samples. Those results showed that the samples once again divided into three clusters (**Figure 2**), each corresponding to mitochondrial haplotype groups. Importantly, the mitochondrial and nuclear data of the 12 genetically/geographically mismatched individuals were consistent, indicating no hybridization between individuals from the different cluster/haplotype groups. Therefore, we conclude that in China there are three *P. sita* lineages (i.e., PS-A, PS-B, and PS-C) that are substantially genetically diverged from one another. There are already two *P. sita* subspecies reported in China, *P. s. sita* in southwestern China and *P. s. niphonica* on Taiwan Island (Chou, 1994). Altogether and based on the sampling regions, we conclude that PS-C is *P. s. niphonica* and there are two lineages in *P. s. sita*.

Combining molecular data and morphological features can provide more powerful taxonomic evidence than either type of data alone (Caesar et al., 2006; Warren et al., 2009; Hu et al., 2018). Genitalia structures are important diagnostic characters in insect taxonomy, and those structures differed greatly among PS-A, PS-B, and PS-C (**Figure 4**). Therefore, based on both molecular divergence and genitalia differences, we can justifiably conclude that PS-A, PS-B, and PS-C lineages are actually three species. According to the published structures of male *P. sita* genitalia (Lang, 2012), PS-A's male genitalia is identical to that of *P. s. sita*. The distribution range of PS-A is almost the same as that of *P. s. sita*, further supporting that PS-A is *P. s. sita*. However, the male genitalia of PS-B is identical to that of *P. pedonga*. *P. pedonga*

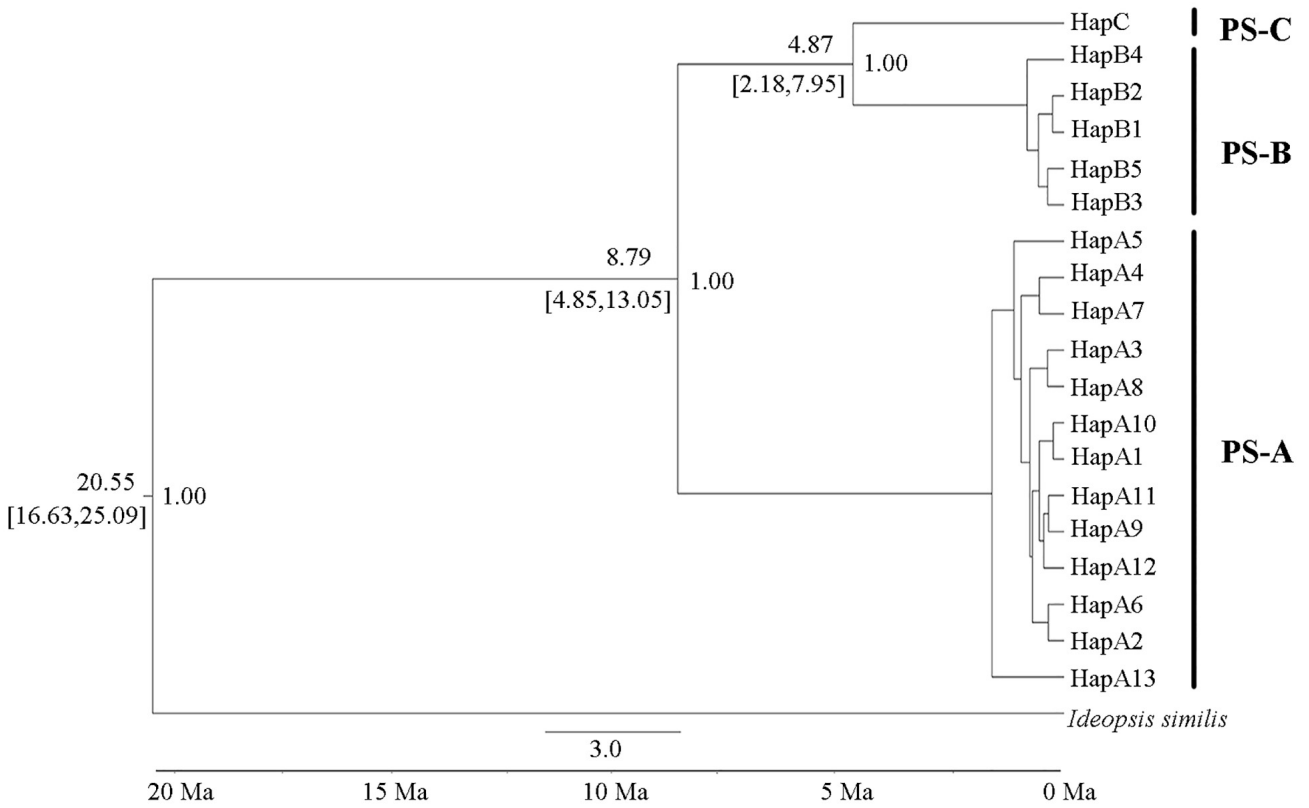


FIGURE 6 | Divergence time estimates of the three *Parantica sita* lineages (PS-A, PS-B, and PS-C). The number above a branch at the left of a node is the divergence time, and that below the branch is its 95% confidence interval. The number to the right of a node is the posterior probability. Ma: million years ago.

had been classified as *P. sita* until Fujioka (1970) recognized it as a new species based on differences in male genitalia. According to Fujioka (1970) description, *P. pedonga* and *P. sita* appear greatly similar, except that the black patch of the male sex brand on *P. pedonga*'s hind wing does not extend beyond the fold in cu_2 space, whereas it goes beyond Cu_2 vein in *P. sita* (Zhang et al., 2008). Here, we found a contradictory result: The black patches of all PS-B males go beyond Cu_2 vein. Fortunately, we collected a sample that had been identified as *P. pedonga* based on the position of its black patch. Through sequencing and microsatellite genotyping, its COI haplotype belonged to HapB, and its microsatellite genotypes were clustered in Cluster2, thus indicating that this *P. pedonga* individual belongs in PS-B. Therefore, we reasonably conclude that PS-B is *P. pedonga*, no matter the extent of the black patch. Possibly, Fujioka (1970) described only those specimens with smaller black patches as *P. pedonga*.

Judging from our molecular and morphological evidence (especially the genitalia characters), we tentatively propose that *P. sita* is actually a species complex of three species: PS-A (the former *P. s. sita*) should be the typical *Parantica sita* [Kollar, (1844)], distributed mainly in southwestern China; PS-C (the former *P. s. niphonica*) should be elevated to full species as *Parantica niphonica* (Moore, 1883), distributed in Taiwan Island and Japan; and PS-B should be recognized as *Parantica pedonga*

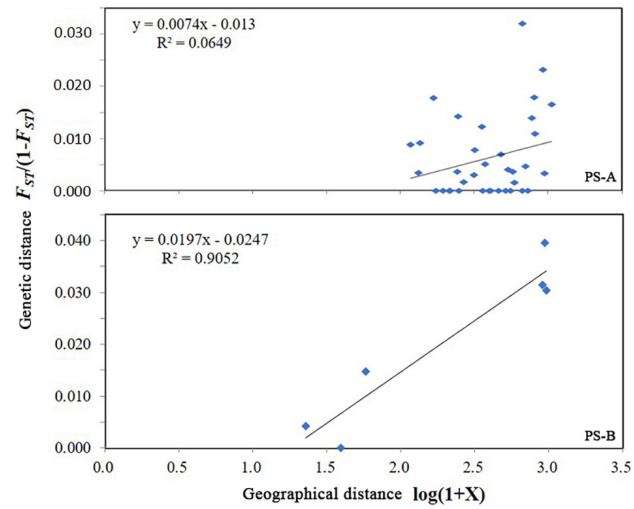


FIGURE 7 | Results of Mantel tests of genetic distance vs. geographical distance of the populations of *Parantica sita* lineages: PS-A and PS-B.

Fujioka, 1970, mainly distributed in Tibet and western Sichuan. **Table 3** summarizes the distinguishable characters for those three species. The wing colors and scale types of the male sex brand can

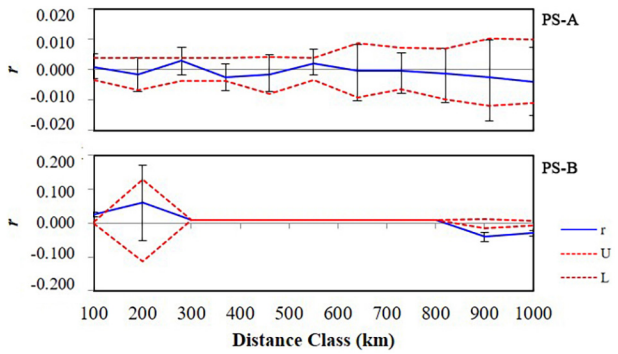


FIGURE 8 | Results of spatial autocorrelation analysis of the PS-A and PS-B *Parantica sita* clades. The blue line indicates the trend of spatial autocorrelation coefficient *r*, the red dotted lines indicate the 95% confidence interval boundaries (U, upper boundary; L, lower boundary).

be used to easily identify most specimens, but not all. However, the significant differences in both genitalia structures and COI sequences among these three species may be used for effective species identification.

Evolutionary History and Related Geological Events

Both PS-A and PS-B are distributed in southwestern China with PS-A mainly in Yunnan, Guizhou, Chongqing, and northern Sichuan, and PS-B in Tibet and western Sichuan (**Figure 1B**). The Hengduan Mountains extend north to south through this region, and the PS-A distribution range covers the eastern side of the Mountains and Tibet to the west. PS-A's separation from PS-B + PS-C at about 8.79 Ma (**Figure 6**) correlates highly with the appreciable increase in the Hengduan Mountains' elevation about 8 (±3) Ma (Molnar et al., 1993; Mulch and Chamberlain, 2006). Based on our gene flow analysis results, PS-A is the ancestor of *P. sita* because asymmetric gene flow always occurs from ancestral to daughter populations/species (Cheng et al., 2019; **Figure 5**). Also, more haplotypes in PS-A than PS-B support PS-A as the ancestor. As the Hengduan Mountains rose, the Tibet population was completely isolated from the others and consequent speciation occurred. The PS-B populations in western Sichuan may have originated from a recent dispersal of the Tibet population across the Hengduan Mountains. We also found some PS-B individuals in the valleys of the Hengduan Mountains. Higher mountains in this region may impede gene flow among geographical populations, so there is isolation by distance in the PS-B populations (**Figure 7**). Similarly, those individuals with "mismatched" distribution ranges likely originated from a recent dispersal.

The PS-B and PS-C lineages separated about 4.87 Ma, soon after the formation of Taiwan Island and concomitant formation of mountains on the Island. The formation resulted from the collision of the Eurasian and Philippine plates about 4–5 Ma (Teng, 1990; Huang et al., 1997). After formation, Taiwan Island underwent repeated connection to and separation from Mainland China during Pleistocene glaciations (Teng, 1990; Huang et al., 1997), so the population/species in Taiwan Island appeared closely related to those in Mainland China, especially those neighboring the Taiwan Strait (Long et al., 2006; Liu et al., 2021). However, PS-C is more closely related to PS-B, the Tibet and western Sichuan populations, rather than PS-A. This implies that the PS-C ancestors did not originate from PS-A and migrate to Taiwan Island through a land bridge. That is likely possible because an extensive plain in eastern China effectively separates the southwestern China populations from Taiwan Island because *P. sita* tends to occur in areas with 1,000–3,000 m elevations (Lu et al., 2020). The potential migration shortcut for the PS-C ancestors may have been southward from Tibet into the Indo-China Peninsula, and then eastward into Taiwan Island. However, since three subspecies (i.e., *P. s. ethologa*, *P. s. oblit*a, and *P. s. melanosticta*) live in the Indo-China Peninsula, Malaysia, and the Philippines, PS-C may have originated from one of those subspecies instead of from PS-B. To clarify PS-C's evolutionary history, those three subspecies must be included in a future investigation; and since *P. luzonensis* was mixed into *P. sita*

TABLE 3 | Summary of the distinguishing characters for the *Parantica sita* lineages: PS-A, PS-B, and PS-C.

	PS-A	PS-B	PS-C
Color of discal cell of wings	transparent	transparent	whitish
Ratio of long hairy scales and short round scales in male sex brand	Approximately 1:1	The majority are long hairy scales	The majority are short round scales
Male genitalia			
Distance of socii	Narrow	Broadest	Broad
Valve	Broad	Narrow	Broad
Saccus	Bent smoothly	Bent upward	Bent smoothly
Blunt projection at valve margin	Absent	Absent	Present
Female genitalia			
Papillae anales	Shorter	Longer	Short
Signum granules	Large	Larger	Small
Ductus bursae	Longer	Shorter	Long
COI haplotype	HapA	HapB	HapC

clades in a phylogenetic analysis (Zhang et al., 2022), it should be included as well.

Whatever the ancestors of PS-C are, it most likely arose from a small number of colonizers, thus leading to a founder effect, which explains the lower genetic diversity of PS-C on Taiwan Island than for the mainland species. On the other hand, PS-C's (*P. s. niphonica*) seasonal, annual migration between Japan and Taiwan Island (Kanazawa et al., 2012; Cheng et al., 2015) likely causes high mortality that could undermine both effective population size and, consequently, high genetic diversity.

Notably, although they diverged about 8.79 Ma, PS-A and PS-B appear morphologically similar, whereas the appearances of PS-C are distinguishable from them. Possibly, because both PS-A and PS-B live in mountainous regions, they may have been under similar selection pressures that resulted in similar appearances as the two lineages adapted to similar ecological conditions. However, PS-C lives in oceanic islands and not mainland mountains, so it had to adapt to different selective pressures that likely resulted in an appearance different from its mainland congeners. To explore these hypotheses, a future study should thoroughly investigate the genes that are under natural selection in the *P. sita* genome. Genomics analysis is also needed to clarify the evolutionary history of the genus *Parantica*.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Microsatellite sequences, https://www.ncbi.nlm.nih.gov/genbank/, MK838447-838459 and MT009235-009237; COI sequences, https://www.ncbi.nlm.nih.gov/genbank/, OM200014-200032.

AUTHOR CONTRIBUTIONS

RW: study designing and manuscript writing. PH and LL: sample collection, molecular analysis, morphology analysis, and writing manuscript. SH and HZ: sample collection and morphology analysis. WD, C-LH, DW, YZ, and YX: sample collection. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the biodiversity conservation program of the Ministry of Ecology and Environment, China (China BON-Butterflies, No. SDZXWJZ012016038) and the Second Tibetan Plateau Scientific Expedition and Research (STEP) program (Grant No. 2019QZKK05010603).

ACKNOWLEDGMENTS

We thank Ran Ma and Yuping Meng from Peking University and Liangyu Yang from Capital Normal University for help with laboratory work. We also thank Wei Zhang from Peking University for suggestions on the whole study, and Xueyan Li from Kunming Institute of Zoology, Chinese Academy of Sciences for manuscript revision. Sample collection was supported by Du Feng from Jilin Agricultural University; Fei Xiong from Leshan, Sichuan Province; Zhiwei Dong and Zhou Chang from Kunming Institute of Zoology, Chinese Academy of Sciences; Pingfan Jia from the Southwest Forestry University; Jichun Xing from Guizhou University; Jiang Hou from Chongqing Museum; and Liang Dou from Sichuan University. We appreciate their help.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2022.846499/full#supplementary-material

REFERENCES

Beerli, P., and Felsenstein, J. (2001). Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *P. Natl. Acad. Sci. U. S. A.* 98, 4563–4568. doi: 10.1073/pnas.081068098

Borchers, T. E., and Marcus, J. M. (2014). Genetic population structure of buckeye butterflies (*Junonia*) from Argentina. *Syst. Entomol.* 39, 242–255. doi: 10.1111/syen.12053

Caesar, R. M., Sörensson, M., and Cognato, A. I. (2006). Integrating DNA data and traditional taxonomy to streamline biodiversity assessment: an example from edaphic beetles in the Klamath ecoregion, California, USA. *Divers. Distrib.* 12, 483–489. doi: 10.1111/j.1366-9516.2006.00237.x

Chazot, N., Wahlberg, N., Lucci Freitas, A. V., Mitter, C., Sohn, J. C., Sahoo, R. K., et al. (2019). Priors and posteriors in Bayesian timing of divergence analyses: the age of butterflies revisited. *Syst. Biol.* 68, 797–813. doi: 10.1093/sysbio/syz002

Chen, C., Durand, E., Forbes, F., and François, O. (2007). Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Mol. Ecol. Notes* 7, 747–756. doi: 10.1111/j.1471-8286.2007.01769.x

Cheng, R., Jiang, N., Xue, D. Y., and Han, H. X. (2019). Species reassessment congruent with the phylogeographical study of the *Biston falcata* species group. *Syst. Entomol.* 44, 886–898. doi: 10.1111/syen.12362

Cheng, R., Jiang, N., Yang, X., Xue, D., Liu, S., and Han, H. (2016). The influence of geological movements on the population differentiation of *Biston panterinaria* (Lepidoptera: Geometridae). *J. Biogeogr.* 43, 691–702. doi: 10.1111/jbi.12676

Cheng, W. W., Pun, H. S., Chung, O., Fukumura, T., and Kanazawa, I. (2015). *Parantica sita niphonica* (Lepidoptera: Nymphalidae) migrated from Japan to Hong Kong, southern China in 2013. *Bull. Osaka Mus. Nat. Hist.* 69, 25–28.

Chou, I. (1994). *Monographia Rhopalocerorum Sinensium (Monograph of Chinese Butterflies)*. Zhengzhou: Henan Scientific and Technological Publishing House, 1.

Drummond, A. J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214. doi: 10.1186/1471-2148-7-214

Excoffier, L., and Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10, 564–567. doi: 10.1111/j.1755-0998.2010.02847.x

Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* 3, 294–299.

Fujioka, T. (1970). Butterflies collecctd by the lepidopterological research expedition to Nepel Himalayan, 1963. Part I Papilionoidea. *Spec. Bull. Lep. Soc. Jap.* 4, 1–125.

Hebert, P. D. N., Cywinska, A., Ball, S. L., and DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proc. R. Soc. B-Biol. Sci.* 270, 313–321. doi: 10.1098/rspb.2002.2218

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature* 405, 907–913. doi: 10.1038/35016000

Howes, B. J., Lindsay, B., and Lougheed, S. C. (2006). Range-wide phylogeography of a temperate lizard, the five-lined skink (*Eumeces fasciatus*). *Mol. Phylogenet. Evol.* 40, 183–194. doi: 10.1016/j.ympev.2006.03.008

Hu, P., Huang, C. L., Luo, M. X., Hsu, Y. F., and Wang, R. J. (2020). Development and characterization of novel microsatellite markers in chestnut tiger butterfly *Parantica sita* (Lepidoptera: Nymphalidae) using next-generation sequencing. *Appl. Entomol. Zool.* 55, 281–286. doi: 10.1007/s13355-020-00675-w

Hu, P., and Wang, R. J. (2019). The complete mitochondrial genome of *Parantica sita sita* (Lepidoptera: Nymphalidae: Danainae) revealing substantial genetic divergence from its sibling subspecies *P. s. niphonica*. *Gene* 686, 76–84. doi: 10.1016/j.gene.2018.10.088

Hu, S. J., Cotton, A. M., Condamine, F. L., Duan, K., Wang, R. J., Hsu, Y. F., et al. (2018). Revision of *Pazala* Moore, 1888: the *Graphium* (*Pazala*) *mandarinus* (Oberthür, 1879) group, with treatments of known taxa and descriptions of new species and new subspecies (Lepidoptera: Papilionidae). *Zootaxa* 4441, 401–446. doi: 10.11646/zootaxa.4441.3.1

Huang, C. Y., Wu, W. Y., Chang, C. P., Tsao, S., Yuan, P. B., Lin, C. W., et al. (1997). Tectonic evolution of accretionary prism in the arc-continent collision terrane of Taiwan. *Tectonophysics* 281, 31–51. doi: 10.1016/S0040-1951(97)00157-1

Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11:94. doi: 10.1186/1471-2156-11-94

Kalinowski, S. T. (2005). HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol. Ecol. Notes* 5, 187–189. doi: 10.1111/j.1471-8286.2004.00845.x

Kalinowski, S. T., Taper, M. L., and Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* 16, 1099–1106. doi: 10.1111/j.1365-294X.2007.03089.x

Kanazawa, I., Chen, C., and Hiyoshi, Y. (2012). A chestnut tiger, *Parantica sita niphonica* (Nymphalidae: Danainae), marked in Japan and recaptured in China in 2006. *News Lepidopterists'. Soc.* 54, 38–39.

Kozak, K. H., Blaine, R. A., and Larson, A. (2006). Gene lineages and eastern North American palaeodrainage basins: phylogeography and speciation in salamanders of the *Eurycea bislineata* species complex. *Mol. Ecol.* 15, 191–207. doi: 10.1111/j.1365-294X.2005.02757.x

Lang, S. Y. (2012). *The Nymphalidae of China (Lepidoptera, Rhopalocear)*, Part I. Pardubice: Tshikolovets Publications.

Ledevin, R., Chevret, P., Helvaci, Z., Michaux, J. R., and Renaud, S. (2018). Bank voles in Southern Eurasia: vicariance and adaptation. *J. Mamm. Evol.* 25, 119–129. doi: 10.1007/s10914-016-9368-3

Lei, F., Qu, Y., and Song, G. (2014). Species diversification and phylogeographical patterns of birds in response to the uplift of the Qinghai-Tibet Plateau and Quaternary glaciations. *Curr. Zool.* 60, 149–161. doi: 10.1093/czoolo/60.2.149

Leigh, J. W., and Bryant, D. (2015). POPART: full-feature software for haplotype network construction. *Methods Ecol. Evol.* 6, 1110–1116. doi: 10.1111/2041-210X.12410

Liu, H. X., Liang, J. Y., Zhou, J. Y., Wang, S. J., and Bu, W. J. (2021). Vicariance and ecological adaptation drive genetic and morphological diversification of a widely distributed bug, *Carbula crassiventris* (Insecta: Hemiptera: Pentatomidae), in South China. *Ecol. Entomol.* 46, 368–382. doi: 10.1111/een.12978

Liu, Y. X., Dietrich, C. H., and Wei, C. (2019). Genetic divergence, population differentiation and phylogeography of the cicada *Subpsaltria yangi* based on molecular and acoustic data: an example of the early stage of speciation? *BMC Evol. Biol.* 19:5. doi: 10.1186/s12862-018-1317-8

Long, Y., Wan, H., Yan, F. M., Xu, C. R., Lei, G. C., Li, S. W., et al. (2006). Glacial effects on sequence divergence of mitochondrial COII of *Polyura eudamippus* (Lepidoptera: Nymphalidae) in China. *Biochem. Genet.* 44, 359–375. doi: 10.1007/s10528-006-9035-0

Lu, L. Z., Hu, P., Zhang, Y. F., Zhang, H. H., Wang, D., Hu, S. J., et al. (2020). Projecting the distribution range of the chestnut tiger butterfly *Parantica sita sita* (Lepidoptera: Nymphalidae: Danainae) in southwestern China. *Appl. Entomol. Zool.* 55, 413–421. doi: 10.1007/s13355-020-00699-2

Miao, Y. F., Herrmann, M., Wu, F. L., Yan, X. L., and Yang, S. L. (2012). What controlled Mid-Late Miocene long-term aridification in Central Asia? – Global cooling or Tibetan Plateau uplift: a review. *Earth-Science Reviews* 112, 155–172.

Molnar, P., England, P., and Martinod, J. (1993). Mantle dynamics, uplift of the tibetan plateau, and the indian monsoon. *Rev. Geophys.* 31, 357–396. doi: 10.1029/93RG02030

Mulch, A., and Chamberlain, C. P. (2006). The rise and growth of Tibet. *Nature* 439, 670–671. doi: 10.1038/439670a

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., and Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature* 403, 853–858. doi: 10.1038/35002501

Nielson, M., Lohman, K., and Sullivan, J. (2001). Phylogeography of the tailed frog (*Ascaphus truei*): Implications for the biogeography of the Pacific Northwest. *Evolution* 55, 147–160. doi: 10.1111/j.0014-3820.2001.tb01280.x

Niu, Y. T., Ye, J. F., Zhang, J. L., Wan, J. Z., Yang, T., Wei, X. X., et al. (2018). Long-distance dispersal or postglacial contraction? Insights into disjunction between Himalaya-Hengduan Mountains and Taiwan in a cold-adapted herbaceous genus, *Triplostegia*. *Ecol. Evol.* 8, 1131–1146. doi: 10.1002/ece3.3719

Peakall, R., and Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research–an update. *Bioinformatics* 28, 2537–2539. doi: 10.1093/bioinformatics/bts460

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1093/genetics/155.2.945

Purvis, A., and Hector, A. (2000). Getting the measure of biodiversity. *Nature* 405, 212–219. doi: 10.1038/35012221

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Tang, H., Micheels, A., Eronen, J. T., Ahrens, B., and Fortelius, M. (2013). Asynchronous responses of East Asian and Indian summer monsoons to mountain uplift shown by regional climate modelling experiments. *Clim. Dynam.* 40, 1531–1549. doi: 10.1007/s00382-012-1603-x

Teng, L. S. (1990). Geotectonic evolution of a late Cenozoic arc-continent collision inTaiwan. *Tectonophysics* 183, 57–76. doi: 10.1016/0040-1951(90)90188-E

Warren, A. D., Ogawa, J. R., and Brower, A. V. Z. (2009). Revised classification of the family Hesperiidae (Lepidoptera: Hesperioidea) based on combined molecular and morphological data. *Syst. Entomol.* 34, 467–523. doi: 10.1111/j.1365-3113.2008.00463.x

Wu, L. W., Lin, L. H., Lees, D. C., and Hsu, Y. F. (2014). Mitogenomic sequences effectively recover relationships within brush-footed butterflies (Lepidoptera: Nymphalidae). *BMC Genomics* 15:468. doi: 10.1186/1471-2164-15-468

Yao, Y. H., Zhang, B. P., Han, F., and Pang, Y. (2010). Spatial pattern and exposure effect of altitudinal belts in the Hengduan Mountains. *J. Mountain Science* 28, 11–20.

Yeh, F. C., and Boyle, T. J. B. (1997). Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belgian J. Bot.* 129, 157.

Zhang, D. Z., Ding, G., Ge, B. M., Zhang, H. B., and Tang, B. P. (2015). Molecular dating of a marine species, *Eriocheir japonica* (Decapoda: Grapsidae),

corroborates cenozoic tectonic events and sea level fluctuation. *Biochem. Syst. Ecol.* 63, 45–50. doi: 10.1016/j.bse.2015.09.023

Zhang, Y. F., Lu, L. Z., Hu, P., Wang, D., and Wang, R. J. (2022). A novel mitochondrial genome haplotype in *Parantica sita sita* (Lepidoptera: Nymphalidae: Danainae) indicates substantial intraspecific genetic divergence. *Appl. Entomol. Zool.* doi: 10.1007/s13355-022-00772-y [Epub ahead of print].

Zhang, Y. L., Fang, L. J., and Chou, I. (2008). Taxonomic study on Chinese species of the genus *Parantica* Moore (Lepidoptera, Nymphalidae, Danainae). *Acta Zootaxon. Sin.* 33, 157–163.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 12 April 2022
doi: 10.3389/fevo.2022.845922

Find My Way to You: A Comparative Study of Antennal Sensilla and Olfactory Genes in Slug Moth With Different Diet Ranges (Lepidoptera: Limacodidae)

Jing Li[1†], Yi-ming Yang[1†], Ying Wang[1], Cai-qing Yang[1,2], Gui-fang Wang[1], Chun-sheng Wu[3] and Ai-bing Zhang[1]*

[1] College of Life Sciences, Capital Normal University, Beijing, China, [2] Key Laboratory of RNA Biology, Center for Big Data Research in Health, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China, [3] National Zoological Museum of China, Institute of Zoology, Chinese Academy of Sciences, Beijing, China

OPEN ACCESS

Edited by:
Hong Lei,
Arizona State University, United States

Reviewed by:
Xin-Cheng Zhao,
Henan Agricultural University, China
Yang Liu,
Institute of Plant Protection (CAAS), China
XiangBo Kong,
Chinese Academy of Forestry, China

***Correspondence:**
Ai-bing Zhang
zhangab2008@mail.cnu.edu.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: 30 December 2021
Accepted: 10 March 2022
Published: 12 April 2022

Citation:
Li J, Yang Y-M, Wang Y, Yang C-Q, Wang G-F, Wu C-S and Zhang A-B (2022) Find My Way to You: A Comparative Study of Antennal Sensilla and Olfactory Genes in Slug Moth With Different Diet Ranges (Lepidoptera: Limacodidae). Front. Ecol. Evol. 10:845922. doi: 10.3389/fevo.2022.845922

Insects and plants that provide them with foods have coexisted for several hundred million years, which leads to various defense approaches and insect-feeding strategies. The host plant provides insects with food sources, shelter materials, and oviposition sites for phytophagous insects. However, they need to find the most suitable host plants in complicated plant communities. The antenna is the main sensory organ of insects, housing different types of sensilla dedicated to detecting chemical cues, motion, humidity, and temperature. Phytophagous insects with different diets may possess various adaptations in their olfactory system. We selected three species of slug moth (*Narosoideus flavidorsalis, Chalcoscelides castaneipars,* and *Setora postornata*) with different diet breadths to detect the structural diversity of antennal sensilla using the scanning electron microscope. A total of nine types of sensilla were identified in these three species, in which two types of sensilla (sensilla uniporous peg and sensilla furcatea) were the first found and reported in Limacodidae. By comparing the number of sensilla types, there was a trend of gradually decreasing the number of sensory types with the gradual expansion of feeding habitats. To better understand the vital roles of olfactory proteins in localizing host plants, we investigated the chemosensory proteins in the antennal transcriptomes of *N. flavidorsalis* and *S. postornata.* However, there was no significant correlation between the number of olfactory genes and the increase of antennal sensilla types. Combining antennal morphology, transcriptome analysis, and the prediction of suitable areas, we better understood the olfactory systems with different feeding preferences, which will provide new prospects for plant–insect interactions and population control methods.

Keywords: Limacodidae, diet range, antennae sensilla, scanning electron microscopy, olfactory proteins, transcriptome, ecological niche modeling

INTRODUCTION

Phytophagous insects rely on the plants as the food sources and shelter materials to support larval performance and survival (Lill et al., 2006). However, the host plant provides caterpillars with nutrients and challenges in defensive chemistry (Mithöfer and Boland, 2012; Checker and Sharma, 2021). Herbivorous insects need to distinguish suitable and unsuitable feeding habitats in complex plant communities (Checker and Sharma, 2021). They have evolved a variety of strategies to cope with sophisticated environments (Stanton, 1983). Diet breadth, ranging from narrow to quite comprehensive, has an important influence on the adaptations to the host plant's defense mechanisms and the exploitation of host recognition (Harris et al., 2003). Consequently, herbivores display various degrees of specificity in their use of plants ranging from strict monophagy to broad polyphagy (Levins and MacArthur, 1969; Thompson, 1998). Mono- and oligophagous insects can adopt only one or a few closely related plant species as their feeding habitats, which allows elaborate adaptations to the plant defense responses (Petschenka and Agrawal, 2015, 2016) and host plant search behaviors (Ahmad, 2012). On the other hand, polyphagous herbivores showed apparent advantages in the face of complex environments where the composition of plant communities had characteristics of temporal and spatial variation in an unpredictable way (Milne and Walter, 2000). However, phytophagous insects have to pay more to adapt to multiple plants, such as withstand various plant secondary metabolites and the cost of mispairing (Cates, 1981; Hunter and McNeil, 1997).

In the evolution of lepidoptera, the specialized feeding behavior evolved into a more general tendency instead of generalization (Bernays, 1997). Rank et al. (1996) further explained this phenomenon from the ecological perspective, which mainly includes three aspects: cost of generalist hypothesis, interspecific competition hypothesis, and predation hypothesis; Nevertheless, a more persuasive explanation is the intense pressure of predators that makes it possible for insects to have more developed nervous systems and sensory functions, thus making it possible to select the specific host plants (Bernays, 1997; Lill et al., 2006). The preference of insects to host plant depends on the acute senses of insects, which include sight, smell, taste, and touch (Salgado and Saastamoinen, 2019). This means that the phytophagous insects can accurately identify the secondary metabolites, which can be detoxified, with their olfactory systems (Visser, 1988; Bruce and Pickett, 2011). Lepidoptera, comprising about 180,000 described species, is the second-largest order of insects; the majority (99%) are phytophagous (Perveen, 2017; Lancaster, 2020). Sensory structure on the antennae of lepidopteran plays a vital role in insect various behaviors, such as orientation, host location, feeding, mating, and identifying oviposition sites (Schneider, 1964; Jaffar-Bandjee et al., 2020). The sensilla are the basic structural and functional unit of insect sensory systems (Schneider, 1964; Chapman, 1998). Given that most of the lepidopterans are herbivores, the studies of the antennal sensilla based on the morphology and molecular levels could provide new prospects for feeding differentiations of lepidopteran insects (Schneider, 1964; Hansson, 1995; Weller et al., 1999).

Within the Lepidoptera, the slug moth family Limacodidae is a part of a monophyletic clade in the superfamily Zygaenoidea (van Nieukerken et al., 2011). Caterpillars in Limacodidae were noted for their colorful and elaborate morphologies, which include aposematic coloration, stinging setae, and spines, which have intrigued researchers for the decades (Lill et al., 2006; Murphy et al., 2010). The larvae of Limacodidae are critical economic pests distributed throughout the world, and they are mainly harmful to fruit and forest trees (Conant et al., 2002). Nonetheless, different species display various degrees of specificity in their use of plants as oviposition substrates and feeding habitats ranging from strict monophagy to broad polyphagy (Duke, 2002). At present, there are scarce studies on the divergences of slug moths (Lin et al., 2019; Bian et al., 2020; Walker et al., 2021). This study selected three species with different diet breadths to conduct the studies on olfactory structures and olfactory-related genes. *Narosoideus flavidorsalis* (Staudinger, 1887) is a typical monophagous species whose host plant is the pear trees, and *Chalcoscelides castaneipars* (Moore, 1866) belong to oligophagous herbivores, whose main host plants are Araliaceae and Lauraceae. The last one was *Setora postornata* (Hampson 1900), a generalist caterpillar feeding on various tree species (e.g., Salicaceae, Rosaceae, Sterculiaceae, Magnoliaceae, Ginkgoaceae, etc.) (Wu, 2010). Using several mitochondrial genes from 35 species of slug moth, we reconstructed the phylogenetic relationships of Limacodidae (unpublished data), and the results showed that the three species had relatively close genetic relationship. Since different types of sensilla recognize different types of information (such as mechanical signals, temperature and humidity changes, and sex pheromones), feeding differentiation may have been associated with the modifications of sensilla morphology or alterations in the relative abundance of sensilla types (Kaupp, 2010). To investigate this, we describe and compare the antennal morphology and sensilla structures of these three moth species that represent different host plant preferences. Nevertheless, the chemical signals entering the sensilla lumen through the sensilla pores are the first step in olfactory recognition processes (Vieira and Rozas, 2011). The sensilla lymph was secreted by non-neuronal support cells. It contained a variety of proteins, which includes the odorant-binding proteins (OBPs) and chemosensory proteins (CSPs), which were highly efficient at recognizing and binding chemical signals (Steinbrecht, 1998). In the Insecta, there are three different types of chemosensory receptors, and the odorant (OR), the gustatory (GR), and the ionotropic (IR) receptors were activated accompanied by the diffusion of odor molecules through the lymph (Sánchez-Gracia et al., 2009; Kaupp, 2010). Although these molecules' full range of functions has not been well established, there is increasing evidence of their importance in chemosensory perception (Tanaka et al., 2009; Bengtsson et al., 2014). Recently, the studies on antennal transcriptomes have led to the identification of olfactory-related genes in several moth species (Yuvaraj et al., 2018; Yang et al., 2020; Jiang et al., 2021), which showed the power of transcriptomic strategies for detecting the high sequence diversity of olfaction-related

genes. However, few comparisons were performed between moths' olfactory genes of different feeding habitats. In this study, we assembled and analyzed the antennal transcriptomes of *N. flavidorsalis* and *S. postornata*, two relatively close relatives, using next-generation sequencing. We report the Gene Ontology (GO) annotation results and sets of putative OBPs, CSPs, sensory neuron membrane proteins (SNMPs), ORs, and IRs in these two species. The studies of the molecular mechanisms of the olfactory system could provide new prospects for host plants' selection of herbivorous insects.

MATERIALS AND METHODS

Insect Rearing and Antenna Collection

A total of three species used in this study, *N. flavidorsalis, C. castaneipars,* and *S. postornata*, were collected from the coniferous and broad-leaved mixed forest by light trapping in Mt Tiantong, Zhejiang Province, China (29°49′N, 121°48′E) (**Supplementary Figure 1**). A number of five male and five female moths were segregated into different cages (40 cm × 40 cm × 60 cm) containing 10% (w/v) sugar solution until the use for microscopy. Antennae from other males and females were excised and stored in RNAlater (Ambion, Austin, TX, United States). Then, all samples were taken back indoors and held at –80°C until RNA isolation. The specimens were identified by Prof. Chun-sheng Wu (Institute of Zoology, Chinese Academy of Sciences, China), and DNA barcoding was also used in species identification. Their voucher specimens are preserved at the Entomological Museum of Capital Normal University, China, and the number was Lep/Lim/191027 to Lep_Lim_191119.

Specimen Preparing for Scanning Electron Microscopy

The adults of the slug moth were first anesthetized by freezing at −20°C for 5 min. Antennae were then dissected and immersed in a freshly prepared fixative solution containing 2% paraformaldehyde and 2.5% glutaraldehyde mixed with a phosphate-buffered solution (PBS) (pH 7.4) for 24 h at 4°C. The antennae were kept at 10% KOH at 80°C for 30 min to remove the scales. Subsequently, the antennae were dehydrated using a graded ethanol series (30, 50, 70, 80, 90, 95, and 100%) followed by the critical-point drying (Leica EM CPD 030, Tokyo, Japan). The dried specimens were carefully glued onto SEM stubs and sputter-coated with gold (Eiko IB-5 Ion Coater, Tokyo, Japan; 45 s, 20 mA). The preparations were examined using a Hitachi SU-8010 cold field emission scanning electron microscope (Japan) under 3 kV voltage.

Statistical Analysis

Given that different terminological terms based on the morphological characters have been used on the antennal sensilla of Lepidoptera, this study primarily followed the study of Schneider (1964) and Jefferson et al. (1970). The antennal sensilla of *N. flavidorsalis, C. castaneipars,* and *S. postornata* were identified, counted, and measured according to the previous measurement method (Onagbola and Fadamiro, 2008; Ivanov and Melnitsky, 2016). Means were based on the measurements (μm) from at least 20 photomicrographs of individual sensilla of the identical type. The types and abundance of sensilla were compared between species and genders and analyzed with the two-way ANOVA with R. The level of significance in all tests was set at 0.05.

cDNA Library Construction and Illumina Sequencing

Narosoideus flavidorsalis and *S. postornata* were selected for antennal transcriptome analysis. Total RNA was extracted from twelve adults' antennae from each species using TRIzol reagent (Invitrogen, Carlsbad, CA, United States) and the RNeasy Plus Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. RNA quantity and purity were checked using the NanoDrop 8000 (Thermo, Waltham, MA, United States). A total amount of 3 μg RNA per sample was used for cDNA library construction. All samples had RIN values above 8. The RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, United States). The cDNA library construction and subsequent Illumina sequencing of samples were performed at Novogene Corporation (Beijing, China). The cDNA libraries were generated using the TruSeq RNA Sample Preparation Kit (Illumina, San Diego, CA, United States). Random hexamer primers were used to synthesize the first-strand cDNA, then synthesizing the second-strand cDNA using buffer, dNTPs, RNase H, and DNA polymerase I at 16°C for 1 h. The remaining overhangs were converted into blunt ends *via* exonuclease or polymerase activities and removed the enzymes. After end repair, A-tailing, and the ligation of adaptors, the products were amplified by PCR and quantified precisely using the Qubit DNA Br Assay Kit (Q10211; Invitrogen, Carlsbad, CA, United States). The library fragments were purified using the MinElute Gel Extraction Kit (Qiagen, Hilden, Germany). The resulting cDNA library preparations were then sequenced on an Illumina HiSeq-2000 platform. The quality of the sequences generated from the PE 200 bp, and all mate-pair libraries were assessed using FastQC (Brown et al., 2017).

De novo Assembly and Functional Annotation

To ensure the accuracy of sequence assembly, the clean data were obtained from raw reads through the following steps: filtered out the reads with adapters, deleted the reads with uncertain bases more than 10%, then removed low-quality (the bases with sequencing error rates greater than 1% are more than 50% in the read) and adaptor sequences by Trimmomatic[1]. At the same time, the Q20, Q30, GC-content, and other related information of the clean data were calculated. All downstream analyses were based on the clean data with high-quality reads. Transcriptome assembly was carried out with program trinity, in which all the

[1]http://www.usadellab.org/cms/index.php?page=Trimmomatic

parameters were set to their defaults (Grabherr et al., 2011). The transcripts longer than 200 bp were performed by NCBI BLASTx searches to databases (Nr, Pfam, Swissprot, KOG, and KEGG); the given transcripts were functionally annotated as the retrieving proteins or nucleic acid with the highest sequence similarity. The blast results were then imported into the Blast2GO pipeline[2] for GO annotation (Conesa et al., 2005; Götz et al., 2008).

Sequences and Phylogenetic Analyses

All candidate chemosensory genes (OBPs, CSPs, SNMPs, ORs, GRs, and IRs) in *N. flavidorsalis* and *S. postornata* and their open reading frames (ORFs) were manually verified by BLASTx searches against custom-made databases and the non-redundant nucleotide collection at NCBI. The ORFs were identified, and the annotation was confirmed by the additional BLAST searches[3]. Transmembrane domains of ORs, IRs, and GRs were predicted using the default parameter of TMHMM2.0 and TMPred, and the N-terminal signal peptide of the candidates, OBPs and CSPs, was predicted by SignalP4.0 (Quevillon et al., 2005).

For further verification of the candidate chemosensory genes and identification of orthologs, phylogenetic analyses were conducted among two slug moths and other related Lepidoptera species, such as *Bombyx mori, Helicoverpa armigera,* and *Holcocerus hippophaecolus.* The available amino acid sequences of chemosensory genes identified in these species were downloaded manually. Maximum likelihood trees were reconstructed using the predicted OR, IR, OBP, CSP, and SNMP protein sequences and orthologs in other species of Lepidoptera and model insects (*Drosophila melanogaster*) to analyze the characteristics of olfactory genes in two species of family Limacodidae.

Amino acid sequences were aligned with MAFFT (version 7.308) (Katoh and Standley, 2013), and the corresponding maximum-likelihood trees were constructed in IQ-TREE (version 2.1.7) using best-fitting substitution model (GTR + I + G) (Trifinopoulos et al., 2016). The tree structure and node support reliability were evaluated by the bootstrap analysis with 1,000 replicates. The phylogenetic trees were colored and arranged in FigTree (version 1.4.2)[4].

Prediction of Suitable Areas With Ecological Niche Modeling

Ecological niche modeling (ENM), a widely accepted method to visualize the distribution patterns (Peterson and Soberón, 2012), was adopted to compare the area of suitable optimal habitats of *N. flavidorsalis* and *S. postornata*. Meanwhile, the habitat suitability of the host plants of these two species was also predicted. *Pyrus sorotina* was selected as the dominant host plant of *N. flavidorsalis*. A number of three plants, *Juglans regia, Populus simonii,* and *Camellia sinensis,* were chosen as representative host plants of euryphagous *S. postornata*. The occurrence records for niche modeling analysis were collected from the online dataset Global Biodiversity Information Facility (GBIF) and published documents. The present bioclimatic variables were downloaded from the WordClim website[5]. The maximum entropy (MaxEnt) approach was employed to predict the species distribution using the ENMTools version 1.0.4 package in the R platform with MaxEnt.jar version 3.4.4 (Warren and Seifert, 2011). A combination of six MaxEnt feature classes [linear (L), linear quadratic (LQ), hinge (H), linear quadratic hinge (LQH), linear quadratic hinge product (LQHP), and linear quadratic hinge product threshold (LQHPT)] was tested to optimize model parameters for the calibration. We trained that the MaxEnt models were trained with the sub-sample method, in which 1/3 of the presence points were set aside for model testing evaluation. The model was trained by replicating each model ten times, and the final output used was the average of all runs. We selected the model with a combination of feature classes and the regularization multiplier that had the lowest akaike information criterion (AICc) value, which could describe the model fit and complexity (Shin et al., 2021).

RESULTS

General Antennae Morphology

The antennae structure of the three species is broadly in line with other lepidopteran insects, which consisted of a scape, a pedicel, and a flagellum. Based on the antennal data from male samples, there was no significant difference in the pedicel and total antennae length among the three species, but only in the side-branches size (**Table 1**). There was no significant difference in antennae length among the three female species, similar to the male data. Still, a significant difference was detected in the length and width of flagellomeres among the three species (**Table 2**). The antennae of all three species bear a pair of sexual dimorphic antennae, the female ones were filiform, and the male ones were bipectinate (**Supplementary Figure 2**). The number of flagellomeres of these three species was different, which was consistent with the difference in antennae length. Meanwhile, the number of flagellomeres of the males was slightly more than that of the females (**Supplementary Table 1**).

Sensilla Types on the Antennae

A total of nine types of sensilla were identified on the antennae of three slug moths, in which five types each having two types of sensilla: sensilla chaetica (SCh I, SCh II), Böhm's bristles (BB I, BB II), sensilla styloconica (SSt I, SSt II), sensilla coeloconica (SCo I, SCo II), and sensilla furcatea (SFu I, SFu II). In these types of sensilla, SB, SSq, and SCo I were detected only in *N. flavidorsalis*, while SCh II and SFu II were detected only in *S. postornata.* In general, *N. flavidorsalis* has more types of sensilla than the other two species (**Table 3**).

Sensilla Trichoderma (STr) was the most widespread sensilla in the three species, densely distributed on the ventral side of the flagellum spindle and lateral branches of both male and female antennae. The surface facial structure of STr among the three species was different: *N. flavidorsalis* was irregular, and the others were arranged in an oblique line (**Supplementary Figure 3**).

[2]http://eggnogdb.embl.de/#/app/home

[3]http://blast.ncbi.nlm.nih.gov/Blast.cgi

[4]http://tree.bio.ed.ac.uk/software/figtree

[5]http://www.worldclim.org/

TABLE 1 | Morphological measurements (mean ± SE, μm) of three parts of antenna in three species of male Limacodidae moths.

Species		Scape	Pedicel	Flagellomeres	Total antenna
N. flavidorsalis	Length	357.4 ± 39.7[a]	162.6 ± 15.4[a]	(110.1 ± 3.4) ~ (210.4 ± 44.5)[a]	6893.8 ± 374.5[a]
	Width	401.1 ± 24.6[a]	283.6 ± 34.1[a]	(51.9 ± 14.8) ~ (182.6 ± 25.3)[a]	6520.2 ± 618.9[a]
	Side-branches Length	—	—	(93.2 ± 23.1) ~ (496.7 ± 58.5)[a]	14687.1 ± 1719.3[a]
C. castaneipars	Length	**465.7 ± 23.7[b]**	156.3 ± 11.9[a]	(76.8 ± 6.1) ~ (205.5 ± 30.6)[a]	7233.7 ± 270.2[a]
	Width	389.2 ± 24.1[a]	274.6 ± 22.9[a]	(59.5 ± 4.4) ~ (202.3 ± 14.2)[b]	**7137.0 ± 108.5[b]**
	Side-branches Length	—	—	(124.8 ± 38.1) ~ (441.4 ± 36.5)[a]	17463.9 ± 1803.0[a]
S. postornata	Length	306.5 ± 24.7[a]	134.7 ± 23.9[a]	(75.6 ± 10.7) ~ (187.3 ± 39.0)[b]	6568.6 ± 324.1[a]
	Width	362.1 ± 5.5[a]	262.1 ± 21.5[a]	(61.9 ± 1.8) ~ (213.6 ± 27.7)[b]	6069.1 ± 694.8[ab]
	Side-branches Length	—	—	**(81.4 ± 9.2) ~ (457.1 ± 25.4)[b]**	14501.1 ± 3802.1[a]

Different letters (a,b) represent significant differences ($p < 0.05$) between species in each column, no data (line). Bold values stand for significant differences.

TABLE 2 | Morphological measurements (mean ± SE, μm) of three parts of antenna in three species of female Limacodidae moths.

Species		Scape	Pedicel	Flagellomeres	Total antenna
N. flavidorsalis	Length	431.3 ± 10.6[a]	167.0 ± 19.9[a]	(112.7 ± 7.3) ~ (216.1 ± 8.6)[a]	7998.1 ± 101.2[a]
	Width	406.5 ± 0.4[a]	286.2 ± 22.6[a]	(81.4 ± 10.5) ~ (254.1 ± 15.1)[a]	6935.7 ± 15.9[a]
C. castaneipars	Length	417.3 ± 19.4[a]	184.6 ± 3.5[a]	(113.8 ± 0.2) ~ (206.6 ± 9.4)[b]	7593.0 ± 267.8[a]
	Width	375.6 ± 19.7[a]	287.9 ± 10.6[a]	(96.8 ± 2.3) ~ (224.1 ± 1.23)[b]	7522.4 ± 475.9[a]
S. postornata	Length	341.5 ± 27.4[a]	219.2 ± 87.9[a]	(72.61 ± 27.3) ~ (184.0 ± 10.7)[c]	6358.9 ± 1334.7[a]
	Width	234.3 ± 47.0[b]	256.6 ± 17.5[a]	(48.5 ± 3.3) ~ (213.3 ± 9.2)[c]	6043.4 ± 1114.3[a]

Different letters (a,b,c) represent significant differences ($p < 0.05$) between species in each row.

Sensilla basiconica (SB) is located vertically in a shallow pit with a wrinkled proximal base (**Supplementary Figure 4**). SCh mainly found on the end of the antennal axis and lateral branches. Two subtypes of SCh (I and II) were identified based on the pattern of these grooves: SCh I were longitudinally striated from base to the end, while in SCh II, the grooves of the stripe gradually changed from longitudinal line to the imbricated texture (**Figure 1**).

TABLE 3 | The distribution of nine types of antennal sensilla among three Limacodidae species.

Antennal sensilla	*N. flavidorsalis*		*C. castaneipars*		*S. postornata*	
	Female	Male	Female	Male	Female	Male
Sensilla trichodea (STr)	+	+	+	+	+	+
Sensilla basiconica (SB)	+	+	–	–	–	–
Sensilla chaetica I (SCh I)	+	+	+	+	–	–
Sensilla chaetica II (SCh II)	–	–	–	–	+	+
Böhm's bristles I (BB I)	+	+	+	+	+	+
Böhm's bristles II (BB II)	–	+	–	+	–	–
Sensilla squamiformia (SSq)	–	+	–	–	–	–
Sensilla styloconica I (SSt I)	+	+	+	+	+	+
Sensilla styloconica II (SSt II)	+	+	+	+	+	+
Sensilla coeloclnica I (SCo I)	+	+	–	–	–	–
Sensilla coeloclnica II (SCo II)	+	+	+	+	+	+
Sensilla uniporous peg (SUp)	+	+	+	–	–	+
Sensilla furcatea I (SFu I)	+	+	+	+	+	+
Sensilla furcatea II (SFu II)	–	–	–	–	+	+
Sum	10	12	9	9	8	9

Present (+), absent (–).

BB sensilla, which were found in clusters at the bases of the scape and pedicel, were quite long with a smooth surface and sharp end. The difference between the BB I and BB II was whether the end of the bristles was bifurcated or not (**Figure 1**). Sensilla squamiformia (SSq) was similar to the antennae scale structure, with no porous structure and conspicuous longitudinal stripes (**Supplementary Figure 5**). SSt with a terminal conical protrusion (without a pore) were on the surface of a decorative pattern of a cylindrical protrusion from the antennal surface. The two subtypes were found based on their structure: SSt I has a long columnar and a small cone, and SSt II has a short columnar with a longer conical body (**Supplementary Figure 6**). SCo were irregularly scattered on the surface of the flagellum, which consists of multiple longitudinal grooves and no stoma distribution. It can be divided into two subtypes according to whether the base is in the circular cavity of the epidermal bulge: SCo I was covered by the socket at the bottom, and SCo II was wholly exposed at the base (**Figure 2**). The sensilla uniporous peg (SUp) base was stuck into the convex epidermal with a smooth surface; the top was blunt and perforated (**Figure 1**). The SFu gradually narrowed from the base to the end and bifurcated at the end. There were two subtypes in *S. postornata*. The SFu II generally had longer poles and shorter bifurcation than SFu I (**Supplementary Figure 7**).

Morphological Measurements of Sensilla

There were significant differences in several sensilla types between female *S. postornata* and female *N. flavidorsalis*, which includes sensilla trichoidea, sensilla coeloclnica (SCo), SSt, BB. The differences between species *C. castaneipars* and the other two species were not noticeable. There was no significant difference of

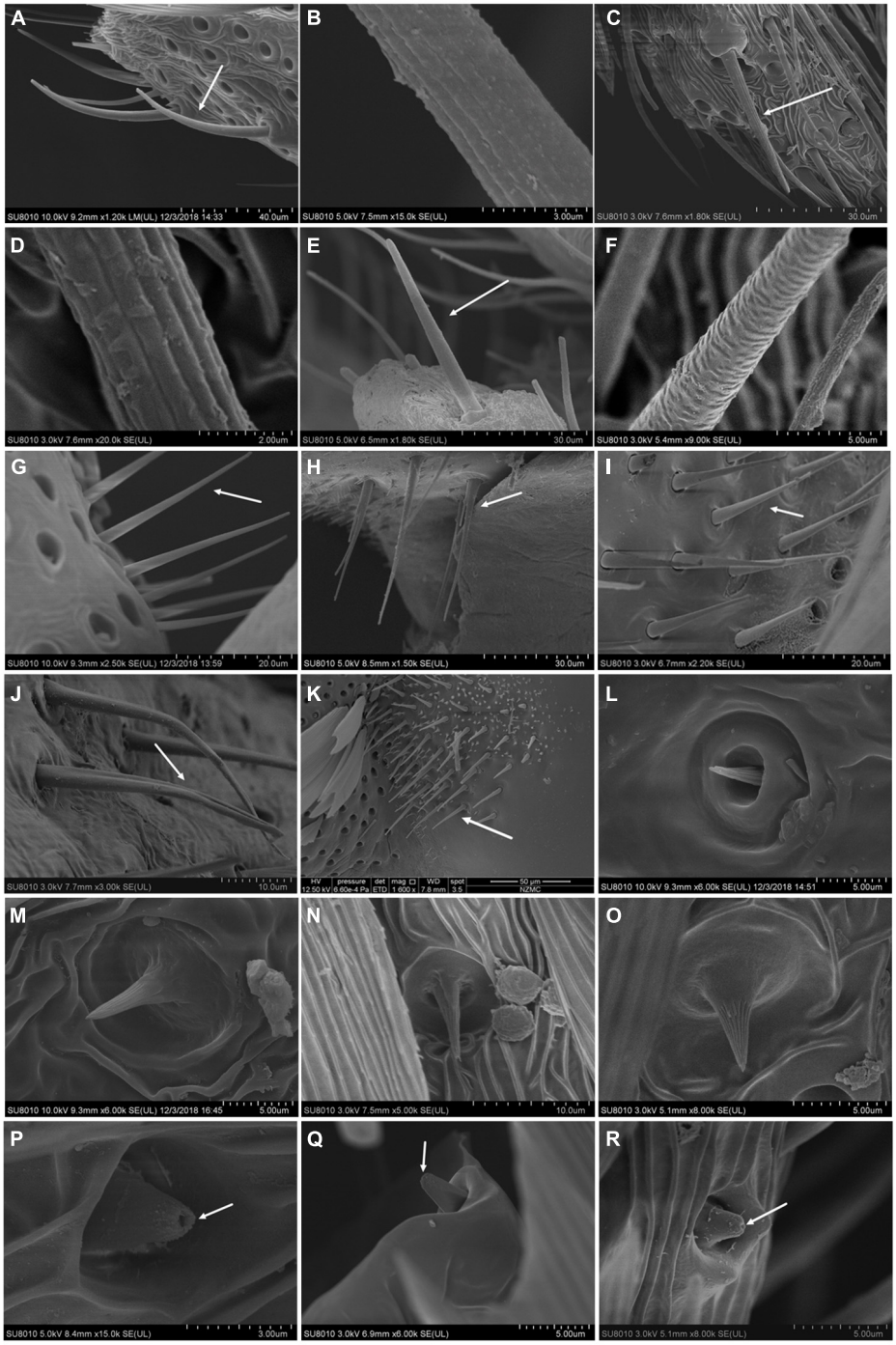


FIGURE 1 | Different types of sensilla in three species of Limacodidae by SEM. **(A–F)** Sensilla chaetica, **(A,B)** SCh I of *N. flavidorsalis*; **(C,D)** SCh I of *C. castaneipars*; **(E,F)** SCh II of *S. postornata*; **(G–K)** Böhm's bristles, **(G)** BB I of *N. flavidorsalis*, **(H)** BB II of *N. flavidorsalis*, **(I)** BB I of *C. castaneipars*, **(J)** BB I of *S. postornata*, **(K)** BB II of *S. postornata*; **(L–O)** Sensilla coeloclnica, **(L)** SCo I of *N. flavidorsalis*, **(M)** SCo II of *N. flavidorsalis*, **(N)** SCo II of *C. castaneipars*; **(O)** SCo II of *S. postornata*. **(P–R)** Sensilla uniporous peg, **(P)** *N. flavidorsalis*; **(Q)** *C. castaneipars*; **(R)** *S. postornata*. The arrow points to the hole at the tip of the sensilla.

SCh I among species, but there was a considerable difference in sensilla length among species and sex; the sensilla from females is significantly longer than males (**Table 4**). The morphological differences of sensilla among species were mainly reflected in females. There were significant differences in the length of sensilla BB I between *S. postornata* and the other two species, which was significantly longer than others in females (**Table 4**). There were significant differences in the *Sst*I between the female *S. postornata*

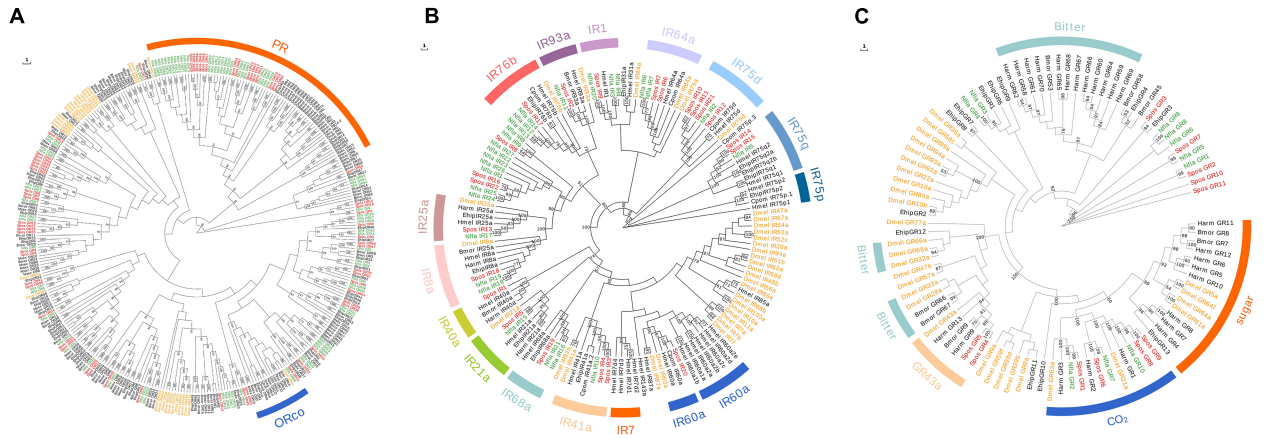


FIGURE 2 | Maximum-likelihood phylogenetic tree based on protein sequences of candidate odorant receptors (ORs) **(A)**, ionotropic receptors (IRs) **(B)**, and gustatory receptors (GRs) **(C)**. **(A)** The ML phylogenetic analysis of ORs of *N. flavidorsalis* (Nfla OR, green) and *S. postornata* (Spos OR red) were performed with reference ORs of *D. melanogaster* (Dmel OR, yellow) and ORs of Lepidoptera species (black). The orange arch refers to PR lineage, and the blue arch refers to ORco. **(B)** The ML phylogenetic analysis of IRs of *N. flavidorsalis* (Nfla IR, green) and *S. postornata* (Spos IR red) were performed with reference ORs of *D. melanogaster* (Dmel IR, yellow) and IRs of Lepidoptera species (black). **(C)** The ML phylogenetic analysis of GRs of *N. flavidorsalis* (Nfla GR, green) and *S. postornata* (Spos GR, red) were performed with reference GRs of *D. melanogaster* (Dmel IR, yellow) and GRs of Lepidoptera species (black). The stability of the nodes was assessed by bootstrap analysis with 1000 replications, and only bootstrap values ≥ 0.6 are shown at the corresponding nodes.

and the other female samples, such as the length, basal width, and the size of conical extremist of sensilla (**Table 4**).

Transcriptome Assembly

The transcriptomic sequence data were generated using the antenna cDNA library and illumina HiSeq2500 sequencing platform. A total of 76,763,082 and 101,749,890 raw reads were obtained from *N. flavidorsalis* and *S. postornata*, respectively. After removing adaptor sequences, low-quality sequences, and N-containing sequences, approximately 73.3 million and 99.0 million clean reads were retained. The assembly of all clean reads together led to the generation of 148,264 and 198,318 unigenes (**Table 5**). The clean reads for *N. flavidorsalis* and *S. postornata.* have been deposited in the NCBI SRA database under the accession numbers SRR15330236 and SRR15330240.

Functional Annotation of Unigenes

We used the unigenes assembled in the transcriptome as the queries in BLASTx searches of the GO database, and 11,543 unigenes of *N. flavidorsalis* and 15,902 unigenes of *S. postornata* were annotated. All of the unigenes were divided into three categories: molecular function, cellular component, or biological process according to the biological processes and functional annotations. GO annotation indicated that the antennal transcriptomes of these two species were highly similar concerning GO terms (**Supplementary Figure 8**). In the biological process terms, single-organism, cellular, and metabolic occupied the majority of both differentially expressed unigenes. Cell and cell parts were the most abundant for all unigenes in the cellular component terms. In the molecular function category, catalytic activity, binding, and transporter activity had huge preponderance; however, unigenes in "signal transducer activity" and "chemoattractant activity" were also present (**Supplementary Figure 8**).

Identification of Putative Genes Related to Transporting Odorant Molecules

There were more putative OBPs identified in *S. postornata* than in *N. flavidorsalis*; however, the numbers of putative CSPs and SNMPs have no significant difference between these two species. A number of fourteen and eighteen OBPs, including common odor-binding protein (GOBPs) and sex pheromone-binding protein (PBPs), were identified in the antennae transcriptome of *N. flavidorsalis* and *S. postornata*, separately. The sequence identities of the OBPs with other lepidopteran insects ranged from 40 to 92% in the NCBI database (**Supplementary Table 2**). A phylogenetic tree of the OBPs was constructed based on the orthologs from *Drosophila melanogaster* (Dmel) and the three lepidopteran species, which are *Bombyx mori* (Bmor), *Helicoverpa armigera* (Harm), and *Lobesia botrana* (Lbot) (**Figure 3A**). The phylogenetic analysis demonstrated that the lepidopteran PBP and GOBP sequences were highly conserved and clustered into three lineage-specific clades according to their different functions. Meanwhile, other OBPs showed an extremely divergent trend. There were two GOBPs (*Spos_OBP15* and *Spos_OBP16*) and one PBP (*Spos_OBP18*) identified in *S. postornata*, while only one GOBP (*Nfla_OBP10*) was detected in *N. flavidorsalis.* In each Limacodidae species, 14 unigenes were identified as CSP genes (**Supplementary Table 3**). All CSPs shared high sequence identities to known lepidopteran CSPs with an average of 67% identity. Of the 14 unigenes in *N. flavidorsalis*, 12 contained full-length ORFs encoding 107–152 amino acid residues. Unlike *N. flavidorsalis*, 9 out of 14 unigenes containing full-length ORFS were detected in *S. postornata.* A phylogenetic tree of the CSP was constructed based on the orthologs from

TABLE 4 | Morphological measurements of different types of antennal sensilla in three species of Limacodidae (Mean ± SE) (*N* = 20).

Antennal sensilla		*N. flavidorsalis*		*C. castaneipars*		*S. postornata*	
		Female	Male	Female	Male	Female	Male
Sensilla trichodea (STr)	L	30.21 ± 1.71^{a}	31.53 ± 2.9^{ab}	40.40 ± 2.51^{c}	34.54 ± 3.61^{abc}	37.81 ± 1.60^{bc}	36.12 ± 2.64^{abc}
	BW	2.72 ± 0.11^{a}	2.82 ± 0.12^{a}	3.80 ± 0.22^{a}	2.84 ± 0.25^{a}	2.81 ± 0.11^{a}	2.85 ± 0.21^{a}
Sensilla basiconica (SB)	L	8.34 ± 0.52^{a}	10.24 ± 1.00^{a}	-	-	-	-
	BW	2.94 ± 0.20^{a}	3.71 ± 0.41^{a}	-	-	-	-
Sensilla chaetica I(SCh I)	L	51.65 ± 2.61^{c}	45.04 ± 3.62^{ab}	50.0 ± 1.92^{bc}	41.5 ± 3.7^{a}	-	-
	BW	3.98 ± 0.11^{a}	5.44 ± 2.82^{a}	4.25 ± 0.21^{a}	3.8 ± 0.2^{a}	-	-
Sensilla chaetica II(SCh II)	L	-	-	-	-	43.61 ± 3.21^{a}	39.32 ± 3.30^{a}
	BW	-	-	-	-	4.44 ± 0.20^{a}	4.34 ± 0.22^{a}
Böhm's bristles I(BB I)	L	26.32 ± 4.51^{a}	35.10 ± 2.22^{b}	26.42 ± 1.51^{a}	32.30 ± 2.30^{ab}	33.54 ± 0.81^{b}	32.21 ± 3.72^{ab}
	BW	3.21 ± 0.11^{c}	3.15 ± 0.12^{bc}	2.91 ± 0.10^{b}	2.65 ± 0.14^{a}	3.08 ± 0.10^{b}	3.05 ± 0.24^{bc}
Böhm's bristles II (BB II)	L	-	32.12 ± 3.22^{a}	-	36.58 ± 2.62^{a}	-	-
	BW	-	3.51 ± 0.21^{a}	-	3.84 ± 0.24^{a}	-	-
Sensilla squamiformia (SSq)	L	-	97.42 ± 3.20	-	-	-	-
	BW	-	4.12 ± 0.40	-	-	-	-
Sensilla styloconica I (SSt I)	L	48.32 ± 2.41^{d}	40.65 ± 0.71^{bcd}	44.65 ± 4.01^{cd}	39.14 ± 3.32^{abcd}	31.71 ± 0.42^{a}	34.19 ± 1.42^{ab}
	CEL	4.83 ± 0.30^{abc}	5.44 ± 0.61^{abc}	5.84 ± 1.00^{c}	5.74 ± 0.32^{bc}	4.27 ± 0.31^{ab}	4.04 ± 0.44^{a}
	BW	8.70 ± 0.51^{b}	6.90 ± 0.61^{ab}	8.28 ± 0.80^{ab}	8.20 ± 0.64^{ab}	6.65 ± 0.30^{a}	8.52 ± 0.51^{ab}
Sensilla styloconica II (SSt II)	L	3.82 ± 0.12^{a}	4.61 ± 0.10^{a}	3.91 ± 0.12^{a}	3.35 ± 0.12^{a}	5.24 ± 0.14^{a}	2.18 ± 0.11^{a}
	CEL	3.91 ± 0.21^{a}	3.95 ± 0.08^{a}	3.01 ± 0.11^{a}	5.12 ± 0.15^{a}	3.81 ± 0.11^{a}	2.67 ± 0.13^{a}
	BWE	2.61 ± 0.17^{a}	2.45 ± 0.12^{a}	1.90 ± 0.14^{a}	2.44 ± 0.04^{a}	6.53 ± 0.10^{a}	1.73 ± 0.12^{a}
Sensilla coeloclnica I(SCo I)	L	4.42 ± 0.51^{a}	5.24 ± 0.73^{a}	-	-	-	-
	BW	2.45 ± 0.20^{a}	2.86 ± 0.55^{a}	-	-	-	-
	DP	4.37 ± 0.33^{a}	4.65 ± 0.81^{a}	-	-	-	-
Sensilla coeloclnica II (SCo II)	L	6.02 ± 0.50^{b}	5.27 ± 0.33^{a}	5.13 ± 0.42^{a}	4.57 ± 0.44^{a}	4.85 ± 0.13^{a}	5.17 ± 0.41^{a}
	BW	2.79 ± 0.21^{a}	2.82 ± 0.34^{a}	3.07 ± 0.63^{a}	2.98 ± 0.25^{a}	2.68 ± 0.22^{a}	2.66 ± 0.12^{a}
	DP	9.88 ± 0.65^{b}	9.37 ± 0.44^{b}	7.48 ± 0.42^{a}	7.88 ± 0.54^{a}	8.27 ± 0.41^{a}	8.29 ± 0.61^{a}
Sensilla uniporous peg (SUp)	L	3.64 ± 0.28^{a}	4.42 ± 0.1^{b}	3.42 ± 0.1^{a}	-	-	-
	BW	2.52 ± 0.17^{a}	2.92 ± 0.14^{a}	2.54 ± 0.12^{a}	-	-	-
Sensilla furcatea I (SFu I)	L	12.72 ± 1.07^{b}	9.66 ± 1.92^{a}	7.82 ± 0.41^{a}	8.22 ± 0.81^{a}	6.54 ± 0.41^{a}	4.34 ± 0.72^{a}
	BW	3.13 ± 0.34^{a}	2.87 ± 0.42^{a}	2.67 ± 0.12^{a}	2.86 ± 0.22^{a}	2.24 ± 0.16^{a}	4.54 ± 1.33^{a}
Sensilla furcatea II (SFu II)	L	-	-	-	-	23.07 ± 0.08^{a}	27.32 ± 0.07^{a}
	BW	-	-	-	-	2.42 ± 0.14^{a}	2.68 ± 0.21^{a}

L, length; BW, basal width; DP, diameter of the pit; BWE, basal width of conical extremity; CEL, length of conical extremity. Different letters (a,b,c,d) represent significant differences ($p < 0.05$) between species in each row.

D. melanogaster (Dmel), *Helicoverpa armigera* (Harm), *B. mori* (Bmor), and *Eogystia hippophaecolus* (Ehip). The amino acid identities between the orthologous CSPs in the two moths were relatively high, and most of the CSPs appeared in pairs on the dendrogram (**Figure 3B**). Only four SNMPs (3 with complete ORFs) were identified in *N. flavidorsalis*. In addition, five SNMPs were detected in *S. postornata*, but only one with complete ORF (**Supplementary Table 4**). As expected, SNMPs were clustered into two branches with other SNMP1 and SNMP2 orthologs from other lepidopteron (**Supplementary Figure 9**).

Identification of Putative Receptor-Encoding Genes

Unlike the results of OBPs, more ORs and IRs were detected in *N. flavidorsalis* than in *S. postornata*. We identified transcripts encoding 76 putative ORs in *N. flavidorsalis*, among which 33 likely represented full-length genes. In *S. postornata*, we identified 61 candidate OR genes comprising 32 full-length genes (**Supplementary Table 5**). Sex pheromone receptors (PRs) and co-receptor (ORco) were marked with orange and blue lines in the phylogenetic tree (**Figure 2**), in which both contain members from Lepidoptera and *Drosophila* (Dmel). This result shows that 15 putative PR genes of *S. postornata* were clustered into the PR subfamily and one into the ORco subfamily. Meanwhile, 27 putative PR genes and 1 putative ORco genes of *N. flavidorsalis* were assigned to PR and ORco subfamilies, respectively. The phylogenetic analysis showed a separate branch of Lepidoptera in OR, and the genes of PR and ORco subfamilies were relatively conservative (**Figure 2A**).

Ionotropic gene family prediction results showed that there were 27 members in the IR family of *N. flavidorsalis*, among which 17 genes were full length (the length of ORF ranged from 137 to 1,057 aa), while there were 22 members of *S. postornata*,

TABLE 5 | An overview of the sequencing and assembly process (after trinity).

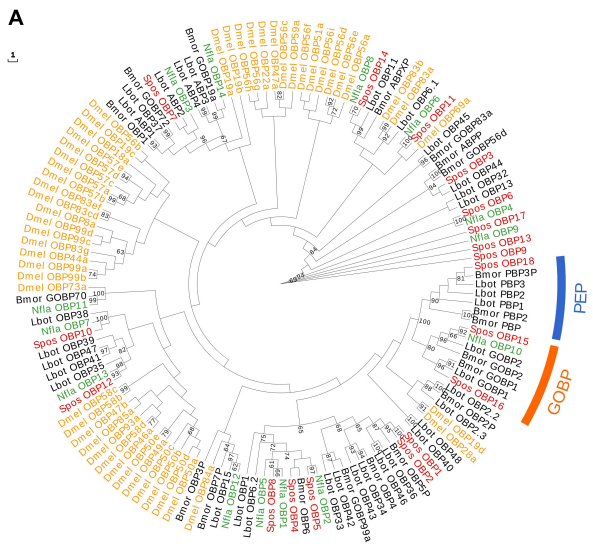
	N. flavidorsalis	*S. postornata*
Read Length	150	150
Total Raw Reads	76,763,082	101,749,890
Total Clean Reads	73,262,766	99,043,274
Total Clean Reads Ratio (%)	99.35	97.34
Clean Reads GC (%)	42.99	42.31
Clean Reads Q20 (%)	97.94	97.76
Clean Reads Q30 (%)	93.72	93.48
No. of Unigene	148,264	198,318
Max length	13,992	26,052
Min Length	201	201
Average Length	791	735
Unigene N90 length	307	260
Unigene N50 length	1,296	1,427
Unigene GC consent (%)	36.54	37.92

in which there were 10 with full length (the size of ORF ranged from 107 to 1,059 aa) (**Supplementary Table 6**). In the phylogenetic tree reconstructed with IR genes, the 16 IR gene member clusters were marked with different colors. The putative IRs in *N. flavidorsalis* were clustered into 11 known branches, while the putative IRs in *S. postornata* were classified into 13 branches. Meanwhile, other IR members from these two species were clustered together, which indicated that the IR family was highly conservative (**Figure 2B**).

Family branches were classified and labeled according to the functions identified by the previous studies (Jiang et al., 2015; Hu et al., 2016; Xu et al., 2016). For example, bMOR_GR7-9, BMOR_66-67 were sugar receptors, Harm_GR9 and 13 were GR_43a receptor (fructose receptor), Dmel_28a and Dmel_32 were bitter receptors, Dmel_64a and Dmel_64F were carbohydrate receptors, and Harm_1 ∼ 3 were CO2 receptors. Spos_GR1, Spos_GR6, Spos_GR8, Spos_GR9, Nfla_GR2, Nfla_GR7, and Nfla_GR10 were clustered to CO2 receptor branches, while Spos_GR4 and Spos_GR5 were clustered to GR43a (**Figure 2C** and **Supplementary Table 7**). These results showed that the evolution of different taste receptors was relatively conservative, while neither bitter receptor nor sugar receptor was found in the two species.

Suitable Habitat Distribution of Two Moths and Their Host Plants

The results of ENM showed areas suitable and unsuitable for the monophagous *N. flavidorsalis* and polyphagous *S. postornata* and the area of low to high habitat suitability for the corresponding host plants (**Figure 4**). The MaxEnt models using several feature classes (L, LQ, H, LQH, LQHP, and LQHPT) determine how predictor variables are transformed for each species or species combination. Based on the value of AUC_{DIFF} and delta AICc, the LQHPT was chosen as the best model. Generally, the suitable distribution area of the *S. postornata* was much larger than that of the *N. flavidorsalis,* which were remarkably consistent with the known distribution of these two species. Meanwhile, the suitable habitat distribution of *N. flavidorsalis* was uniform with the area of high suitability of its host plants, which indicated that the sympatric distribution of herbivorous insects and their host plants was obvious.



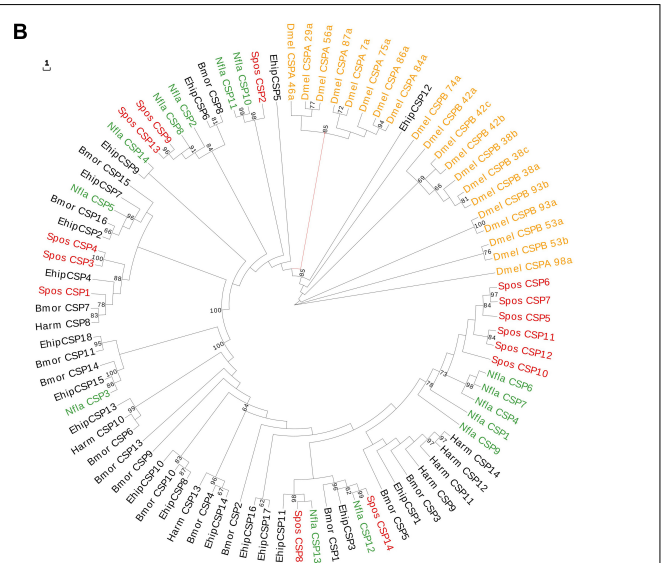


FIGURE 3 | Maximum-likelihood phylogenetic tree based on protein sequences of candidate odorant-binding proteins (OBPs) **(A)** and chemosensory proteins (CSPs) **(B)**. **(A)** The ML phylogenetic analysis of OBPs of *N. flavidorsalis* (Nfla OBP, green) and *S. postornata* (Spos OBP, red) were performed with reference OBPs of *D. melanogaster* (Dmel OBP, yellow) and OBPs of Lepidoptera species (black). The orange arch refers to PBP/GOBP lineage, and the blue arch refers to PEP. **(B)** The ML phylogenetic analysis of OBPs of *N. flavidorsalis* (Nfla CSP, green) and *S. postornata* (Spos CSP, red) were performed with reference CSPs of *D. melanogaster* (Dmel CSP, yellow) and CSPs of Lepidoptera species (black). The stability of the nodes was assessed by bootstrap analysis with 1000 replications, and only bootstrap values ≥ 0.6 are shown at the corresponding nodes.

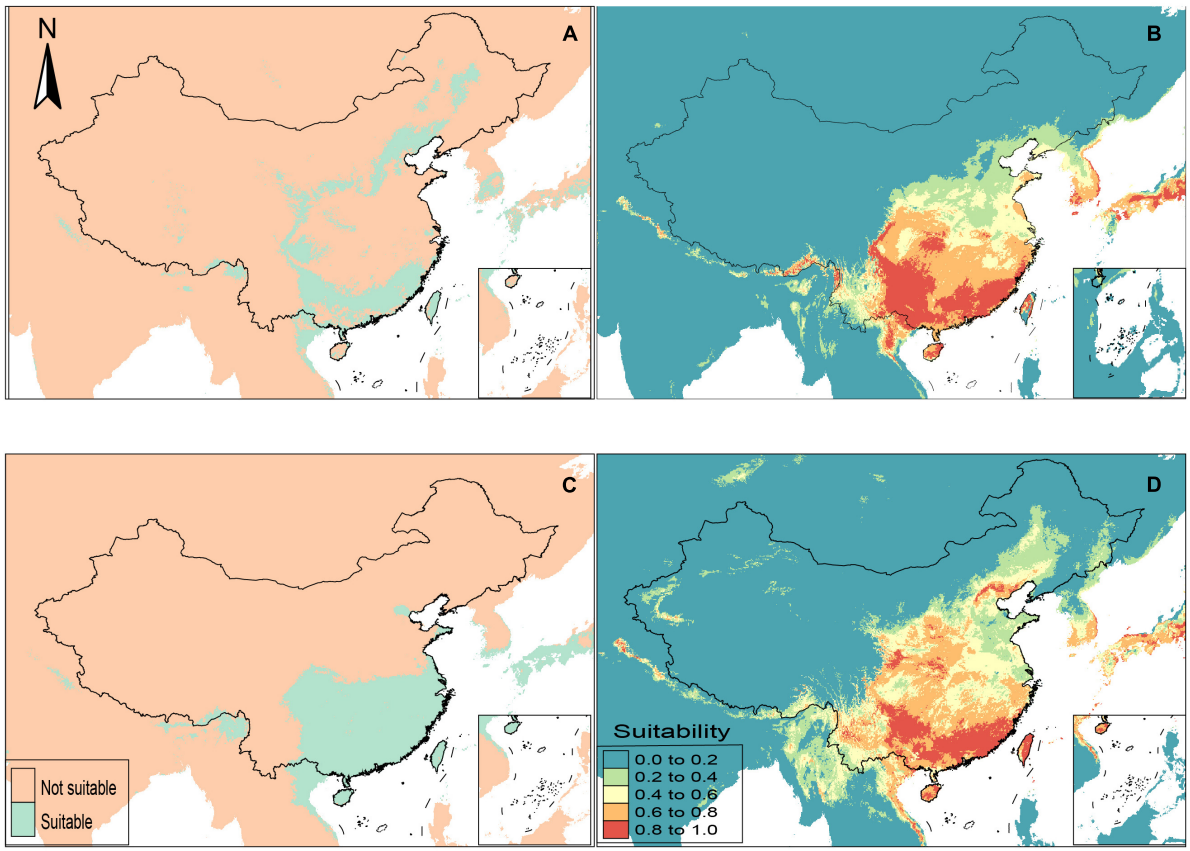


FIGURE 4 | Potential suitable distribution ranges of two slug moths (*N. flavidorsalis*, **A**; *S. postornata*, **B**), and their host plants predicted using climate as predictive factors in MaxEnt. **(A)** Potential distribution ranges of *N. flavidorsalis* in China; **(B)** potential distribution ranges of *Pyrus sorotina*, which is the host plant of *N. flavidorsalis*; **(C)** potential distribution ranges of *S. postornata* in China; **(D)** potential distribution ranges of three plants, *Juglans regia, Populus simonii*, and *Camellia sinensis*, which were selected as representative host plants of *S. postornata*.

DISCUSSION

Phytophagous insects rely on plant vegetative tissues as a food source to support larval development. Diet breadth, ranging from monophagy to polyphagy, plays significant roles in adapting to the host's defense mechanisms (s) and the exploitation of host-recognition cues. In this study, sensilla on the male and female antennae of three species of slug moths with significantly feeding habits differentiation were examined using a scanning electron microscope. Generally, the antennae of the three species are the same in shape and structure in the same gender. There were identified a total of nine types of sensilla. The two types of sensilla, SUp and SFu, had not been found in the previous studies of the slug moth *Monema flavescens* and *Iragoides fasciata* (Huang et al., 2012; Yang et al., 2017). This study is the first report on these two types of sensilla in species of Limacodidae. There are five types of sensilla observed with two subtypes (SCh, BB, SSt, SCo, and SFu), in which BB II and SFu II were not found in Limacodidae based on the previous studies. This phenomenon suggested that sensilla differentiation was related to adaptability to the environments.

Compared to the other two species, SB and SSq were only found in *N. flavidorsalis*. The SSq belonged to the mechanical sensilla and was related to the buffering gravity during flight (McIver, 1975; Sun et al., 2011). Moreover, both subtypes of SCo were detected in *N. flavidorsalis,* mostly thought to be related to sensing fluctuations in temperature and humidity (Yokohari et al., 1982). By comparing the number of sensillum types in the three herbivorous species, there was a trend of gradually decreasing the number of sensory types with the gradual expansion of feeding habitats. We speculated that the evolution of diverse sensillum types was more advantageous for detecting specific hosts based on this phenomenon. Thus, it was possible to find single hosts more accurately (Jermy, 1984; Agosta, 2006; Dermauw et al., 2018). In contrast, euryphagous insects tend to perceive only the odor molecules prevalent in multiple hosts, while a small variety of the sensillum types were needed (Zwölfer, 1982; Jermy, 1988). The antennal sensilla samples of superfamily Noctuoidea, which were studied more recently, were selected and compared to verify the speculation further (Wang et al., 2002, 2015; Jin et al., 2008; Seada, 2015). The results showed more sensilla types in omnivorous species than in oligophagous ones (**Supplementary Table 6**).

Meanwhile, the size and number of sensillum types of *N. flavidorsalis* and *S. postornata* differed. By comparing the

antennal sensilla of these two species, the results showed that oligophagous insects have relatively more abundant sensillum types than polyphagous one. For example, SB, SSq, and SCo I were detected only in *N. flavidorsalis,* while SCo II and SFu II were detected only in *S. postornata*. Among them, SB belongs to mechanoreceptor, while SB has the function of receptor pheromone. Moreover, the diameter and length of the base of SCh, SCo I, SSt I, and SFu I of *N. flavidorsalis* were all large and more than *S. postornata*. This phenomenon might be related to the species specificity and may also be related to the diet range, which needs further research. The external morphology of the antennal sensilla was examined using scanning electron microscopy, which could provide a better understanding of the mechanisms of insect–insect and insect–plant chemical communications. At present, the researchers are mainly focused on the taxonomy classification and phylogeny analysis (Tan et al., 2012; Wang et al., 2018), and the comparative studies with different feeding preferences were limited. However, our results are limited to quite a few species in Limacodidae, and further exploration should be carried out in more species and combined with electrophysiological methods.

Although most of the researches related to the olfactory system of Lepidoptera have proceeded, less is known about the olfactory mechanisms of slug moth (family Limacodidae), which are severe pests of cash crops and harmful to human health (Lin et al., 2019; Plata-Rueda et al., 2020). To better understand the vital role of olfactory proteins in localizing host plants, we investigated CSPs in the antennal transcriptomes of *N. flavidorsalis* and *S. postornata* using next-generation sequencing technology. The number of ORs and IRs family members of *N. flavidorsalis* was slightly higher than that of *S. postornata*. OBPs were considered the first gate in the odorant recognition process, especially for hydrophobic odors; they bind and transport odors, which includes pheromones and plant volatiles, across the lymph in the sensillum (Krieger and Breer, 1999). We identified 14 and 18 putative OBPs, respectively, of which we studied the expression of nine in antennae and other chemosensory tissues. In the dendrogram of OBPs, the PBP lineages and GOBP lineage comprised the PBP/GOBP complex, which supports the monophyletic of PBP/GOBP and PBP with more dynamic evolution than GOBP (Steinbrecht et al., 1995). The differences in the total number of putative OBP genes could be potentially attributed to their specialized ecology. It is also possible that some OBPs have not been accurately identified because of the limitation of library representation. Similar to OBPs, CSP-encoding genes were also expressed ubiquitously in insects and have been thought to be involved in chemoreception and participate in other physiological processes (Iovinella et al., 2013). Based on the phylogenetic topology, almost all CSPs of Diptera formed a taxon-specific clade. According to the divergence of insect orders, diversification has also been observed in other lepidopteron. The numbers of putative CSP-genes members of *N. flavidorsalis* and *S. postornata* were both 14. Besides, due to the evolutionarily conserved CSP gene family, the putative CSP genes of the two species were also clustered together. The SNMPs are conserved throughout holometabolous insects (Suh et al., 2014); this analysis suggests two SNMP sub-clades: SNMP1 and SNMP2. Moreover, SNMP1 and SNMP2 cluster in a monophyletic group, respectively, which was consistent with the previous Lepidoptera studies (Xu et al., 2021).

odorant receptors connect binding proteins with olfactory sensory neurons and conduct olfactory signal transduction. The putative OR gene members were the highest in *N. flavidorsalis* (76) and *S. postornata* (61), respectively. According to the phylogenetic topology, an atypical odor receptor (Orco) was identified in both species, which was consistent with most members of the insect OR family and showed highly evolutionary conservation. According to phylogenetic analysis, the lineage of IR gene members was highly conserved, and one IR25a member was identified in both species that could participate in sensing temperature changes (Chen et al., 2015). The three IR76b, which was sensitive to low concentration salt, were detected in *N. flavidorsalis*, yet only one member of IR76b was found in *S. postornata* (Zhang et al., 2013; Lee et al., 2017). This implied that *N. flavidorsalis* might be more sensitive to salt ions at lower concentrations, which may be related to its host specificity (Chen et al., 2019). Notably, transcripts putatively encoding IR8a, which are thought to function as IR co-receptors, were also found in slug moth (Ai et al., 2013; Rytz et al., 2013; Zhang et al., 2019). GR gene family is a typical function in sensing sugar, CO_2, and bitter molecules (Wanner and Robertson, 2008). We also detected 10 and 11 putatively GRs in the antennal transcriptomes in *N. flavidorsalis* and *S. postornata*, respectively, which provided important sequence information. In the phylogenetic tree, GRs involved in detecting CO_2 molecules were clustered in a group with other species. GRs engaged in detecting sugar and bitter molecules were not detected. Unfortunately, this phenomenon may be related to the species-specific distribution of GR receptors and sample sequencing depth.

To further clarify the effects of host plant distribution on herbivorous insects, the ecological niche model analyses proceeded on *N. flavidorsalis*, *S. postornata*, and their host plants in China. Their occurrence data were collected mainly through extensive field sampling across China and GBIF. The results indicate that there was a clear correlation with the availability of host plants and the suitable distribution area of herbivorous insects. The oligophagous insects may accurately locate the high-density distribution area of their host plants with the help of their complex olfactory sensing system. Not only that plant diversity could affect insect diet breadth (Forister et al., 2015), which may, in turn, feed back onto plant diversity with either coevolutionary or ecological interactions. As a matter of fact, in addition to host plants, species distributions of insects are also affected by multiple factors such as climate, precipitation, soil, and natural enemies (Dang et al., 2021). In the future, a refined model would provide more accurate information to predict the relationship of the potential distribution between the insect and host plants.

In conclusion, we compared the antennal sensilla structures of three species of slug moth with different diet breadth. A total of 9 types of sensilla were identified, in which SUp and SFu were first reported in the family Limacodidae. Furthermore, there was a trend of gradually decreasing the number of sensillum types with the gradual expansion of feeding habitats, which was consistent with that found in Noctuidae insects. However,

there was no correlation between the number of olfactory-related genes (including receptor-encoding genes and genes related to transport odorant molecules) and the increase of antennal sensilla types. There is no doubt that further research is needed to test this phenomenon. Our studies will provide novel ideas for developing bioinsecticides and facilitate further study on the plant–insect interactions.

DATA AVAILABILITY STATEMENT

The clean reads for *N. flavidorsalis* and *S. postornata* have been deposited in the NCBI SRA database under the accession numbers SRR15330236 and SRR15330240.

AUTHOR CONTRIBUTIONS

JL and ABZ designed the study. YMY and JL collected and analyzed the data. YW, CQY, and GFW contributed to explaining the outputs from software and models. CSW contributed to identifying species. JL wrote the manuscript together with YMY and ABZ. All authors participated in the preparation of this manuscript.

FUNDING

This work was supported by the Natural Science Foundation of China (Grant Nos. 32170421 and 31772501), the China National Funds for Distinguished Young Scientists (Grant No. 31425023), the Program of Ministry of Science and Technology of China (Grant No. 2018FY100403), and the Academy for Multidisciplinary Studies, Capital Normal University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2022.845922/full#supplementary-material

REFERENCES

Ai, M., Blais, S., Park, J. Y., Min, S., Neubert, T. A., and Suh, G. S. (2013). Ionotropic glutamate receptors IR64a and IR8a form a functional odorant receptor complex in vivo in *Drosophila*. *J. Neurosci.* 33, 10741–10749. doi: 10.1523/JNEUROSCI.5419-12.2013

Agosta, S. J. (2006). On ecological fitting, plant–insect associations, herbivore host shifts, and host plant selection. *Oikos* 114, 556–565. doi: 10.1111/ele.12555

Ahmad, S. (ed.) (2012). *Herbivorous Insects: Host-Seeking Behavior And Mechanisms*. Amsterdam: Elsevier.

Bengtsson, J. M., Gonzalez, F., Cattaneo, A. M., Montagné, N., Walker, W. B., Bengtsson, M., et al. (2014). A predicted sex pheromone receptor of codling moth *Cydia pomonella* detects the plant volatile pear ester. *Front. Ecol. Evol.* 2:33. doi: 10.3389/fevo.2014.00033

Bernays, E. A. (1997). Feeding by lepidopteran larvae is dangerous. *Ecol. Entomol.* 22, 121–123. doi: 10.1046/j.1365-2311.1997.00042.x

Bian, D., Ye, W., Dai, M., Lu, Z., Li, M., Fang, Y., et al. (2020). Phylogenetic relationships of Limacodidae and insights into the higher phylogeny of Lepidoptera. *Int. J. Biol. Macromol.* 159, 356–363. doi: 10.1016/j.ijbiomac.2020.05.023

Brown, J., Pirrung, M., and McCue, L. A. (2017). FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics* 33, 3137–3139. doi: 10.1093/bioinformatics/btx373

Bruce, T. J., and Pickett, J. A. (2011). Perception of plant volatile blends by herbivorous insects–finding the right mix. *Phytochemistry* 72, 1605–1611. doi: 10.1016/j.phytochem.2011.04.011

Cates, R. G. (1981). Host plant predictability and the feeding patterns of monophagous, oligophagous, and polyphagous insect herbivores. *Oecologia* 48, 319–326. doi: 10.1007/BF00346488

Chapman, R. F. (1998). *The Insects: Structure And Function*. Cambridge: Cambridge university press.

Checker, V. G., and Sharma, M. (2021). "Signalling during insect plant interaction," in *Plant-Pest Interactions: From Molecular Mechanisms to Chemical Ecology*, eds I. K. Singh and A. Singh (Singapore: Springer), 193–214. doi: 10.1007/978-981-15-2467-7_9

Chen, C., Buhl, E., Xu, M., Croset, V., Rees, J. S., Lilley, K. S., et al. (2015). *Drosophila* ionotropic receptor 25a mediates circadian clock resetting by temperature. *Nature* 527, 516–520. doi: 10.1038/nature16148

Chen, H. L., Stern, U., and Yang, C. H. (2019). Molecular control limiting sensitivity of sweet taste neurons in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A* 116, 20158–20168. doi: 10.1073/pnas.1911583116

Conant, P., Hara, A. H., Nagamine, W. T., Kishimoto, C. M., and Heu, R. A. (2002). *Nettle Caterpillar Darna pallivitta Moore (Lepidoptera: Limacodidae). New Pest Advisory No. 01-03.* Honolulu, HI: State of Hawai'i, Department of Agriculture, 27–30.

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610

Dang, Y. Q., Zhang, Y. L., Wang, X. Y., Xin, B., Quinn, N. F., and Duan, J. J. (2021). Retrospective analysis of factors affecting the distribution of an invasive wood-boring insect using native range data: the importance of host plants. *J. Pest Sci.* 94, 981–990. doi: 10.1007/s10340-020-01308-5

Dermauw, W., Pym, A., Bass, C., Van Leeuwen, T., and Feyereisen, R. (2018). Does host plant adaptation lead to pesticide resistance in generalist herbivores? *Curr. Opin. Insect Sci.* 26, 25–33. doi: 10.1016/j.cois.2018.01.001

Duke, N. C. (2002). Sustained high levels of foliar herbivory of the mangrove Rhizophora stylosa by a moth larva Doratifera stenosa (Limacodidae) in north-eastern Australia. *Wetl. Ecol. Manag.* 10, 403–419. doi: 10.1007/bf03263357

Forister, M. L., Novotny, V., Panorska, A. K., Baje, L., Basset, Y., Butterill, P. T., et al. (2015). The global distribution of diet breadth in insect herbivores. *Proc. Natl. Acad. Sci. U.S.A* 112, 442–447. doi: 10.1073/pnas.1423042112

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., et al. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic. Acids. Res.* 36, 3420–3435. doi: 10.1093/nar/gkn176

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Hansson, B. S. (1995). Olfaction in lepidoptera. *Experientia* 51, 1003–1027. doi: 10.1007/bf01946910

Harris, M. O., Stuart, J. J., Mohan, M., Nair, S., Lamb, R. J., and Rohfritsch, O. (2003). Grasses and gall midges: plant defense and insect adaptation. *Annu. Rev. Entomol.* 48, 549–577. doi: 10.1146/annurev.ento.48.091801.112559

Hu, P., Tao, J., Cui, M., Gao, C., Lu, P., and Luo, Y. (2016). Antennal transcriptome analysis and expression profiles of odorant binding proteins in *Eogystia hippophaecolus* (Lepidoptera: Cossidae). *BMC Genomics* 17:651. doi: 10.1186/s12864-016-3008-4

Huang, A. P., Bao, X. C., Liu, B. Y., Wang, Y. J., Zhou, L. Y., Ning, J., et al. (2012). Electroantennogram responses of the tea slug moth, Iragoides fasciata to some plant volatiles associated with tea, Camellia sinensis. *J. Insect Sci.* 12:75. doi: 10.1673/031.012.7501

Hunter, M. D., and McNeil, J. N. (1997). Host-plant quality influences diapause and voltinism in a polyphagous insect herbivore. *Ecology* 78, 977–986. doi: 10.1890/0012-9658(1997)078[0977:hpqida]2.0.co;2

Iovinella, I., Bozza, F., Caputo, B., Della Torre, A., and Pelosi, P. (2013). Ligand-binding study of Anopheles gambiae chemosensory proteins. *Chem. Senses.* 38, 409–419. doi: 10.1093/chemse/bjt012

Ivanov, V. D., and Melnitsky, S. I. (2016). Diversity of the olfactory sensilla in caddisflies (Trichoptera). *Zoosymposia* 10, 224–233. doi: 10.11646/zoosymposia.10.1.20

Jaffar-Bandjee, M., Steinmann, T., Krijnen, G., and Casas, J. (2020). Insect pectinate antennae maximize odor capture efficiency at intermediate flight speeds. *Proc. Natl. Acad. Sci. U.S.A* 117, 28126–28133. doi: 10.1073/pnas.2007871117

Jefferson, R. N., Rubin, R. E., McFarland, S. U., and Shorey, H. H. (1970). Sex pheromones of noctuid moths. XXII. The external morphology of the antennae of *Trichoplusia* ni, *Heliothis* zea, Prodenia *ornithogalli*, and *Spodoptera exigua*. *Ann. Entomol. Soc. Am.* 63, 1227–1238. doi: 10.1093/aesa/63.5.1227

Jermy, T. (1984). Evolution of insect/host plant relationships. *Am. Nat.* 124, 609–630. doi: 10.1086/284302

Jermy, T. (1988). Can predation lead to narrow food specialization in phytophagous insects? *Ecology* 69, 902–904. doi: 10.2307/1941241

Jiang, X. C., Liu, S., Jiang, X. Y., Wang, Z. W., Xiao, J. J., Gao, Q., et al. (2021). Identification of olfactory genes from the greater wax moth by antennal transcriptome analysis. *Front. Physiol.* 12:663040. doi: 10.3389/fphys.2021.663040

Jiang, X. J., Ning, C., Guo, H., Jia, Y. Y., Huang, L. Q., Qu, M. J., et al. (2015). A gustatory receptor tuned to D-fructose in antennal sensilla chaetica of *Helicoverpa armigera. Insect Biochem. Mol. Biol.* 60, 39–46.

Jin, X., Li, Y. P., and Cui, W. N. (2008). Morphological studies on the types and distribution of antennal sensilla in *Spodoptera exigua* (Hübner). *J. Northwest Sci-Tech Univ. Agri. For.* 036, 189–193.

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kaupp, U. B. (2010). Olfactory signalling in vertebrates and insects: differences and commonalities. *Nat. Rev. Neurosci.* 11, 188–200. doi: 10.1038/nrn2789

Krieger, J., and Breer, H. (1999). Olfactory reception in invertebrates. *Science* 286, 720–723. doi: 10.1126/science.286.5440.720

Lancaster, L. T. (2020). Host use diversification during range shifts shapes global variation in Lepidopteran dietary breadth. *Nat. Ecol. Evol.* 4, 963–969. doi: 10.1038/s41559-020-1199-1

Lee, M. J., Sung, H. Y., Jo, H., Kim, H. W., Choi, M. S., Kwon, J. Y., et al. (2017). Ionotropic receptor 76b is required for gustatory aversion to excessive Na+ in *Drosophila*. *Mol. Cells* 40, 787–790. doi: 10.14348/molcells.2017.0160

Levins, R., and MacArthur, R. (1969). A hypothesis to explain the incidence of monophagy. *Ecology* 50, 910–911. doi: 10.2307/1933709

Lill, J. T., Marquis, R. J., Forkner, R. E., Le Corff, J., Holmberg, N., and Barber, N. A. (2006). Leaf pubescence affects distribution and abundance of generalist slug caterpillars (Lepidoptera: Limacodidae). *Environ. Entomol.* 35, 797–806. doi: 10.1603/0046-225x-35.3.797

Lin, Y. C., Lin, R. J., Braby, M. F., and Hsu, Y. F. (2019). Evolution and losses of spines in slug caterpillars (Lepidoptera: Limacodidae). *Ecol. Evol.* 9, 9827–9840. doi: 10.1002/ece3.5524

McIver, S. B. (1975). Structure of cuticular mechanoreceptors of arthropods. *Annu. Rev. Entomol.* 20, 381–397. doi: 10.1146/annurev.en.20.010175.002121

Milne, M., and Walter, G. H. (2000). Feeding and breeding across host plants within a locality by the widespread thrips *Frankliniella schultzei*, and the invasive potential of polyphagous herbivores. *Divers. Distrib.* 6, 243–257. doi: 10.1046/j.1472-4642.2000.00089.x

Mithöfer, A., and Boland, W. (2012). Plant defense against herbivores: chemical aspects. *Annu. Rev. Plant Biol.* 63, 431–450.

Murphy, S. M., Leahy, S. M., Williams, L. S., and Lill, J. T. (2010). Stinging spines protect slug caterpillars (Limacodidae) from multiple generalist predators. *Behav. Ecol.* 21, 153–160. doi: 10.1093/beheco/arp166

van Nieukerken, E. J., Kaila, L., Kitching, I. J., Kristensen, N. P., Lees, D. C., Minet, J., et al. (2011). "Order Lepidoptera Linnaeus, 1758," in *Animal Biodiversity: An Outline Of Higher-Level Classification And Survey of Taxonomic Richness*, ed. Z.-Q. Zhang (Auckland: Magnolia Press), 212–221.

Onagbola, E. O., and Fadamiro, H. Y. (2008). Scanning electron microscopy studies of antennal sensilla of Pteromalus cerealellae (Hymenoptera: Pteromalidae). *Micron* 39, 526–535. doi: 10.1016/j.micron.2007.08.001

Perveen, F. K. (2017). "Lepidoptera," in *BoD–Books on Demand*, Vol. 1, ed. F. K. Perveen (London: IntechOpen), 3–17.

Peterson, A. T., and Soberón, J. (2012). Species distribution modeling and ecological niche modeling: getting the concepts right. *Nat. Conserv.* 10, 102–107.

Petschenka, G., and Agrawal, A. A. (2015). Milkweed butterfly resistance to plant toxins is linked to sequestration, not coping with a toxic diet. *Proc. R. Soc. B-Biol. Sci.* 282:20151865. doi: 10.1098/rspb.2015.1865

Petschenka, G., and Agrawal, A. A. (2016). How herbivores coopt plant defenses: natural selection, specialization, and sequestration. *Curr. Opin. Insect Sci.* 14, 17–24. doi: 10.1016/j.cois.2015.12.004

Plata-Rueda, A., Quintero, H. A., Serrão, J. E., and Martínez, L. C. (2020). Insecticidal activity of Bacillus thuringiensis strains on the nettle caterpillar, Euprosterna elaeasa (Lepidoptera: Limacodidae). *Insects* 11:310. doi: 10.3390/insects11050310

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., et al. (2005). InterProScan: protein domains identifier. *Nucleic Acids Res.* 33, W116–W120. doi: 10.1093/nar/gki442

Rank, N., Smiley, J., and Alfered, K. (1996). "Natural enemies and host plant relationship for Chrysomeline leaf beetles feeding on Salicaea," in *Chrysomelidae Biology, Volume 2:Ecological Studies*, eds P. Jolivet and M. L. Cox (Amsterdam: SPB Academic Publishing), 147–171.

Rytz, R., Croset, V., and Benton, R. (2013). Ionotropic receptors (IRs): chemosensory ionotropic glutamate receptors in *Drosophila* and beyond. *Insect Biochem. Mol. Biol.* 43, 888–897. doi: 10.1016/j.ibmb.2013.02.007

Salgado, A. L., and Saastamoinen, M. (2019). Developmental stage-dependent response and preference for host plant quality in an insect herbivore. *Anim. Behav.* 150, 27–38. doi: 10.1016/j.anbehav.2019.01.018

Sánchez-Gracia, A., Vieira, F. G., and Rozas, J. (2009). Molecular evolution of the major chemosensory gene families in insects. *Heredity* 103, 208–216. doi: 10.1038/hdy.2009.55

Seada, M. A. (2015). Antennal morphology and sensillum distribution of female cotton leaf worm Spodoptera littoralis (Lepidoptera: Noctuidae). *J. Basic Appl. Zool.* 68, 10–18. doi: 10.1016/j.jobaz.2015.01.005

Schneider, D. (1964). Insect antennae. *Ann. Rev. Entomol.* 9, 103–122.

Shin, Y., Min, M. S., and Borzée, A. (2021). Driven to the edge: species distribution modeling of a Clawed Salamander (Hynobiidae: Onychodactylus koreanus) predicts range shifts and drastic decrease of suitable habitats in response to climate change. *Ecolo. Evol.* 11, 14669–14688. doi: 10.1002/ece3.8155

Stanton, M. L. (1983). "Spatial patterns in the plant community and their effects upon insect search," in *Herbivorous Insects: Host-Seeking Behavior and Mechanisms*, ed. S. Ahmad (Amsterdam: Elsevier), 125–157. doi: 10.1111/plb.12644

Steinbrecht, R. A. (1998). Odorant-binding proteins: expression and function. *Ann. NY Acad. Sci.* 855, 323–332. doi: 10.1111/j.1749-6632.1998.tb10591.x

Steinbrecht, R. A., Laue, M., and Ziegelberger, G. (1995). Immunolocalization of pheromone-binding protein and general odorant-binding protein in olfactory sensilla of the silk moths *Antheraea* and *Bombyx*. *Cell Tissue Res.* 282, 203–217. doi: 10.1007/bf00319112

Suh, E., Bohbot, J. D., and Zwiebel, L. J. (2014). Peripheral olfactory signaling in insects. *Curr. Opin. Insect Sci.* 6, 86–92. doi: 10.1016/j.cois.2014.10.006

Sun, X., Wang, M. Q., and Zhang, G. (2011). Ultrastructural observations on antennal sensilla of *Cnaphalocrocis medinalis* (Lepidoptera: Pyralidae). *Microsc. Res. Techniq.* 74, 113–121. doi: 10.1002/jemt.20880

Tan, Q., Yan, X. F., Wen, J. B., and Li, Z. Y. (2012). Phylogenetic relationship of seven *Dendrolimus* (Lepidoptera: Lasiocampidae) species based on the ultrastructure of male moths' antennae and antennal sensilla. *Microsc. Res. Techniq.* 75, 1700–1710. doi: 10.1002/jemt.22118

Tanaka, K., Uda, Y., Ono, Y., Nakagawa, T., Suwa, M., Yamaoka, R., et al. (2009). Highly selective tuning of a silkworm olfactory receptor to a key mulberry leaf volatile. *Curr. Biol.* 19, 881–890. doi: 10.1016/j.cub.2009.04.035

Thompson, J. N. (1998). The evolution of diet breadth: monophagy and polyphagy in swallowtail butterflies. *J. Evol. Biol.* 11, 563–578. doi: 10.1007/s000360050106

Trifinopoulos, J., Nguyen, L. T., von Haeseler, A., and Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* 44, W232–W235. doi: 10.1093/nar/gkw256

Vieira, F. G., and Rozas, J. (2011). Comparative genomics of the odorant-binding and chemosensory protein gene families across the Arthropoda: origin and evolutionary history of the chemosensory system. *Genome Biol. Evol.* 3, 476–490. doi: 10.1093/gbe/evr033

Visser, J. H. (1988). Host-plant finding by insects: orientation, sensory input and search patterns. *J. Insect Physiol.* 34, 259–268. doi: 10.1007/s10886-010-9766-6

Walker, A. A., Robinson, S. D., Paluzzi, J. P. V., Merritt, D. J., Nixon, S. A., Schroeder, C. I., et al. (2021). Production, composition, and mode of action of the painful defensive venom produced by a limacodid caterpillar, *Doratifera vulnerans*. *Proc. Natl. Acad. Sci. U.S.A* 118:e2023815118. doi: 10.1073/pnas.2023815118

Wang, G. R., Guo, Y. Y., and Wu, K. M. (2002). Study on the ultrastructures of antennal sensilla in Helicoverpa armigera. *Agri. Sci. China* 1, 896–899.

Wang, Q., Shang, Y., Hilton, D. S., Inthavong, K., Zhang, D., and Elgar, M. A. (2018). Antennal scales improve signal detection efficiency in moths. *Proc. R. Soc. B Biol. Sci.* 285:20172832. doi: 10.1098/rspb.2017.2832

Wang, Y., Qian, H., Han, L. L., Zhao, K. J., and Jiang, X. X. (2015). Scanning electron microscopic observation on antennal sensilla of *Heliothis viriplaca*. *J. Anhui Agri. Sci.* 43, 7–11.

Wanner, K. W., and Robertson, H. M. (2008). The gustatory receptor family in the silkworm moth Bombyx mori is characterized by a large expansion of a single lineage of putative bitter receptors. *Insect Mol. Biol.* 17, 621–629. doi: 10.1111/j.1365-2583.2008.00836.x

Warren, D. L., and Seifert, S. N. (2011). Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecol. Appl.* 21, 335–342.

Weller, S. J., Jacobson, N. L., and Conner, W. E. (1999). The evolution of chemical defences and mating systems in tiger moths (Lepidoptera: Arctiidae). *Biol. J. Linn. Soc.* 68, 557–578. doi: 10.1111/j.1095-8312.1999.tb01188.x

Wu, C. S. (2010). Analysis on the host plant diversity of slug caterpillar moths in China. *Fore. Pest Dis.* 29, 25–29.

Xu, M., Guo, H., Hou, C., Wu, H., Huang, L. Q., and Wang, C. Z. (2016). Olfactory perception and behavioral effects of sex pheromone gland components in *Helicoverpa armigera* and *Helicoverpa assulta*. *Sci. Rep.* 6:22998.

Xu, W., Zhang, H., Liao, Y., and Papanicolaou, A. (2021). Characterization of sensory neuron membrane proteins (SNMPs) in cotton bollworm *Helicoverpa armigera* (Lepidoptera: Noctuidae). *Insect Sci.* 28, 769–779. doi: 10.1111/1744-7917.12816

Yang, H., Dong, J., Sun, Y., Hu, Z., Lv, Q., and Li, D. (2020). Antennal transcriptome analysis and expression profiles of putative chemosensory soluble proteins in *Histia rhodope Cramer* (Lepidoptera: Zygaenidae). *Comp. Biochem. Phys. B* 33:100654. doi: 10.1016/j.cbd.2020.100654

Yang, S., Liu, H., Zhang, J. T., Liu, J., Zheng, H., and Ren, Y. (2017). Scanning electron microscopy study of the antennal sensilla of *Monema flavescens* Walker (Lepidoptera: Limacodidae). *Neotrop. Entomol.* 46, 175–181. doi: 10.1007/s13744-016-0450-6

Yokohari, F., Tominaga, Y., and Tateda, H. (1982). Antennal hygroreceptors of the honey bee. *Apis Mellifera L. Cell. Tissue. Res.* 226, 63–73. doi: 10.1007/BF00217082

Yuvaraj, J. K., Andersson, M. N., Zhang, D. D., and Löfstedt, C. (2018). Antennal transcriptome analysis of the chemosensory gene families from Trichoptera and basal Lepidoptera. *Front. Physiol.* 9:1365. doi: 10.3389/fphys.2018.01365

Zhang, J., Bisch-Knaden, S., Fandino, R. A., Yan, S., Obiero, G. F., Grosse-Wilde, E., et al. (2019). The olfactory coreceptor IR8a governs larval feces-mediated competition avoidance in a hawkmoth. *Proc. Natl. Acad. Sci. U.S.A* 116, 21828–21833. doi: 10.1073/pnas.1913485116

Zhang, Y. V., Ni, J., and Montell, C. (2013). The molecular basis for attractive salt-taste coding in *Drosophila*. *Science* 340, 1334–1338. doi: 10.1126/science.1234133

Zwölfer, H. (1982). "Patterns and driving forces in the evolution of plant-insect systems," in *In Proceedings of the 5th International Symposium Insect-Plant Relationships*, eds J. H. Visser and A. K. Minks (Wageningen), 287–296.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Ecology and Evolution

REVIEW
published: 19 April 2022
doi: 10.3389/fevo.2022.787560

The SITE-100 Project: Site-Based Biodiversity Genomics for Species Discovery, Community Ecology, and a Global Tree-of-Life

*Xueni Bian[1,2], Beulah H. Garner[1,2], Huaxi Liu[1,2] and Alfried P. Vogler[1,2]**

[1] Department of Life Sciences, Silwood Park Campus, Imperial College London, Ascot, United Kingdom, [2] Department of Life Sciences, Natural History Museum, London, United Kingdom

Most insect communities are composed of evolutionarily diverse lineages, but detailed phylogenetic analyses of whole communities are lacking, in particular in species-rich tropical faunas. Likewise, our knowledge of the Tree-of-Life to document evolutionary diversity of organisms remains highly incomplete and especially requires the inclusion of unstudied lineages from species-rich ecosystems. Here we present the SITE-100 program, which is an attempt at building the Tree-of-Life from whole-community sampling of high-biodiversity sites around the globe. Combining the local site-based sets into a global tree produces an increasingly comprehensive estimate of organismal phylogeny, while also re-tracing evolutionary history of lineages constituting the local community. Local sets are collected in bulk in standardized passive traps and imaged with large-scale high-resolution cameras, which is followed by a parataxonomy step for the preliminary separation of morphospecies and selection of specimens for phylogenetic analysis. Selected specimens are used for individual DNA extraction and sequencing, usually to sequence mitochondrial genomes. All remaining specimens are bulk extracted and subjected to metabarcoding. Phylogenetic analysis on the mitogenomes produces a reference tree to which short barcode sequences are added in a secondary analysis using phylogenetic placement methods or backbone constrained tree searches. However, the approach may be hampered because (1) mitogenomes are limited in phylogenetic informativeness, and (2) site-based sampling may produce poor taxon coverage which causes challenges for phylogenetic inference. To mitigate these problems, we first assemble nuclear shotgun data from taxonomically chosen lineages to resolve the base of the tree, and add site-based mitogenome and DNA barcode data in three hierarchical steps. We posit that site-based sampling, though not meeting the criterion of "taxon-completeness," has great merits given preliminary studies showing representativeness and evenness of taxa sampled. We therefore argue in favor of site-based sampling as an unorthodox but logistically efficient way to construct large phylogenetic trees.

Keywords: metagenomics, Tree-of-Life, biodiversity, communities, phylogenetics

OPEN ACCESS

Edited by:
Xin Zhou,
China Agricultural University, China

Reviewed by:
Miquel A. Arnedo,
University of Barcelona, Spain
Haw Chuan Lim,
George Mason University,
United States

***Correspondence:**
Alfried P. Vogler
a.vogler@imperial.ac.uk

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *30 September 2021*
Accepted: *10 January 2022*
Published: *19 April 2022*

Citation:
Bian X, Garner BH, Liu H and Vogler AP (2022) The SITE-100 Project: Site-Based Biodiversity Genomics for Species Discovery, Community Ecology, and a Global Tree-of-Life. Front. Ecol. Evol. 10:787560. doi: 10.3389/fevo.2022.787560

INTRODUCTION

A solid taxonomy of species on Earth is the basis of all biology. Community ecology depends on detailed knowledge of species in the assemblage and their traits that determine each species' functional role, its interactions with other species, and its ecological and geographic distributions. These species traits are acquired in a line of descent from a common ancestor, and knowing the phylogenetic placement of a species can inform us of its traits (Wiley and Lieberman, 2011). The power of phylogenetics to test hypotheses of ecological trait evolution within communities has long been recognized (Losos, 1996). In studies of community ecology, information on trait evolution is used in a variety of ways, e.g., for distinguishing between scenarios of environmental filtering and competition inferred from phylogenetic clustering or overdispersion, respectively, and the processes driving community assembly generally (Webb et al., 2002; Choo et al., 2017). This paper considers the utility of phylogenetics in community ecology from the perspective of the huge unknown species diversity encountered in many groups of invertebrates. The core focus of our approach aims at sampling understudied and yet to be described species from multiple sites worldwide and adding them to the phylogenetic tree. With each additional sample, the phylogenetic tree becomes more complete in representing the global diversity, while in turn the growing global tree informs on the phylogenetic composition of each local community.

A comprehensive catalog of all or most living species and their phylogenetic relationships appeared largely unattainable with conventional taxonomic methods, but potentially is made possible with the tools of genomics and fast algorithms for phylogenetic tree construction (Hinchliff et al., 2015; Lewin et al., 2018). However, this vision of a complete Tree-of-Life has yet to become a realistic prospect on account of the multiple constraints to taxonomy over the past two centuries. As these constraints shifted over time, they highlight the limitation to our current knowledge and the possibility that the diversity of life on Earth will never be fully known, as the extinction crisis leaves little available time (Dirzo and Raven, 2003). Any approach aimed at speeding up the taxonomic process therefore needs to maximize its contribution in the face of methodological limitations and constraints to resources. Of particular concern are recent restrictions to access to biological samples in a complex regulatory framework for specimen acquisition (Laird et al., 2020). This begs the question about the most efficient manner for obtaining a phylogenetic tree that represents the highest number of species possible. We propose that a community ecology approach could greatly contribute to the taxonomic endeavor while at the same time producing key insights into the forces that shape the assembly and maintenance of the communities themselves. As we take on the task of producing ever larger phylogenetic trees as more communities are added, we also need to understand the potential pitfalls of phylogenetic-tree building that come with a community centered approach.

Over the past two centuries, continuing efforts of species description and phylogenetic analyses have produced a classification that captures the state of knowledge about the huge diversity of living and extinct species and their evolutionary history (Grandcolas and Pellens, 2016). Most conventional approaches of filling the gaps in taxonomic knowledge target particular focal groups or species that are described with reference to close relatives, either already known or also newly described, and only once the group is sufficiently well known is the taxonomic information made available for use in ecological or biodiversity studies. This monographic process of taxonomy can take decades from the time of the initial discovery of a specimen (the so-called "species shelf-life"), and thus such *lineage-based* approaches are not efficient for the use of phylogenetics in community ecology (Fontaine et al., 2012). They also do not make efficient use of recent methods of DNA sequencing and large-scale digital imaging, which allows processing of numerous species in large batches irrespective of the membership in a particular lineage. Various approaches to automated sorting and imaging of morphospecies for downstream extractions and sequencing have already been developed (e.g., Folk et al., 2021; Srivathsan et al., 2021). We here propose the SITE-100 initiative as an approach specifically designed to deal with the dire need for higher-throughput methods in taxonomy, to both assess species diversity and trace its origins at community level. SITE-100 takes a *site-based* approach to sampling the Tree-of-Life, by collecting extensively at accessible high-biodiversity localities and processing all or most specimens encountered with standard imaging and sequencing methods (Arribas et al., 2021), to be incorporated in the global phylogenetic tree. The initial ambition of this project is to obtain local samples for 100 sites around the globe, to represent the major biogeographic realms and ecoregions, with focus on forest biomes (Olson et al., 2001). With a site-based approach, communities are captured—as they are ecologically encountered, therefore keeping communities intact—in one sampling event. Once incorporated in a phylogenetic tree, the community data sets provide the basis for addressing questions about local ecological interactions, habitat associations, environmental filters, and others, while at the global level they contribute to address macroevolutionary and macroecological questions about lineage evolution, rates of speciation, historical biogeography, and global species distribution patterns.

Traditionally, alpha taxonomy, being the delimitation and description of species, and phylogenetics, being the inference of relationships, combine into the field of systematics, which enables our classification and understanding of biodiversity. With the wide use of DNA sequencing, these approaches are no longer separate, as population-level sequence data designed to determine the species boundaries or population structure can equally be used to link these species into a wider database and phylogenetic tree (Bocak et al., 2016). Community sampling has immense potential to contribute to phylogenetic inference, as it adds the knowledge of species diversity, while the phylogenetic inference is improved with the increasing taxon density and growing amount of DNA data per species (Tautz et al., 2003; Vogler and Monaghan, 2007). It is worth noting that mapping out the historical processes of evolution through phylogenetics vs. deciphering current compositional aspects of community species diversity are no longer separate endeavors but present

themselves as a singular challenge. All aspects of defining and placing a species within an evolutionary framework (systematics) is continually refined by the addition of new data (Lipscomb et al., 2003). To the degree that a tree depicting the relationships of extant species can reflect the true evolutionary history, this growing database and use of sophisticated phylogenetic inference methods would eventually approach an evolutionary tree that includes most of the extant biodiversity at the species level. There has yet to be an organized endeavor that aims to (i) sample the globe at multiple sites and (ii) co-assess both what is present in the local fauna (community scale), and what are the relationships of the local faunas with one another (biogeographic scale). The challenge of bringing together these aspects of both alpha and beta diversity to iteratively begin sampling the globe, and gain a true sense of the diversity within biomes, can be met with the application of high throughput phylogenetics.

Existing databases, even in large collaborative projects are far from achieving a comprehensive tree (Rees and Cranston, 2017). First and foremost, the greatest complication arises from the sheer magnitude of unknown species (Mora et al., 2011). Even if we focus only on Metazoa (multicellular eukaryotes), possibly 90% of species remain unknown. Mora et al. (2011) predicts global eukaryotic diversity to be in the region of 8.7 million (± 1.3 million SE) species, of which ∼2.2 million (± 0.18 million SE) are marine. However, only a fraction of this number has been named and incorporated in the Tree-of-Life. For example, 2.3 million names are present in the Open Tree-of-Life, currently the most comprehensive database of this kind, but the great majority of nodes remain unresolved and only correspond to the Linnaean taxonomy (Hinchliff et al., 2015). In many lineages such as protozoans the proportion of unknown diversity may be a lot higher (Larsen et al., 2017). In addition, these species counts are generally based on Linnaean names and morphological species delimitations that are not easily linked to a DNA-based taxonomic system on which phylogenetic inference is predominantly based (Zamani et al., 2021). Extensive sequencing efforts at the species level are underway, generating inventories of standardized short "DNA barcodes" for each species (Hebert et al., 2003). However, this leads us to the second issue of the type and number of genes available: while barcodes are readily generated, they have limited phylogenetic power (Wiemers and Fiedler, 2007; Quicke et al., 2012; DeSalle and Goldstein, 2019). Resolving deeper levels requires more genes, i.e., genomics, but data collection and analysis are not practical for all species. Compounding this are challenges created by gene flow and horizontal gene transfer that complicate the inference of the species tree from a limited set of markers. Overall, most ambitious genome sequencing projects that ultimately aim for genome sequences for all species, namely the i5K (Robinson et al., 2011), Darwin Tree of Life[1], and the EarthBiogenome, remain in their infancy, although in insects the availability of genome sequences is now increasing exponentially (Feron and Waterhouse, 2021).

The SITE-100 initiative implements a hybrid approach, where well-identified specimens are used to bridge classical morpho-taxonomy and phylogenetic inference by subjecting them to a combined phylogenetic analysis of taxon-poor genomic data and link them to taxon-dense metabarcodes (**Figure 1**). This hybrid approach that integrates few taxa represented by many genes, and many taxa represented by few genes, is probably the only realistic prospect for building the species-level Tree-of-Life in the near future (Chesters, 2017). Disparate activities over the past circa 300 years reporting on Earth's diversity have been lacking until the advent of phylogenetics, which enhanced the taxonomic informativeness that goes beyond a catalog of species and individual observations, and now provides a synthesis of lineage and trait evolution. Vice versa, for the Tree-of-Life to be fully understood ecological factors must be investigated. The phylo-ecology of community (site) based studies presented here provides an increasingly necessary adjunct to the gargantuan effort of taxonomic lineage-based studies. Combining these approaches, we can then move toward a methodology embedded within the discipline of systematics, within the modern-day constraints to accessing the "completeness" of all living species on Earth.

THE SITE-100 METHODOLOGY

We first discuss the SITE-100 approach in regard to the protocols for data generation and processing. The approach borrows from the idea of "genomic observatories," i.e., sites that are part of a global network for, ideally, long-term surveillance using genomic methods (Davies et al., 2014). Even if limited to short periods of intense collecting, the SITE-100 approach aspires to employ standardized field methodology repeatable across biomes. A georeferenced one-hectare plot populated with collecting traps enables consistency of data collection across biomes and habitats, with an initial focus on known biodiversity hotspots in tropical forest sites. Specimen selection for phylogenetic analysis relies on an alpha-taxonomic overview for morphospecies capture, particularly with regard to complex community assemblages where putative species are many and cryptic. High throughput sequencing (HTS) provides data for the different hierarchical levels at which the analysis of diversity and turnover is conducted: (1) Genome sequencing for resolving basal relationships; (2) large-scale mitogenome sequencing for the phylogenetics of local communities; (3) (meta) barcoding for the sequencing of all species (or clusters of sequence variation, referred to as Operational Taxonomic Units, OTUs) for studies of species diversity; and (4) ultimately providing resolution of haplotypes (Amplicon Sequence Variants, ASVs) for phylogeographic and population genetics studies (**Figure 1**).

The sampling and sequencing strategy of the SITE-100 protocol coincides with this hierarchical structure of the data (**Figure 2**). In the field, we make use of a suite of sampling methods including passive traps (flight interception, pitfall, malaise, light traps), sweeping, and canopy-fogging. Sites may contain sub-sites to cover as many kinds of habitat as possible within a given area. These methods are replicated through time, e.g., 1 month collecting period using a particular trap type for 3-day intervals, providing a modular design that helps with comparability across sites

[1]https://www.darwintreeoflife.org

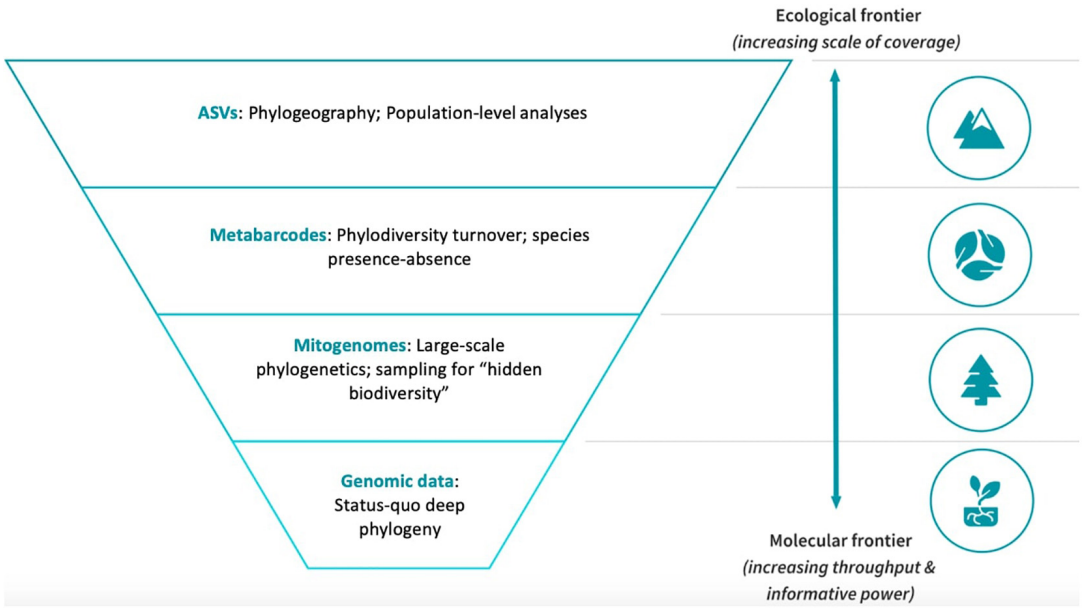


FIGURE 1 | The SITE-100 initiative implements a hybrid approach with different levels of phylogenetic sequencing depth and taxon density. Full genome data are used to represent a limited set of major lineages chosen on taxonomic grounds, while mitogenomes and metabarcodes represent increasing levels of sampling density but with lower sequencing amount. The top level represents intraspecific variation that is assessed using filtered unique ASVs from multiple metabarcoding libraries.

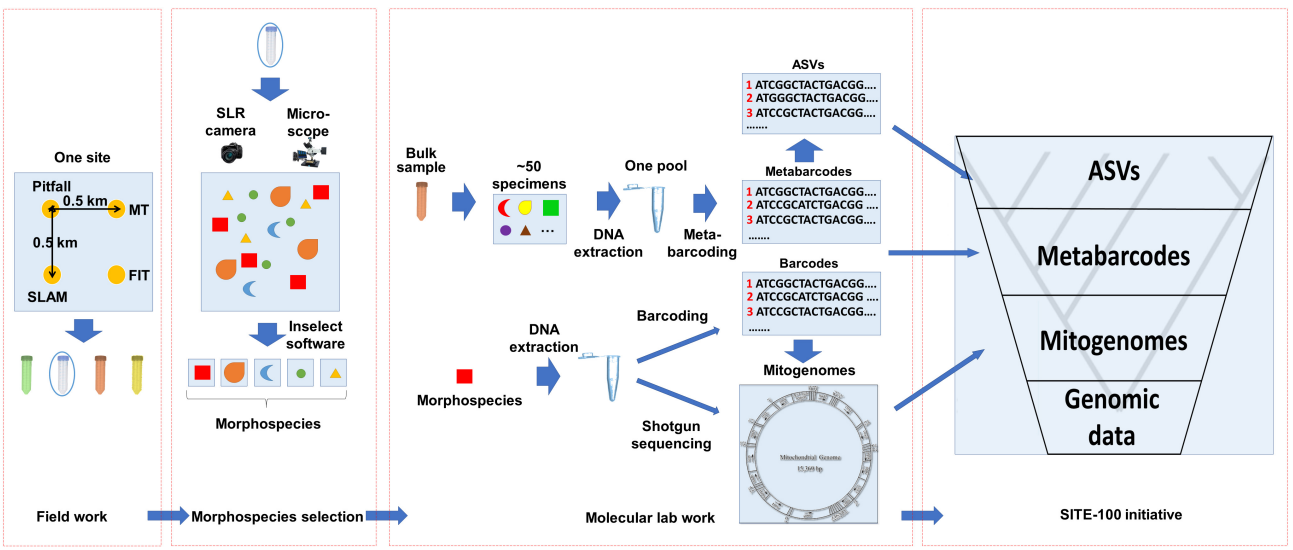


FIGURE 2 | General workflow of the SITE-100 initiative. (1) Field work: Sampling from one site using various standardized collection methods: pitfall trap, Malaise trap (MT), Flight interception trap (FIT) and sea, land, air, and malaise trap (SLAM) to sample from the various strata of a given habitat. (2) Morphospecies selection: Specimens selected in batches of 50 of similar size and digitally imaged with high-resolution imagery. The resulting composite image is separated to morphospecies by employing open-source software Inselect (Hudson et al., 2015). (3) Molecular lab work: A hybrid approach of *cox1* metabarcoding of bulk samples and mitochondrial metagenomics of individual specimens representing unique morphospecies. For the latter, *cox1* barcodes are produced in parallel as baits for recovery of mitogenomes from mixed sequencing libraries. (4) SITE-100 database and analyses: Phylogeny reconstruction of hierarchical datasets and metabarcode placement for Tree-of-Life construction and biodiversity research.

and trapping methods (Arribas et al., 2021). Once gathered in the field, the pool of specimens is imaged using large-scale, high resolution photography (**Appendix 1**). In a further step these images are used to select representatives of all recognizable morphospecies or of key specimens of interest to phylogenetics spanning the assumed phylogenetic diversity of

a site. This selection of specimens conducted in "real-time" requires the expertise of parataxonomists able to recognize higher-level taxa and to separate morphospecies. However, as DNA extractions are generally non-destructive the detailed evaluation of specimens can be conducted by taxonomic specialists at a later stage.

At the core of the sequencing strategy is the community-wide analysis and assembly of genomic information of mixed specimens using either metagenomics (PCR-free) or metabarcoding (PCR-based). Phylogenetically informative markers may be obtained by genome skimming, i.e., the low coverage shotgun sequencing of pooled samples and assembly of high-copy markers using standard genome assemblers, which in insects provides the reliable acquisition of mitogenomes for all specimens in the mixture (Zhou et al., 2013; Papadopoulou et al., 2015; Crampton-Platt et al., 2016). Genome skimming can add to the number of available mitogenomes rapidly at a sequencing depth of about 10 species per Gb of DNA shotgun data. This step can be conducted on the unsorted specimen mixture from the trap sample (the "insect soup"; Ji et al., 2013) or after the presorting of individual target specimens (Crampton-Platt et al., 2016). Prior to DNA extraction, bulk specimens are imaged with large-scale, high resolution imaging systems, such as the Zeiss AXIO Zoom, featuring a motorized focus drive and motorized stage that enable the field to be divided into regular tile-images which are subsequently *xyz* stitched. Individual specimen images are cropped from the composite photograph, e.g., using the Inselect software (Hudson et al., 2015), and uploaded to public image databases for storage and downstream taxonomic identification. This image database also provides the for selection of specimens for individual DNA extraction for shotgun sequencing in the genome skimming step generating the mitogenomes (or potentially generating nuclear ortholog sets if sequenced more deeply).

In the currently used protocols, mitochondrial genomes are obtained by shotgun sequencing on the Illumina platform and bioinformatically separated, usually mixing the DNAs from ∼200 specimens to reduce costs. This approach routinely generates 50–80% of complete or nearly complete mitogenome assembly (e.g., Breeschoten et al., 2016; Choo et al., 2017), with limited risk of chimera formation (Gómez-Rodríguez et al., 2017). The resulting mitogenome assemblies are then assigned to a particular specimen by a DNA barcoding step carried out in parallel. The DNA extract from each specimen in the library is used for a separate amplification and sequencing of a fragment of *cox1*, and the most abundant read extracted with the NAPtime pipeline (Creedy et al., 2019) is used as bait to match a particular mitogenome contigs. However, as sequencing library costs are becoming cheaper, separate genome skims of each individual may be preferable over the sequencing of a specimen mixture. All remaining specimens not selected in this step are bulk-processed and subjected to metabarcoding using the *cox1* gene. Of the total pool of *cox1* sampled, we cluster sequence reads into OTUs to obtain entities equivalent to the species level using VSEARCH (Rognes et al., 2016). Alternatively, stringent filtering can produce the presumed true haplotypes (ASVs). Phylogeny reconstruction by hierarchical datasets then follows suit, with metabarcode placement on the tips of the phylogeny. For specific details please refer to **Figure 2** and **Appendix 1**.

BUILDING THE TREE-OF-LIFE USING SITE-BASED METAGENOMIC SEQUENCING

Each "community," i.e., the specimens encountered at the 1-ha sites potentially including many hundreds of species, is a largely fortuitous selection of species and deeper lineages. They can be expected to include taxa previously unknown and thus fill the gaps in global clade coverage, obviating traditional approaches specifically seeking to "complete" the sampling of a target lineage. Metabarcoding and metagenomics will extend taxonomic research in particular to the "hidden" biodiversity of small-bodied and poorly known groups whilst populating the tree at the deeper nodes. With the resulting phylogenetic tree, the evolutionary history of the community can be known. For example, a community may be characterized as the inhabitants of a single tree of any given biome, or the population of a geologically young island. Taking into account the growing numbers of samples, DNA data from local sites are combined for an ever more complete sampling of the global Tree-of-Life. Placement of the members of each community on the global tree becomes instrumental to understanding the evolutionary history of local communities, including the biogeographic context that is obtained automatically from the site information (i.e., data retain information of the sites from which they are sampled).

Sequencing and sampling methods of SITE-100 thus deviate from standard practice in phylogenetics, which aims at a set of taxonomic exemplars deemed adequate representatives of a focal group (McKenna et al., 2019). On the other hand the phylogenetics of site-based selection regimes remains in its infancy. By linking two disciplines—phylogenetics and community ecology—it holds great potential for them to converge, provided there is general awareness of possible implications of this type of taxon choice on inferring phylogenetic relationships.

THE IMPACT OF SITE-BASED SAMPLING ON TREE INFERENCE: WHAT IS THE PHYLOGENETIC PROBLEM?

The potential problems of site-based phylogenetics revolve around the incomplete sampling of a local site and the limited representation of lineages to inform the topological reconstruction, which exacerbate long-branch attraction (LBA) (Bergsten, 2005). Community sampling intends to capture the phylogenetic diversity within a region. Essentially, it relies on the assumption that many small site-based samples are adequate to capture the phylogenetic diversity of large geographic regions (**Figure 3**). However, single-sites sampling may skew the taxon selection because: (1) Single sites contain species that are phylogenetically clumped relative to the global tree because of their conserved biogeographic distributions

(the locality is phylogenetically correlated) or the common ancestry of ecological traits of co-occurring species (traits are phylogenetically conserved). (2) Single sites contain lineages that are phylogenetically isolated given the incomplete representation of the global tree at any one site. Thus, site-based studies would be expected to suffer from long branches in phylogenetically isolated taxa, which may confound the tree construction in particular when using DNA data with high levels of homoplasy. However, as site-based sampling proceeds, increasing numbers of communities are fed into the same global dataset and phylogenetic analysis so that branches are more densely sampled. Between multiple sites, the following can be expected based on preliminary empirical observations: (1) It takes very few sites globally for most deep lineages to be represented. For example, using communities of Coleoptera (beetles) a few sites in the Neotropics and a single site in Borneo already recovered most clades known from a global taxonomic effort (most families and even subfamilies are represented), while the Neotropical sites combined define the depth of clades unique to this biogeographic region (unpublished). (2) Under the assumption that a more uniform sampling of the tree along the root-to-tip axis can improve the estimates of character variation and avoid long branches (Bergsten, 2005), sampling sites need to be chosen to represent the taxonomically most distinct sets, e.g., by selecting biogeographically or ecologically distant sites. Thus the taxonomic process is then starting at a state that is comparable to taxon selection, that is, where phylogenetic knowledge is directed toward targeted lineages, except here the sampling is for the sites most valuable for taxonomic gap filling. This can be complemented with a final taxon-based selection for isolated lineages that cannot be obtained by bulk sampling.

The challenge of "incomplete" sampling for phylogenetics is actually a challenge of "uneven" sampling. Completeness is an impossible goal. Trees are fundamentally incomplete catalogs of (mostly) extant taxa as well as an incomplete record of branching events that have supposedly led to the taxa represented. A multitude of events permanently mar the shape or "completeness" of a tree, namely extinction dynamics and ecological coverage (Warnock et al., 2020). On the other hand, "evenness" in sampling can be tested and can be intuited as sampling that is randomly spaced out across the root-to-tip axis of the tree. Currently we know too little about the geographic distribution of phylogenetic lineages in most groups of insects to assess the error resulting from the unorthodox sampling. However, in at least one case it has been observed that site-based sampling of dung beetles (Scarabaeinae) from the Neotropical and Oriental regions do indeed capture a set similar to a random sub-selection of global scarabaeine lineage diversity available on GenBank, indicating that even minimal site-based sampling provides high lineage representation (Tansley, 2020).

The benefits of increasing the number of lineages (as opposed to the number of genes) has perhaps been dwarfed by colossal advances in sequencing throughput (**Figure 4**). Various studies conducted at around the turn of the last century have addressed the question of whether increased sampling in either sequence length or taxon count improves

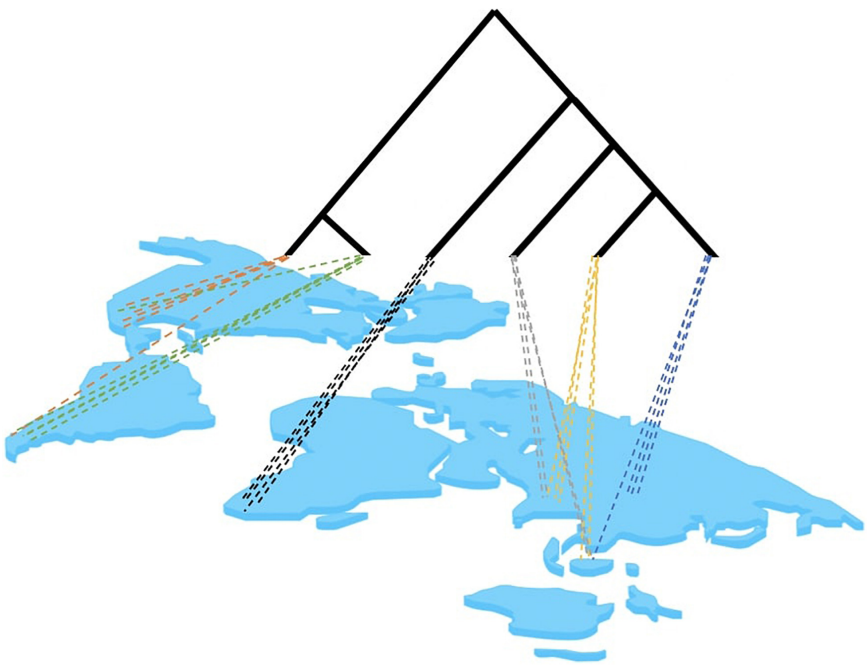


FIGURE 3 | Hypothetical distribution of lineages in a multi-site analysis. Site-based sampling may only capture a certain proportion of the full tree, depending on the level of geographic structure of lineages and the geographic proximity of sampling sites. Colors represent closely related lineages. Note that particular lineages at any hierarchical level may be captured across multiple sites. This overlap can be expected to increase with closer biogeographic proximity of sampling sites and with deeper origin of lineages near the base of the tree.

phylogenetic accuracy, using either simulation studies where the "true tree" is known (Graybeal, 1998; Pollock et al., 2002), or by subsampling increasingly larger subsets from a large starter set (Rosenberg and Kumar, 2001). These studies generally suggest that for a given total data matrix size, phylogenetic accuracy improved with the increase in taxa, and this increase was more rapid than with the same amount of data added to the sequence length per taxon. There are caveats to the conclusion about the benefits of adding more terminals: a certain minimum of sequence length for the added taxa is required to provide phylogenetic power, and accuracy did not improve as much if taxa are added near the tips rather than the base of the trees (Graybeal, 1998).

To sum up the phylogenetic defensibility of the SITE-100 protocol: (1) The sampling design requires a clear image of the effect of sampling density to ensure between-site relief of LBA-related issues; (2) completeness is a catchall term where the relevance of "evenness" to avoidance of LBA is understated; and (3) the effects of increasing the number of terminals in a phylogenetic analysis is unpredictable and case-specific to the total amount of sequence data involved. It is inevitable that all studies are caught in a tradeoff between detail and scale (Barraclough and Nee, 2001). We take this chance to introduce site-based sampling as a haphazard but logistically efficient and long-game strategy to document the biodiversity of poorly known lineages on Earth.

HIERARCHICAL DATASETS

The product of combined nuclear, mitochondrial and metabarcode sequencing is a highly skewed matrix composed of a comparatively small number of nuclear genomes, an intermediate number of full mitogenomes, and a very large number of short metabarcode sequences (see Chesters, 2017). Thus, beyond taxon density, by sequencing a subset of individuals for whole or partial genomes, the database is effectively expanded along other axes of information that (1) contribute a greater number of characters for increased support and (2) are minimally affected by misleading phylogenetic signal. We distinguish four hierarchical levels that

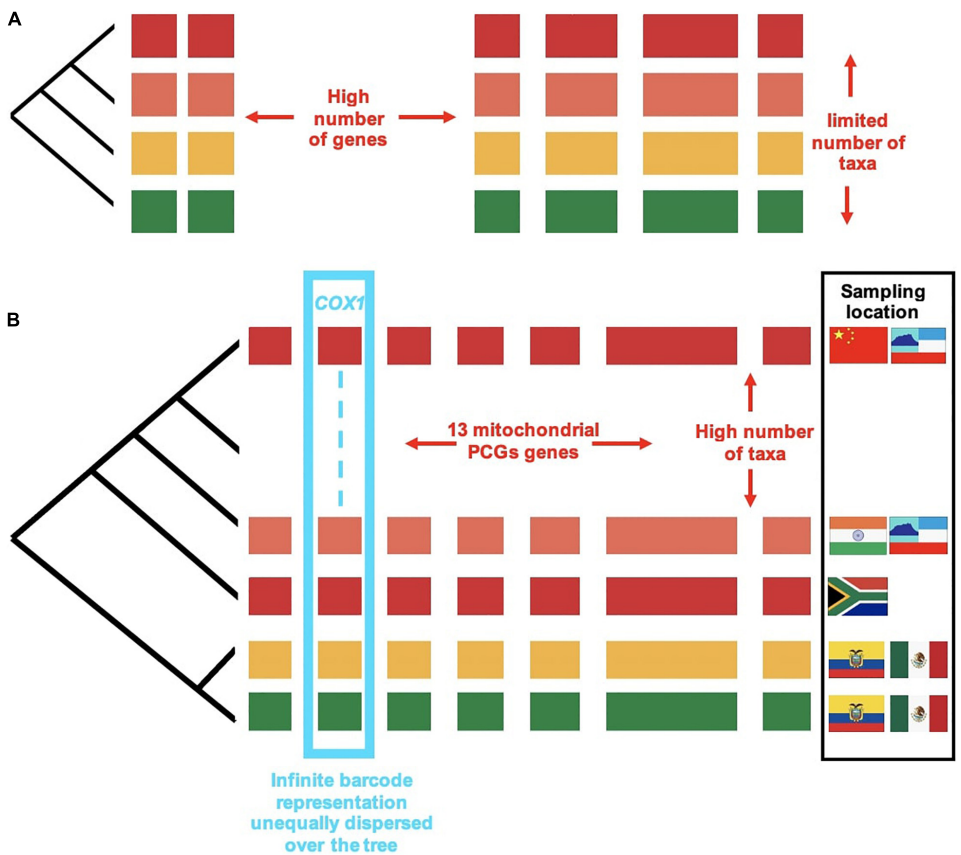


FIGURE 4 | The trade-off between number of lineages and number of genes: **(A)** A set of taxa chosen for taxonomic purposes and sequenced for a large number of genes, e.g., nuclear orthologs from whole-genome or transcriptome sequencing. **(B)** The addition of site-based data for full mitogenomes resulting in 13 protein coding genes (PCGs) obtained by genome skimming, and the much larger number of taxa sequenced for the *cox1* gene from metabarcoding. Colored blocks represent the individual genes sampled. Different colors represent the lineages sampled. Designing a sampling regime for phylogenetics is necessarily a balancing act between number of genes (blocks) and number of lineages (colors). The availability of *cox1* metabarcodes represented in discontinuous lines (cyan) dwarfs all other data types.

differ in scope of biological enquiry and depth of sequencing effort (**Figure 4**), commensurate with the four hierarchical levels of biological inquiry described in **Figure 1**. In this section we elaborate on this methodology in order to achieve the increasing scale of taxonomic coverage at each level.

First, nuclear genome data for taxonomically chosen key entries provide a scaffold for the "status-quo" (lineage-based) approach of inferring deep relationships among major groups. Nuclear datasets can be obtained via raw genome and transcriptome sequences such as the Genbank SRA database. A set of 2,000–4,000 universal orthologs can be extracted readily using the Benchmarking Universal Single-Copy Orthologs (BUSCO) (Simão et al., 2015; Waterhouse et al., 2018) pipeline to search for orthologs against an appropriate reference dataset, e.g., for Endopterygota (Waterhouse et al., 2018). Phylogenetic analyses require alignment of each ortholog, followed by concatenation and tree searches under partitioned maximum likelihood models or using multispecies coalescent models on the individual gene trees to address the effects of incomplete lineage sorting (ILS) (Zhang et al., 2018). These analyses can be performed under different missing-data ratios, e.g., creating matrices of 50, 75, and 90% completeness. LBA in these datasets can be partly ameliorated by conducting phylogenetic analyses at the amino acid level and using models incorporating differences in site frequencies (Wang et al., 2018). Tree searches at these scales are computationally intensive; for example, in our hands a matrix of 120 terminals of Coleoptera and 530,000 amino positions took nearly 5,000 CPU hours and this time increases quickly with greater taxon number (Ding, Y., unpublished).

Secondly, mitochondrial genomes greatly increase taxon coverage. Mitogenomes are small and compact in genomic architecture, with short intergenic regions, as does befit an autonomously replicating entity with replicatory signals under strict nuclear control. Unlike in the case of many nuclear genes, orthology is unequivocal for the 13 mitochondrial protein coding and 2 rRNA genes. Due to higher rates of variation compared to most nuclear protein coding regions, the mitochondrial genome lends itself well to systematic studies at intermediate levels of taxonomic classifications, addressing various taxonomic puzzles previously only considered by morphology alone (Rubinoff and Holland, 2005). Mitochondrial genomes lack recombination (all genes have the same history), which eliminates a potential source of character incongruence, but unfortunately also makes it impossible to recognize the effects of ILS when sequencing this marker alone. In addition, character evolution in mitogenomes is complex, and thus mis-specifications of the model and incorrect tree inference may be common. Tree inferences may be confounded by (1) heterogeneity in rate of nucleotide change, (2) heterogeneity in nucleotide composition among the terminals, and (3) multiple superimposed character changes due to elevated mutation rates (Song et al., 2016; Timmermans et al., 2016), which can only partially be ameliorated by parameter-rich (i.e., high-complexity) models of molecular evolution, e.g., using site-heterogeneous mixture models as those implemented in PhyloBayes (Lartillot and Philippe, 2004). Whole mitogenomes partly overcome specific idiosyncrasies of molecular evolution affecting each individual gene, in particular the *cox1* barcoding marker showing unique features of variation, which confounds the trees (Pons et al., 2010). However, across the insect phylogeny and even at the order and family levels, heterogeneity of rates and composition lead to biases that affect the propensity for inferring deep relationships. To overcome these issues, mitogenome trees may be constructed with the nuclear tree as a backbone to resolve deep relationships. Combined with these nuclear data, in the balancing act between densely sampling for shorter sequences and frugally curating a select catalog of long sequences, the mitogenome datasets sits at the Goldilock zone as a core dataset in the effort to "complete" the Tree-of-Life by sequencing community samples from high-biodiversity sites.

Finally, we use barcodes and metabarcodes for representation at the tip of a stabilized tree, rather than for phylogenetic information (Min and Hickey, 2007). Barcodes and metabarcodes present a limited number of characters, which in addition are highly homoplastic. Foisting these short sequences over a mitogenomic tree is very different from how metabarcodes are normally used: for rapid phylogenetic placement of a sequence based on similarity searches against a reference database (usually GenBank or BOLD) using the Blast algorithm or *k-mer* based methods (Huson et al., 2007; Linard et al., 2019). It remains debatable if short metabarcodes alone are sufficient to determine phylogenetic position, in particular if trees are very big and numbers of taxa greatly exceed the number of characters (nucleotide positions), but longer full mitogenome sequences that match these metabarcodes could further validate phylogenetic placements. The generation of ASVs (i.e., we filter for genuine mitochondrial haplotypes) and the recovery of multiple individuals of a given species, potentially from multiple sites, allow us to ask further questions to do with phylogeography and population-level genetics.

Via the methodology outlined above, we harness both the phylogenetic informativeness of nuclear genome data and the easy accessibility of mitogenomes and metabarcodes. Topologies from the previous level constrain and provide scaffold for tree-building the next level down, allowing for the placement of unidentified, short sequences into the well-established phylogenetic tree (Chesters, 2017). The combined evidence of a hierarchical data set provides stability and phylogenetic power for tree-construction, given that the process of tree-construction accounts for the aforementioned problems relevant to each type of gene (for example the problem of ILS in nuclear genes, and heterogeneity in composition and high mutation rates for mitogenomes). This hierarchical approach scales up phylogenetic power and moves us a step closer to the approximation of the insect Tree-of-Life complete at the species level.

CONCLUSION

The previously separate endeavors of deciphering historical and present-day biodiversity patterns, broadly defined by macroevolution and macroecology, respectively, now present themselves as a unified challenge, enabled by large-scale DNA sequencing focused on a limited number of sites. The phylogenetic framework of biodiversity envisioned here is

designed for the study of any poorly known, highly diverse group currently lacking in taxonomic representation. The effort of gathering the Earth's biodiversity, community by community, has immense potential to contribute to phylogenetic inference, but equally this methodology endeavors to organize the sampling of the terrestrial biotas to establish the distribution of the world's species, as part of efforts to match the biodiversity extinction crisis. Standardized repeatable field protocols give the structure for deep sampling across habitat strata, from soil to canopy. With recurring sampling and rapid metabarcoding to determine species presence/absence (and possibly abundance) at a site over time, the methodology allows for the truest sense of community dynamics at the clade level to emerge, which can then be applied to multiple sites worldwide and inform the Tree-of-Life in ways that lineage-based methodologies cannot.

The hybrid approach makes use of increased availability of genome data and mitogenomes, which solidify the base and middle portion of the tree, respectively, and the inclusion of metabarcodes will create minimal error as basal relationships are fixed. By sampling insect communities with this method, we can answer questions about evolution, ecological dynamics, biogeography, and others, on an unprecedented scale. This will further the close integration of community ecology and phylogenetics, in particular to understand the role of trait and niche diversity for community assembly and responses to environmental change (Choo et al., 2017; Merckx et al., 2015). There is a need to clearly integrate these efforts with the SITE-100 protocol, its logistical merits, the many ways in which it is unorthodox yet long-game, as well as the phylogenetic problems from which it is and isn't exempt.

AUTHOR CONTRIBUTIONS

XB, BG, HL, and AV conceived the study and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

Funding was provided by the NHM Biodiversity Initiative and the NERC ARBOLES grant.

ACKNOWLEDGMENTS

We thank Yin-huan Ding (Nanjing Agricultural University), Michael Tansley (Imperial College London), and Rui-e Nie (Chinese Academy of Sciences) for access to preliminary studies discussed in this text and to members of the Vogler lab (2020–2021) for participation in relevant discussions. We are grateful to our reviewers for helpful comments and improvements to the text.

REFERENCES

Arribas, P., Andújar, C., Bidartondo, M. I., Bohmann, K., Coissac, É., Creer, S., et al. (2021). Connecting high-throughput biodiversity inventories: opportunities for a site-based genomic framework for global integration and synthesis. *Mol. Ecol.* 30, 1120–1135. doi: 10.1111/mec.15797

Barraclough, T. G., and Nee, S. (2001). Phylogenetics and speciation. *Trends Ecol. Evol.* 16, 391–399. doi: 10.1016/s0169-5347(01)02161-9

Bergsten, J. (2005). A review of long-branch attraction. *Cladistics* 21, 163–193. doi: 10.1111/j.1096-0031.2005.00059.x

Bocak, L., Kundrata, R., Fernández, C. A., and Vogler, A. P. (2016). The discovery of Iberobaeniidae (Coleoptera: Elateroidea): a new family of beetles from Spain, with immatures detected by environmental DNA sequencing. *Proc. R. Soc.* 283:20152350. doi: 10.1098/rspb.2015.2350

Breeschoten, T., Doorenweerd, C., Tarasov, S., and Vogler, A. P. (2016). Phylogenetics and biogeography of the dung beetle genus *Onthophagus* inferred from mitochondrial genomes. *Mol. Phylogen. Evol.* 105, 86–95. doi: 10.1016/j.ympev.2016.08.016

Cavender-Bares, J., Kozak, K. H., Fine, P. V. A., and Kembel, S. W. (2009). The merging of community ecology and phylogenetic biology. *Ecol. Lett.* 12, 693–715. doi: 10.1111/j.1461-0248.2009.01314.x

Chesters, D. (2017). Construction of a species-level Tree of Life for the insects and utility in taxonomic profiling. *Syst. Biol.* 66, 426–439. doi: 10.1093/sysbio/syw099

Choo, L. Q., Crampton-Platt, A., and Vogler, A. P. (2017). Shotgun mitogenomics across body size classes in a local assemblage of tropical Diptera: Phylogeny, species diversity and mitochondrial abundance spectrum. *Mol. Ecol.* 26, 5086–5098. doi: doi:10.1111/mec.14258

Crampton-Platt, A., Yu, D. W., Zhou, X., and Vogler, A. P. (2016). Mitochondrial metagenomics: letting the genes out of the bottle. *Gigascience* 5:15. doi: 10.1186/s13742-016-0120-y

Creedy, T. J., Andújar, C., Meramveliotakis, E., Noguerales, V., Overcast, I., Papadopoulou, A., et al. (2022). Coming of age for COI metabarcoding of whole organism community DNA: towards bioinformatic harmonisation. *Mol. Ecol. Resour.* 22, 847–861. doi: 10.1111/1755-0998.13502

Creedy, T. J., Norman, H., Tang, C. Q., Chin, K. Q., Andujar, C., Arribas, P., et al. (2019). A validated workflow for rapid taxonomic assignment and monitoring of a national fauna of bees (Apiformes) using high throughput DNA barcoding. *Mol. Ecol. Resour.* 20, 40–53. doi: 10.1111/1755-0998.13056

Davies, N., Field, D., Amaral-Zettler, L., Clark, M. S., Deck, J., Drummond, A., et al. (2014). The founding charter of the Genomic Observatories Network. *Gigascience* 3:2. doi: 10.1186/2047-217X-3-2

DeSalle, R., and Goldstein, P. Z. (2019). Review and interpretation of trends in DNA barcoding. *Front. Ecol. Evol.* 7:302. doi: 10.3389/fevo.2019.00302

Dirzo, R., and Raven, P. H. (2003). Global state of biodiversity and loss. *Annu. Rev. Env. Resour.* 28, 137–167. doi: 10.1146/annurev.energy.28.050302.105532

Feron, R., and Waterhouse, R. M. (2021). Assessing species coverage and assembly quality of rapidly accumulating sequenced genomes. *bioRxiv* [Preprint]. doi: 10.1101/2021.10.15.464561

Folk, R. A., Kates, H. R., LaFrance, R., Soltis, D. E., Soltis, P. S., and Guralnick, R. P. (2021). High-throughput methods for efficiently building massive phylogenies from natural history collections. *Appl. Plant. Sci.* 9:e11410. doi: 10.1002/aps3.11410

Fontaine, B., Perrard, A., and Bouchet, P. (2012). Twenty-one years of shelf life between discovery and description of new species. *Curr. Biol.* 22, 943–944. doi: 10.1016/j.cub.2012.10.029

Gómez-Rodríguez, C., Crampton-Platt, A., Timmermans, M. J. T. N., Baselga, A., and Vogler, A. P. (2017). Intraspecific genetic variation in complex assemblages from mitochondrial metagenomics: comparison with DNA barcodes. *Methods Ecol. Evol.* 8, 248–256. doi: 10.1111/2041-210x.12667

Grandcolas, P., and Pellens, R. (2016). *Biodiversity Conservation and Phylogenetic Systematics: Preserving Our Evolutionary Heritage in An Extinction Crisis*, 1st Edn. New York, NY: Springer. doi: 10.1007/978-3-319-22461-9

Graybeal, A. (1998). Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst. Biol.* 47, 9–17. doi: 10.1080/106351598260996

Hebert, P. D., Cywinska, A., Ball, S. L., and deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proc. Biol. Sci.* 270, 313–321. doi: 10.1098/rspb.2002.2218

Hinchliff, C. E., Smith, S. A., Allman, J. F., Burleigh, J. G., Chaudhary, R., Coghill, L. M., et al. (2015). Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proc. Natl. Acad. Sci. U.S.A.* 112, 12764–12769. doi: 10.1073/pnas.1423041112

Hudson, L. N., Blagoderov, V., Heaton, A., Holtzhausen, P., Livermore, L., Price, B. W., et al. (2015). Inselect: automating the digitization of natural history collections. *PLoS One* 10:e0143402. doi: 10.1371/journal.pone.0143402

Huson, D. H., Auch, A. F., Qi, J., and Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Res.* 17, 377–386. doi: 10.1101/gr.5969107

Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., et al. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecol. Lett.* 16, 1245–1257. doi: 10.1111/ele.12162

Laird, S., Wynberg, R., Rourke, M., Humphries, F., Muller, M. R., and Lawson, C. (2020). Rethink the expansion of access and benefit sharing. *Science* 367, 1200–1202. doi: 10.1126/science.aba9609

Larsen, B. B., Miller, E. C., Rhodes, M. K., and Wiens, J. J. (2017). Inordinate fondness multiplied and redistributed: the number of species on earth and the new pie of life. *Q. Rev. Biol.* 92, 229–265. doi: 10.1086/693564

Lartillot, N., and Philippe, H. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* 21, 1095–1109. doi: 10.1093/molbev/msh112

Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., et al. (2018). Earth BioGenome Project: sequencing life for the future of life. *Proc. Natl. Acad. Sci. U.S.A.* 115, 4325–4333. doi: 10.1073/pnas.1720115115

Linard, B., Swenson, K., and Pardi, F. (2019). Rapid alignment-free phylogenetic identification of metagenomic sequences. *Bioinformatics* 35, 3303–3312. doi: 10.1093/bioinformatics/btz068

Lipscomb, D., Platnick, N. I., and Wheeler, Q. (2003). The Intellectual content of taxonomy: a comment on DNA taxonomy. *Trends Ecol. Evol.* 18, 65–66. doi: 10.1016/S0169-5347(02)00060-5

Losos, J. B. (1996). Phylogenetic perspectives on community ecology. *Ecology* 77, 1344–1354. doi: 10.2307/2265532

McKenna, D. D., Shin, S., Ahrens, D., Balke, M., Beza-Beza, C., Clarke, D. J., et al. (2019). The evolution and genomic basis of beetle diversity. *Proc. Natl. Acad. Sci. U.S.A.* 116, 24729–24737. doi: 10.1073/pnas.1909655116

Merckx, V. S., Hendriks, K. P., Beentjes, K. K., Mennes, C. B., Becking, L. E., Peijnenburg, K. T., et al. (2015). Evolution of endemism on a young tropical mountain. *Nature* 524, 347–350. doi: 10.1111/zsc.12501

Min, X. J., and Hickey, D. A. (2007). Assessing the effect of varying sequence length on DNA barcoding of fungi. *Mol. Ecol. Notes* 7, 365–373. doi: 10.1111/j.1471-8286.2007.01698.x

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G., and Worm, B. (2011). How many species are there on Earth and in the ocean? *PLoS Biol.* 9:e1001127. doi: 10.1371/journal.pbio.1001127

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., et al. (2001). Terrestrial ecoregions of the world: a new map of life on earth: a new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *Bioscience* 51, 933–938. doi: 10.1641/0006-3568(2001)051[0933:teotwa]2.0.co;2

Papadopoulou, A., Taberlet, P., and Zinger, L. (2015). Metagenome skimming for phylogenetic community ecology: a new era in biodiversity research. *Mol. Ecol.* 24, 3515–3517. doi: 10.1111/mec.13263

Pollock, D. D., Zwickl, D. J., McGuire, J. A., and Hillis, D. M. (2002). Increased taxon sampling is advantageous for phylogenetic inference. *Syst. Biol.* 51, 664–671. doi: 10.1080/10635150290102357

Pons, J., Ribera, I., Bertranpetit, J., and Balke, M. (2010). Nucleotide substitution rates for the full set of mitochondrial protein-coding genes in Coleoptera. *Mol. Phylogenet. Evol.* 56, 796–807. doi: 10.1016/j.ympev.2010.02.007

Quicke, D. L., Smith, M. A., Janzen, D. H., Hallwachs, W., Fernandez-Triana, J., Laurenne, N. M., et al. (2012). Utility of the DNA barcoding gene fragment for parasitic wasp phylogeny (Hymenoptera: Ichneumonoidea): data release and new measure of taxonomic congruence. *Mol. Ecol. Resour.* 12, 676–685. doi: 10.1111/j.1755-0998.2012.03143.x

Rees, J. A., and Cranston, K. (2017). Automated assembly of a reference taxonomy for phylogenetic data synthesis. *Biodivers. Data J.* 5:e12581. doi: 10.3897/BDJ.5.e12581

Robinson, G. E., Hacket, K. J., Purcell-Miramontes, M., Brown, S. J., Evans, J. D., Goldsmith, M. R., et al. (2011). Creating a buzz about insect genomes. *Science* 331:1386. doi: 10.1126/science.331.6023.1386

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584. doi: 10.7717/peerj.2584

Rosenberg, M. S., and Kumar, S. (2001). Incomplete taxon sampling is not a problem for phylogenetic inference. *Proc. Natl. Acad. Sci. U.S.A.* 98, 10751–10756. doi: 10.1073/pnas.191248498

Rubinoff, D., and Holland, B. S. (2005). Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Syst. Biol.* 54, 952–961. doi: 10.1080/10635150500234674

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351

Song, F., Li, H., Jiang, P., Zhou, X., Liu, J., Sun, C., et al. (2016). Capturing the phylogeny of Holometabola with mitochondrial genome data and Bayesian site-heterogeneous mixture models. *Genome Biol. Evol.* 8, 1411–1426. doi: 10.1093/gbe/evw086

Srivathsan, A., Lee, L., Katoh, K., Hartop, E., Kutty, S., Wong, J., et al. (2021). ONTbarcoder and MinION barcodes aid biodiversity discovery and identification by everyone, for everyone. *BMC Biol.* 19:217. doi: 10.5281/zenodo.5115258

Tansley, M. (2020). *A Site-Based Phylogenetic Analysis and Biogeographic Reconstruction of the Dung Beetles (Scarabaeinae)*. Master's thesis. London: Imperial College London.

Tautz, D., Arctander, P., Minelli, A., Thomas, R. H., and Vogler, A. P. (2003). A plea for DNA taxonomy. *Trends Ecol. Evol.* 18, 71–74. doi: 10.1016/S0169-5347(02)00041-1

Timmermans, M. J. T. N., Barton, C., Haran, J., Ahrens, D., Culverwell, L., Ollikainen, A., et al. (2016). Family-level sampling of mitochondrial genomes in Coleoptera: compositional heterogeneity and phylogenetics. *Genome Biol. Evol.* 8, 161–175. doi: 10.1093/gbe/evv241

Vogler, A. P., and Monaghan, M. T. (2007). Recent advances in DNA taxonomy. *J. Zool. Syst. Evol. Res.* 45, 1–10. doi: 10.1111/j.1439-0469.2006.00384.x

Wang, H. C., Minh, B. Q., Susko, E., and Roger, A. J. (2018). Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst. Biol.* 67, 216–235. doi: 10.1093/sysbio/syx068

Warnock, R. C. M., Heath, T. A., and Stadler, T. (2020). Assessing the impact of incomplete species sampling on estimates of speciation and extinction rates. *Paleobiology* 46, 137–157. doi: 10.1017/pab.2020.12

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., et al. (2018). BUSCO Applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 353, 543–548. doi: 10.1093/molbev/msx319

Webb, C. O., Ackerly, D. D., McPeek, M. A., and Donoghue, M. J. (2002). Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.* 33, 475–505. doi: 10.1146/annurev.ecolsys.33.010802.150448

Wiemers, M., and Fiedler, K. (2007). Does the DNA barcoding gap exist? - A case study in blue butterflies (Lepidoptera: Lycaenidae). *Front. Zool.* 4:8. doi: 10.1186/1742-9994-4-8

Wiley, E. O., and Lieberman, B. S. (2011). *Phylogenetics: Theory and Practice of Phylogenetic Systematics*, 2nd Edn. Hoboken, NJ: Wiley and Black.

Zamani, A., Vahtera, V., Sääksjärvi, I. E., and Scherz, M. D. (2021). The omission of critical data in the pursuit of 'revolutionary' methods to accelerate the description of species. *Syst. Entomol.* 46, 1–4. doi: 10.1111/syen.12444

Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19(Suppl.6):153. doi: 10.1186/s12859-018-2129-y

Zhou, X., Li, Y., Liu, S., Yang, Q., Su, X., Zhou, L., et al. (2013). Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *Gigascience* 16:2047. doi: 10.1186/2047-217X-2-4

Conflict of Interest: APV is a co-founder and scientific advisor of NatureMetrics, a company providing commercial services for DNA-based biomonitoring.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

APPENDIX 1

1. Field work protocol
    a. SITE-100 employs standardized site-based sampling, using a set of trapping methods and trap type replication, for example 1 month sampling of pitfall traps collected from every 2 days is employed on one site, e.g., a one hectare plot, or, to cover as many kinds of habitat as possible within a given site, or to focus on collections for community studies, such as canopy fogging transects;
    b. At a minimum each site is sampled with a standardized number of traps, including flight interception traps (FIT), pitfall trap, malaise trap (MT); in addition, other sampling types are used, including sweeping, light trapping, canopy fogging, suspended land and air malaise (SLAM), Berlese traps, etc. These methods are replicated through time, e.g., 1 month of trap collection every 2 days;
    c. Specimen preservation: specimens should be labeled as per each trapping method and dated, and then transferred to 96% ethanol as quickly as possible, and stored at –20 C until further processing.
2. Specimen processing
    a. All specimens are imaged by SLR camera (Canon EOS 650D) and/or Zeiss Axio Zoom.v16 motorized stereo zoom microscope;
    b. Morphospecies are selected through the images by Inselect software (Hudson et al., 2015).
    c. Morphospecies are extracted individually and non-destructively using DNeasy 96 Blood and Tissue Kit (Qiagen, Venlo, Netherlands) and sequenced by shotgun sequencing NGS to obtain mitogenomes of morphospecies;
    d. Bulk samples are extracted non-destructively using the same kit as morphospecies extraction, and amplified for metabarcoding (418 bp portion of the *cox1* region) and sequenced by Illumina MiSeq v.3 (2 × 300 bp paired−end). Mi-seq.
3. Data processing
    a. Raw sequence data are processed through the pipeline [demultiplexing—trimming—merging—quality, size, frequency, and chimera filtered through NAPtime (Creedy et al., 2019, 2022) and VSEARCH (Rognes et al., 2016)]. This pipeline generates the ASVs which are then clustered to obtain metabarcodes (OTUs);
    b. Phylogeny reconstruction by hierarchical datasets;
    c. Metabarcodes placement on the phylogeny.

frontiers | Frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 17 June 2022
doi: 10.3389/fevo.2022.863080

Effects of Habitat and Competition on Niche Partitioning and Community Structure in Neotropical Ants

Alex Salas-López[1]†, Cyrille Violle[2], François Munoz[3], Florian Menzel[4] and Jérôme Orivel[1]*

[1] CNRS, UMR Ecologie de Forêts de Guyane (AgroParisTech, CIRAD, INRAE, Université de Guyane, Université des Antilles), Campus Agronomique, Kourou, France, [2] CEFE, Université Montpellier, CNRS, EPHE, IRD, Montpellier, France, [3] University of Grenoble-Alpes, Laboratoire Interdisciplinaire de Physique (LIPhy), Saint-Martin-d'Hères, France, [4] Institute of Organismic and Molecular Evolution, Johannes Gutenberg University, Mainz, Germany

OPEN ACCESS

Edited by:
Gengping Zhu,
Tianjin Normal University, China

Reviewed by:
Flavia Maria Darcie Marquitti,
State University of Campinas, Brazil
Henintsoa Onivola Minoarivelo,
University of Cape Town, South Africa

***Correspondence:**
Alex Salas-López
alex.salaslopez@gmail.com

†Present address:
Alex Salas-López,
Ávila, Spain

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *26 January 2022*
Accepted: *16 May 2022*
Published: *17 June 2022*

Citation:
Salas-López A, Violle C, Munoz F, Menzel F and Orivel J (2022) Effects of Habitat and Competition on Niche Partitioning and Community Structure in Neotropical Ants. Front. Ecol. Evol. 10:863080. doi: 10.3389/fevo.2022.863080

Competition for limited resources can yield two contrasting outcomes in community structure, namely, either (i) dominance of most competitive species (with functional convergence of the traits conferring this ability), or (ii) niche partitioning of species using distinct resources. In addition, varying resource availability in different environmental contexts is expected to yield varying community dynamics and composition between the contexts (habitat filtering). We addressed resource-based ant community structure in a tropical ecosystem. We expected ant species to display varying trophic preferences and foraging behaviors, allowing habitat selection and niche differentiation in ant assemblages. Furthermore, we expected habitat filtering to occur between open and forested areas in the landscape mosaic, and competition to further influence local species co-occurrence. We assessed resource use in nine ant assemblages distributed in two habitats (i.e., forests and croplands), devising two separate experiments using bait-traps to characterize ant species' trophic preference (e.g., eating prey, seeds, sugars) and their ability to obtain a same resource in heterogeneous forms (e.g., on vegetation, litter, with variable amounts. . .). The majority of baits offered were rapidly exploited in the two habitats suggesting important resource limitations. Forest and cropland ant communities differed, however, in the proportions of resources exploited, suggesting different competitive pressures toward specific resources between habitats. Within each habitat, ants preferentially exploited the same resources, suggesting habitat filtering, but locally, interspecific resource partitioning resulted in a reduction of resource overlap compared to habitat scale. Our study provides evidence of the effects of habitat filtering and competition for resource in tropical ant community structure. Our findings also suggest that niche filtering and niche partitioning are co-variant forces determining the identity of the species present in local assemblages.

Keywords: assembly rules, niche partitioning, habitat filtering, Formicidae, resource limitation, community ecology, foraging ecology

INTRODUCTION

A major goal of community ecology has been to explain the occurrence patterns of species within a region as a consequence of assembly rules (Keddy, 1992; Weiher et al., 2011; Götzenberger et al., 2012; Hille Ris Lambers et al., 2012). The expectation is that interspecific niche differences determine the suitability of species to different habitat patches as well as their interactions in local communities. Niche filtering is explained by common adaptations of species sharing a habitat, relevant for their survival and competitive performance. Particularly, species foraging effort is expected to primarily target resources and strategies providing maximal reward in a given environments (Pyke et al., 1977). Consequently, within a same habitat, species are in average more likely foraging for the same abundant/high quality resources, resulting in interspecific resource overlap (Fox and Vasseur, 2008). In absence of mechanisms regulating species' populations (e.g., natural enemies), resources may become limiting (Wiens, 1977; Schoener, 1983). In that case, dominant species (i.e., competitively superior phenotypes) may displace other species from local assemblages acting as a biological filter. The most frequently invoked mechanism preventing competitive exclusion is interspecific niche partitioning as explained by the limiting similarity theory (MacArthur and Levins, 1967; Abrams, 1983). Indeed, species differing in their resource requirements, or in their performance to exploit different resources, are more likely to co-occur, because competition intensity is greater between ecologically redundant species than between complementary ones. Another major concern is how assembly rules operate at different organizational levels within a region. Habitat filters may favor niche overlap among species sharing a particular habitat, but similarity may be limited within assemblages as a consequence of competitive interactions (**Figure 1**).

In ants, competition has been considered the key-stone of ant assembly rules (Andersen, 1992; Davidson, 1998; Parr and Gibb, 2010). Ant abundance is huge, particularly in the tropics, where ants represent an important fraction of animal biomass (Davidson and Patrell-kim, 1996). Consequently, the foraging activity of ants for food resources is very intense, as demonstrated by many studies using baits to attract ant species (Parr and Gibb, 2010). Indeed, food baits are frequently exploited in a few hours or even minutes, and aggressive behaviors between species over food resources are frequently observed (Savolainen and Vepsäläinen, 1988; Hölldobler and Wilson, 1990; Andersen, 1992). Besides, due to the sessile nature of ant nests, the majority of ant species have a limited territory to forage for their food resources (Andersen, 1991). The intensity different resources are exploited is frequently considered as a consequence of environmental limitation in nutrient availability (Kaspari and Yanoviak, 2001; Hahn and Wheeler, 2002; Bihn et al., 2008; Peters et al., 2014). This is basically deduced from the stoichiometric ecology framework, where species require a balanced set of nutrients, and compete for the most limiting. Moreover, in addition to specific nutritional requirements, the access to nutrients is affected by how these become available *in natura*. For instance, the same nutrients are accessible to different species depending on whether they are contained on dead organic matter, prey, plants, or depending on microhabitat variations (Cerdá et al., 1998; McGlynn and Kirksey, 2000; Agarwal and Rastogi, 2009). In particular, the abundance of a thick litter layer or the vegetation cover, are important factors filtering different ant species (Gibb and Parr, 2013; Gibb et al., 2015; Nooten et al., 2019). Also, some ants forage on vegetation while many others search resources within the litter or foraging on the ground (Sarty et al., 2006; Brandão et al., 2012). Finally, some species are fierce defenders of food resources, while others survive by finding and exploiting faster such resources (Pearce-Duvet et al., 2011; Cerdá et al., 2013). Therefore, interspecific niche differences in nutritional requirements or in the ability to exploit different resources may be essential to explain species co-occurrence within assemblages (Luque and Reyes López, 2007; Lanan, 2014; Houadria et al., 2015). In this matter, a special place in ant community assembly has been attributed to dominant species. Indeed, some ant species excel at exploiting very important amounts of resources and aggressively displace competitors affecting the structure of assemblages (Andersen, 1992; Holway, 1999; Dejean et al., 2007; Parr, 2008). This may lead to intraspecific resource use variations depending on the outcomes of competition (Savolainen and Vepsäläinen, 1988; Sanders and Gordon, 2003; Blüthgen and Fiedler, 2004). Such competitive outputs may be influenced by the abundance of different microhabitats and resources affecting the foraging conditions and available resources (Fowler et al., 2014; Ipser and Gardner, 2020).

In this study we investigated interspecific niche partitioning in ant assemblages in relation to habitat condition and hypothesized that local assembly results from niche processes related to the use of food resources. We assessed ant occurrence patterns at food baits in nine ant assemblages in two habitat types within the same region. In two separate experiments, we devised two major dimensions of niche partitioning, namely, trophic differences and resource acquisition strategies. The first can be affected by the nutritional requirements of different species as well as ant ability to handle different food types. Resource acquisition strategies refer to the ability of ant species to exploit a same type of resource in different displays representing the heterogeneous access to food resources in the environment. First, we assessed and compared the exploitation intensity of different food resources within assemblages. The objective here was to confirm that resource limitations exist, as a preliminary condition to hypothesize that competition may be important; and to provide evidence of resource limitation differences between the two habitats to support habitat filtering. Second, we investigated intraspecific niche differences among assemblages, and interspecific niche differences within them. We examined general ant assemblages and particularly dominant species (i.e., those exploiting the largest fraction of resources). The aim was again two-fold, to confirm that intraspecific resource use is not contingent to local conditions, and that niche differences at intraspecific level are lower than between species, as expected from the limiting similarity theory. Finally, we tested whether resource-based filtering existed at the habitat level, and at the assemblage level. We expected a significant niche overlap among species sharing a habitat, but this can be reversed at

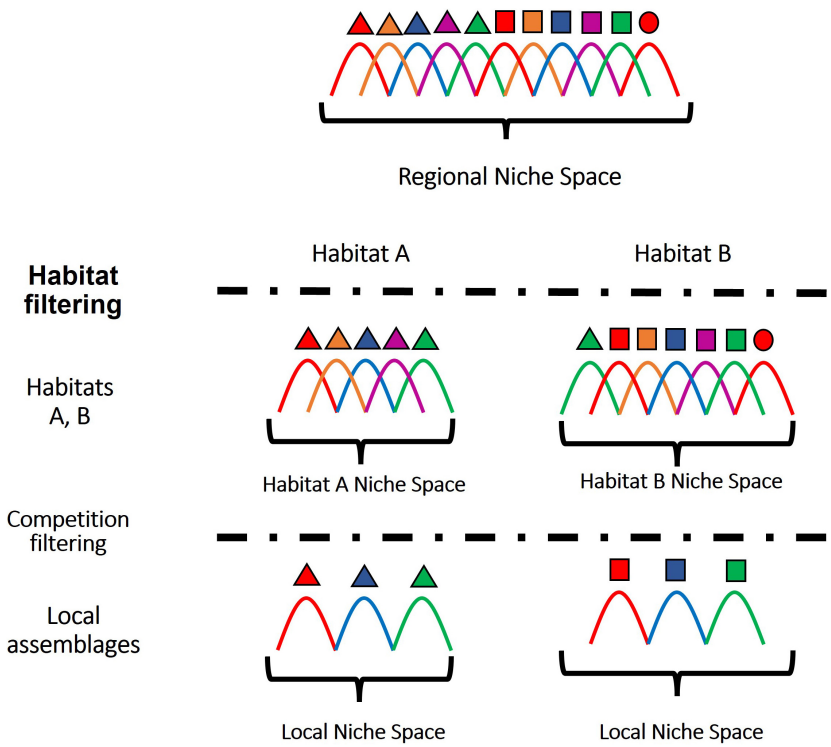


FIGURE 1 | Habitat and competition filters simultaneously shape the niche space of local assemblages. Species are represented with different shapes and colors. Each species niche width and abundance is represented with a distribution curve positioned along an ecological gradient that represents the regional niche space. Within each habitat, only a narrower set of species is able to survive in relation to their traits. In local assemblages, competition may result in the local extinction of similar competitors, due to the limiting similarity principle, selecting for an overdispersed niche space compared to the potential niche range within that habitat.

the assemblage level due to the effects of competition, in which case, the strength of the filtering should be lower than at the habitat level.

MATERIALS AND METHODS

Study Site and Sampling Scheme

The study area is located in the coastal part of French Guiana at *La Montagne des Singes* (5°04'19"N; 52°41'42"W). In total, nine plots were sampled in two habitat types: five forests and four slash-and-burn croplands. Forests presented mixed vegetation representative of rainforests of the facies *Fabaceae*, *Chrysobalanaceae*, *Lecythidaceae* of the coastal part of French Guiana. Croplands consisted of traditional young mixed crops (3–5 years) of different plant species, mainly manioc, pineapple, sugar cane, pepper, and fruit trees. The two habitats were selected to represent contrasting environmental conditions regarding habitat complexity, litter density, shade, and disturbance regularity, resulting in an almost full turnover of ant community composition across habitats (personal observation). Still, assemblages within the same habitat type shared an important fraction of species. Plots consisted of 20 sampling locations separated by 10 m were set up on a rectangular grid (1200 m^2). This sample size was chosen to represent local assemblages, and the number of points based on a recommended minimum for characterizing ant communities (Agosti and Alonso, 2000). A similar display in previous experiments dealing with trophic differentiation of ants showed, that this number of samples adequately captured the largest fraction of ants foraging for food resources on the ground (Houadria et al., 2015; Salas-López et al., 2017). A distance of at least 250 m was left between the plots, and all of them were situated within a radius of 3 km and relatively interspersed [described in Salas-López et al. (2017, 2018)]. The distance between plots is enough to limit autocorrelation, and the plots are still not too far away to prevent large-scale environmental variation, and we could consider a single, consistent pool of species. Sampling was carried out between March and October 2013, alternating croplands and forests in dates, always in the morning (8:00–11:00 a.m.) and only in dry conditions (i.e., in the absence of rain and/or flooding) to avoid, as much as possible, sampling biases linked to soil conditions and thermal stress.

Ant Communities and Foraging Strategies

The sampling scheme enabled to obtain at the same time data about community composition, and interspecific niche differences.

First, we devised a food type experiment (hereafter referred as trophic) to investigate ant foraging behavior depending on the

nature of the resource. Seven different food items representing some of the most frequently available sources of essential nutritional requirements for ants, including carbohydrates, proteins or lipids [see Houadria et al. (2015)]. Sucrose and melezitose constitute the main sugar sources for ants obtained from plant nectar and sap-sucking insects, respectively (3 ml 25% w/w dropped onto paper toweling). Insects constitute the main source of proteins and lipids for most ant species; however, insects can be differently accessible depending on their size or whether they are dead or alive. We offered dead insects (3 g of dead, crushed *Tenebrio molitor* mealworms), small living prey (at least 20 *Anoplotermes* sp. termites) and large living prey (two, living 1.5 and 3 cm-long mealworms). Seeds were offered as a mixture of peanuts and rice to represent a highly lipidic and a mainly starch composed seed. Finally, excrements are exploited by some species, and therefore we used bird droppings (3 g of chicken excrement).

Second, we devised a resource heterogeneity experiment (hereafter referred as acquisition strategies) to investigate ant foraging behavior depending on the location of the resource. We offered dead insects (i.e., the most attractive resource considering the number of ant genera attracted and of recruited individuals in the first experiment) in different bait displays (**Figure 2**). These baits were presented as a standard resource on the leaf litter (3 g of dead *Tenebrio molitor* as in the previous experiment), as interstitial litter resource [a plastic container where an open Eppendorf containing the resource was placed inside containing with six 0.5 cm diameter holes providing access to the ants; e.g., Sarty et al. (2006)], as a resource on vegetation (presented in an open Eppendorf which was tacked to the trunks of trees at a height of 1.3 m; Kaspari and Yanoviak, 2001), as small particles (close to 0.1 g of dried mealworms ground into powder; McGlynn and Kirksey, 2000), and rapid discovery (3 g of DT were placed in Petri dishes, surveyed every 5 min, and then closed as soon as any ant activity was observed; Pearce-Duvet et al., 2011).

With the exception of the rapid discovery bait, the other bait traps were surveyed every 5–10 min to monitor ant activity (i.e., the species consuming the items were noted), and collected after 1 h. After each sampling session, the bait traps were closed to capture the ants. The traps were then taken to the laboratory, and the specimens were killed by freezing at −20°C. The ant specimens were then counted and sorted to genera using the identification key by Bolton (2003) and then to morphospecies based on morphological differences. A reference collection was built and can be found at the UMR Ecofog, Kourou, France.

Statistical Tests

We used incidence data for analyses (e.g., the number of times each of the species was found at a given bait type for an observation scale considered). Within each assemblage, a species can be present as much as 20 times for a given resource (i.e., the number of sampling locations). The trophic experiment enabled the capture of a total of 81 species in the forest habitat and 26 in the cropland habitat. In the acquisition strategies experiment, 66 species were captured in the forest habitat and 15 in the cropland habitat. Species captured in five or more occasions were retained for analyses. This was the minimal information to discriminate differences in the resource use patterns between two species according to the Chi-square (χ^2) test (McCrum-Gardner, 2008). All the tests were performed separately for the two data sets (trophic experiment and resource acquisition). The statistical tests used consist on standard procedures suitable to investigate niche patterns on incidence data in discrete resource categories (Gotelli and Graves, 1996).

Resource Limitations

We first assessed the intensity of exploitation of different food types (using the trophic experiment data) as the percentage of baits exploited and the average number of species exploiting baits in the lapse of an hour. To test whether resource limitations differed between habitats and among assemblages,

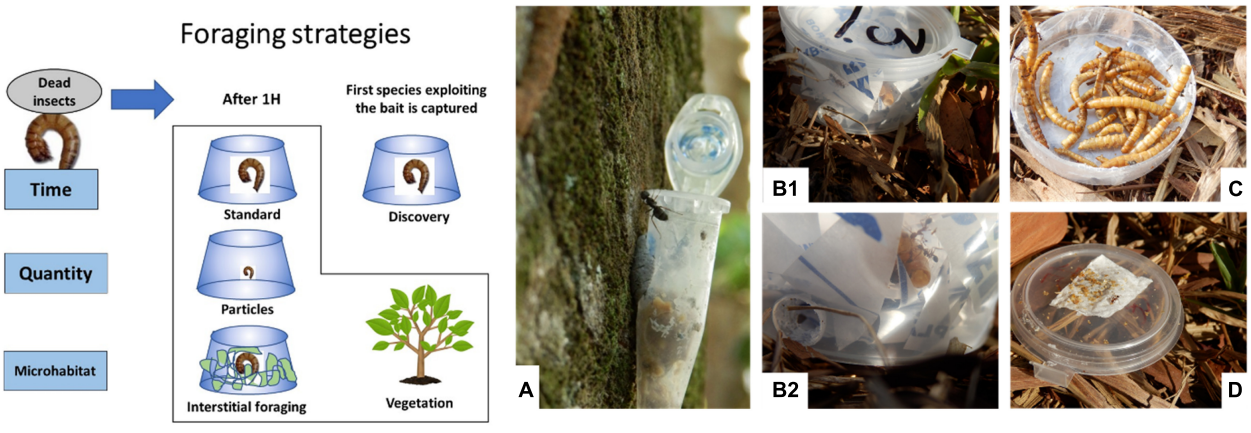


FIGURE 2 | Foraging strategies sampling protocol. The left side of the figure illustrates which environmental factors and niche partitioning dimensions were aimed. At the right, the sampling devices used to emulate different resource acquisition strategies. **(A)** Vegetation foraging. **(B1,B2)** Interstitial foraging. **(C)** The standard bait display also used for the trophic experiment, and applied for both discovery and dominance, and **(D)** the small particles.

the incidence data of all species was summed up to account for exploitation intensity of the different resources. Then a contingency test based on the Chi2 distribution was conducted (here after proportionality test). For each habitat type, the algorithm generated random incidence data by keeping constant the sums in the rows (assemblages/habitat) and columns (resource types). A significant result in the test indicates that resources were consumed in different proportions between the elements compared.

Species Level Resource Use Differences

We characterized ant niches based on their incidence on different resource categories. Then intraspecific and interspecific resource use variations were assessed using the proportionality test. Intraspecific resource use variations were examined on species sufficiently represented at different assemblages (i.e., 10 or more incidences were found for at least three plots). Based on these conditions, a total of 24 tests were performed, where rows were assemblages, and resource categories were columns. Then, interspecific niche differences were investigated within assemblages. Only species with an incidence of five or more were retained for analyses. Rows were species and columns resources. In keeping with the limiting similarity principle, we expected interspecific differences within assemblages but not intraspecific resource use variations. Moreover, the finding of intraspecific level variations within a same habitat would compromise the interpretation of species suitability to different environments as a consequence of their resource use, as well as its competitive effect on other species.

Niche Overlap at the Habitat and the Assemblage Levels

We assessed the extent of habitat filtering among ant species within communities and habitats, based on the null expectation that species can use random proportions of resources (Gotelli and Graves, 1996). We used Ecosim routines of niche overlap implemented in R, at the habitat level first, and then on the different assemblages. The options RA3 algorithm and the Pianka overlap index were chosen as the most frequent procedure for resource overlap hypothesis (Gotelli and Entsminger, 2001). Observed niche overlap in different levels of organization was compared to the null distributions using a two-tailed statistical test. Any deviation from the null models indicated an influence of niche variation among species. Values lower than the 2.5% quantile indicated that species differ more in their resource use than expected by chance. Values over the 97.5% indicated more similarity than expected by chance. In addition, we calculated the standard effect size (SES) of observed overlap values. The absolute values of SES represent the standardized deviation from the null distribution. SES was particularly useful to compare resource use overlap patterns at the habitat and the assemblage levels.

Multiple Comparison Correction

To deal with the problem of the multiple comparisons, we used the method of false discovery rates (FDR), and the algorithms described in Pike (2011). This *post-hoc* method is based on the evaluation of the distribution of *P*-values to evaluate their signification, instead of a normative reduction of the confidence interval in more standard methods (e.g., Bonferroni). The key benefit of these methods is that they are much more powerful than Bonferroni-type comparisons to determine which results are true positives. In addition, they are also capable of detecting false negatives. In the presentation of our data, we maintain the *P*-values provided by the original statistical tests, and only report FDR results when discordance was found.

Ordinations

To identify the major components of species niche differences within habitats, we performed Principal Component Analyses using the software Past 3.0 (Hammer et al., 2001). Species of both habitats were ordinated together in a resource space, where the components were the different resources projected. Only species presenting an incidence of five or more at the habitat level were retained for the ordination. Accordingly, in the trophic experiment, 47 and 15 species were retained for forests and croplands, respectively and 40 and 12 in the resource acquisition experiment. Dominant species were considered apart. We chose those exploiting together the 75% or more of resources at the habitat level, which distinguished 7 species in the forest habitat and 3 species in the cropland habitat. These species are expected to reflect the habitat effect more clearly in favoring specific ecological strategies but also competition effects on ecological similarity.

Species and principal components were normalized prior to ordination, to avoid inertia biases related to differences in the exploitation intensity of different resources or related to species abundances. The niche space occupied by the ants in each habitat, and by dominant ants, was delimited with convex hulls. These join the peripheral points of the ecological volume of a specified group, and are frequently related to ecological filtering (Cornwell et al., 2006).

RESULTS

Resource Limitations

The percentage of baits exploited was very high for most resources. Particularly in just 1 h, the great majority of baits containing dead insects, sugars, seeds, and termites were exploited, frequently, by several species simultaneously (**Table 1**). Despite the lower species diversity in croplands at assemblage and habitat level, the average number of species per bait was similar than in forests.

We found no differences in the exploitation intensity of different food resources among assemblages of the same habitat type, neither for the trophic experiment nor foraging strategies according to the proportionality test. Significant differences were found, however, between forests and croplands ($\chi^2 = 85.23$, $P < 0.001$) supporting the idea that the ants in each habitat differed in their resource requirements. Additionally, the Principal Component Analysis (**Figure 3**) illustrated that the dominant species of each habitat type consumed relatively different resources. Species in croplands, occupied a more central place, and consumed insects in several forms, while forest species

TABLE 1 | Resource exploitation intensity.

	Resource	Percentage of exploited baits	Average number of species per bait	Average number of species per plot
Forest	Big prey	53 ± 1	0.61 ± 0.09	4.8 ± 1.9
	Termite	92 ± 1	1.34 ± 0.21	13.4 ± 1.6
	Dead T.	97 ± 4	1.93 ± 0.44	16.8 ± 3.4
	Excrement	44 ± 1	0.62 ± 0.11	9.2 ± 2.6
	Seeds	99 ± 2	2.35 ± 0.26	18.6 ± 3.3
	Sucrose	82 ± 7	1.75 ± 0.43	14.6 ± 3.6
	Melezitose	80 ± 8	1.36 ± 0.26	11.8 ± 2.7
Cropland	Big prey	92 ± 8	1.14 ± 0.19	3.5 ± 1.5
	Termite	100 ± 0	1.86 ± 0.36	6.5 ± 1.1
	Dead T.	100 ± 0	1.76 ± 0.25	8.5 ± 3.2
	Excrement	49 ± 2	0.59 ± 0.27	4.8 ± 1.3
	Seeds	96 ± 7	2.08 ± 0.2	9 ± 1.4
	Sucrose	100 ± 0	1.95 ± 0.19	9.8 ± 1.8
	Melezitose	91 ± 7	1.43 ± 0.31	7.5 ± 1.1

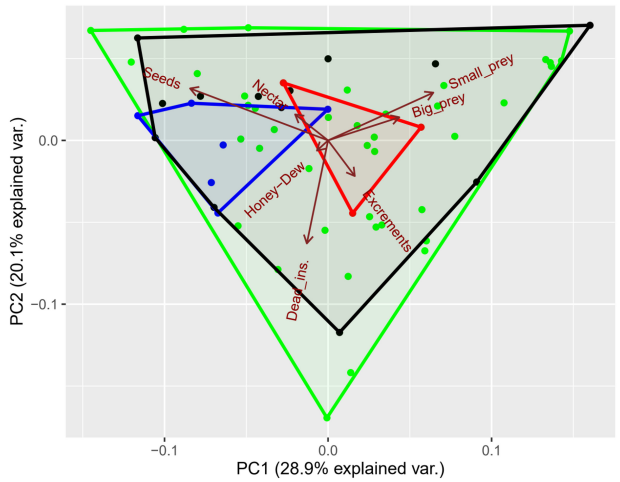


FIGURE 3 | Principal Component Analysis of the trophic niche space. The first and second axes represent the 28.9% and 20% of the variance, respectively. Dots and crosses correspond to forest and cropland species, respectively, distributed along resource axes. Convex hulls encompass four groups of species: in red and blue are dominant species in croplands and forests, respectively. In black and green are the remaining species for croplands and forests. The food types are indicated as follows: termites (Small), mealworms (Large), excrements (Excrements), dead mealworms (Dead), melezitose (Honey-dew), sugar (Nectar), and Seeds.

presented a greater tendency to consume sugars and seeds. In contrast, no differences were found in the intensity of resource acquisition strategies by dominant species between habitats despite the important difference in the niche space delimited by the convex hulls in these two habitats when considering the whole pool of species. The absence of differences can be explained, however, because the reduced acquisition strategies niche space observed in croplands constitute a subassembly of those observed in forests by both dominants and subordinate ants.

Species Level Resource Use Differences

Variations in resource use patterns at the intraspecific level were extremely rare for the two experiments performed. Out of a total of 24 proportionality tests (**Table 2**), significant differences were only found in the means of the resource acquisition of *Nylanderia* sp.1. This result disappeared after applying a correction for multiple comparisons (FDR adjusted-P = 0.16). The absence of intraspecific resource use variations provides robust evidence of the constancy of the species' response to resource limitations and effect of other species.

Interspecific trophic partitioning was found in all the assemblages after applying multiple comparison corrections (**Table 3**). These patterns were also suggested by the wide trophic space delimited with the convex hulls in the Principal Component Analysis (PCA) in both habitats (**Figure 3**). Trophic partitioning was also important among dominant species in all the cropland assemblages, but only in a 40% of the cases in forest assemblages (**Table 3**). These patterns are also reflected by the PCA, which reveals that the three species dominant in croplands were distantly positioned from one another (**Figure 3**). Conversely, forest dominant species were very closely positioned in the trophic space, and only one of them appears as relatively different to the others exploiting insects with greater frequency (**Figure 3**).

Partitioning in resource acquisition strategies was detected in all the forest assemblages after applying multiple comparisons correction, but in only one of the cropland assemblages (**Table 4**). The number dropped to four (80%) in forest assemblages when only the dominant species were considered and remained equal (25%) for croplands. Indeed, it can be observed (**Figure 4**) that species in forest habitat present a greater diversity of foraging preferences in relation to resource heterogeneity, and particularly dominant species which were quite evenly distributed covering an important fraction of this volume.

High Niche Overlap at the Habitat Level, Decreases in Local Assemblages

At the habitat scale (pooled assemblages), the observed trophic overlap was greater than expected by chance for forests ($Pianka_{obs}$ = 0.54 vs. $Pianka_{null}$ = 0.50; $P < 0.001$) and croplands ($Pianka_{obs}$ = 0.65 vs. $Pianka_{null}$ = 0.57; $P < 0.001$) (**Table 5**). This indicates that some types of resources were more frequently exploited than others by most of the species. At the assemblage level, trophic overlap was also larger than expected by chance in four of the five forest assemblages (**Table 5**), but only in two of the cropland assemblages. Indeed, in croplands, the degree of deviation from the null expectation was considerably larger at the habitat scale than for the average assemblage ($SES_{Habitat}$ = 5.53 vs. a local mean of 1.39). The same result was observed in forest habitat compared to local assemblages ($SES_{Habitat}$ = 6.83 vs. a local mean of 2.73).

In the acquisition strategies experiment, forest species overlapped more than expected by chance in the displays used at the habitat scale ($Pianka_{obs}$ = 0.66 vs. $Pianka_{null}$ = 0.60,

TABLE 2 | Contingency tests calculated on resource use matrices by dominant species.

		Trophic experiment				Acquisition strategy			
Habitat	**Species**	**Plots**	**n**	**χ^2**	***P***	**Plots**	**n**	**χ^2**	***P***
Forest	*Wasmannia auropunctata*	5	79	22.48	0.31 ns	5	65	16.2	0.45 ns
	Ochetomyrmex neopolitus	5	69	21.53	0.6 ns	3	45	7.67	0.48 ns
	Pheidole sp.8	4	72	8.54	0.9 ns	–	–	–	–
	Pheidole sp.16	2	29	0.57	0.9 ns	5	20	12.84	0.85 ns
	Pheidole sp.45	4	43	11.57	0.52 ns	5	16	16.17	0.17 ns
	Nylanderia sp.1	3	38	8.49	0.33 ns	4	27	5.6	0.90 ns
	Solenopsis sp.1	5	38	23.47	0.27 ns	4	16	11.82	0.44 ns
	Crematogaster sp.2	3	31	12.08	0.48 ns	–	–	–	–
Cropland	*Solenopsis saevissima*	4	238	12.78	0.81 ns	4	195	7.88	0.8 ns
	Pheidole fallax	4	187	21.6	0.24 ns	4	77	5.08	0.96 ns
	Crematogaster sp.5	3	126	13.3	0.35 ns	4	100	18.2	0.09 ns
	Camponotus sp.3	3	27	5.67	0.68 ns	3	38	14.67	0.06 ns
	Nylanderia sp.4	3	26	9.69	0.73 ns	3	31	15.7	0.045*!

The number of assemblages considered (Plots) and incidence (n) are indicated. Significant differences from a random distribution are based on $\chi 2$ distribution and are indicated as follows: $^{*}0.01 < P < 0.05$; ns, non-significant; !, Non-significant after FDR correction.

TABLE 3 | Contingency tests calculated on trophic partitioning matrices on five forest and four cropland assemblages.

Habitat	**Plot**	**Number of spp.**	**Df**	**% Expl.**	**χ 2**	**Dominant spp.**	**Df**	**% Expl.**	**χ 2**
Forest	For1	13	72	72.1	90.813 •	5	24	45.3	31.254 ns
	For2	12	66	79.4	186.84***	5	24	50.0	47.24***
	For3	10	54	78.4	95.04**	5	24	57.0	31.496 ns
	For4	14	78	78.2	132.13***	5	24	48.5	13.082 ns
	For5	7	36	60.8	109.72***	5	24	53.1	89.2***
Cropland	Crop1	9	48	94.2	80.816**	3	12	69.1	25.788**
	Crop2	8	42	95.1	98.713***	3	12	76.5	49.118***
	Crop3	5	24	86.6	49.193**	3	12	86.0	29.064**
	Crop4	8	42	94.0	90.291***	3	12	65.0	28.841**

At the left, the analyses correspond to the entire assemblages. At the right, only dominant species are retained. The total number of species, and the number of dominant species (Dominant spp.), the percentage of exploited resources, and the degrees of freedom (Df) are indicated. Significant differences from a random distribution are based on $\chi 2$ distribution and are indicated as follows: • significant test after FDR correction; $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; ns, non-significant.

TABLE 4 | Contingency tests calculated on acquisition strategy partitioning matrices on five forest and four cropland assemblages.

Habitat	**Plot**	**Number of spp.**	**Df**	**% Expl.**	**χ^2**	**Dominant spp.**	**Df**	**% Expl.**	**χ^2**
Forest	For1	11	40	58.2	60.73*	5	16	33.6	32.17**
	For2	7	24	68.8	42.071**	5	16	58.9	30.887**
	For3	8	28	72.9	47.943**	5	16	57.9	26.714*
	For4	8	28	68.4	38.72 •	5	16	55.1	20.683 •
	For5	5	16	67.0	29.73*	5	16	67.0	29.73*
Cropland	Crop1	6	20	87.5	38.7**	3	8	57.2	10.24 ns
	Crop2	3	8	84.1	4.74 ns	3	8	84.1	4.74 ns
	Crop3	4	12	93.1	13.965 ns	3	8	85.1	11.70 ns
	Crop4	6	20	91.3	27.564 ns	3	8	67.1	16.24*

At the left, the analyses correspond to the entire assemblages. At the right, only dominant species are retained. The total number of species, and the number of dominant species (Dominant spp.), the percentage of exploited resources, and the degrees of freedom (Df) are indicated. Significant differences from a random distribution are based on $\chi 2$ distribution and are indicated as follows: • significant test after FDR correction; $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; ns, non-significant.

$P < 0.001$), but this pattern was only maintained in one of the local assemblages (**Table 5**). Accordingly, the standardized effect size at the habitat scale was more than three times larger than in any of the local assemblages. In the croplands, the species presented a random pattern of resource use at the habitat and assemblage scales (**Table 5**).

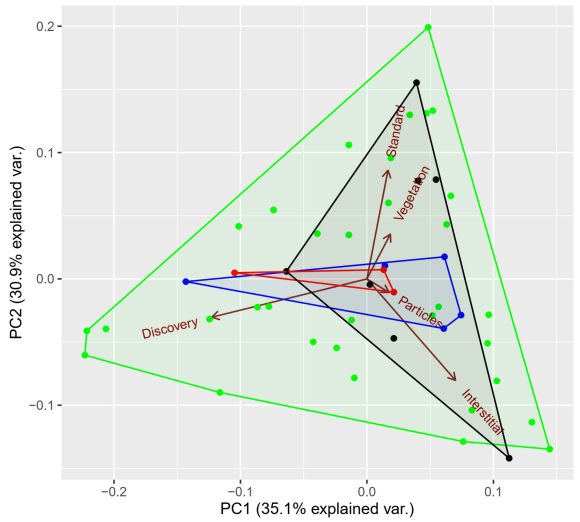


FIGURE 4 | Principal Component Analysis of the acquisition strategies niche space. The first and second axes represent the 35 and 31% of the variance, respectively. Dots and crosses correspond to forest and cropland species respectively distributed along resource axes. Convex hulls encompass four groups of species: in red and blue are dominant species in croplands and forests, respectively. In black and green are the remaining species for croplands and forests. The abbreviations correspond to acquisition strategies as follows: Standard, standard reference bait; Interstitial, litter-surrounded bait; Vegetation, baits placed on vegetation; Particles, small amounts; Discovery, rapid exploitation.

DISCUSSION

In this work we investigated the patterns of resource use in nine ant assemblages in two contrasted habitats. We used two different experiments representing important dimensions of niche partitioning, namely food type and resource acquisition strategies. In all the assemblages studied we found intense ant activity at food baits, which suggested resource limitations. Also, interspecific niche differences were found in all the habitats studied. Finally, species associated to forests and to croplands used more frequently the same resources and/or resource displays at the habitat level as well as at the assemblage level, suggesting niche filtering for each of these habitats. While we found filtering and partitioning in forests and croplands, they affected different ecological dimensions. We try to explain the observed patterns as a consequence of habitat filters and competitive interactions.

Prior to interpretations of assembly patterns, we investigated whether resource use intensity was a consistent property of individual species and entire assemblages. This was confirmed by each of the species examined displaying similar resource exploitation patterns in the different assemblages where they were present. Also, the characteristic assemblages of ants associated to forests and croplands respectively, presented specific patterns of resource exploitation that were consistent between assemblages of the same habitat. Because we are using resource use patterns to characterize the niche of species, as well as an indicator of resource limitations between assemblages, it seems necessary to confirm that resource use proportions are constant for a given species within assemblages of the same type. For example, different resource limitations may alter species competitive outputs (Kaspari et al., 2012; Correa and Winemiller, 2014; Nooten et al., 2019). And also, intraspecific resource use variations have been reported in response to environmental and competition effects (Savolainen and Vepsäläinen, 1988; Pfeiffer et al., 2014).

In our study, the intensity food resources were exploited differed between forests and croplands. An important trophic overlap was found in both habitats, although for different resources. Particularly, dominant species in forests consumed more sugars and seeds, while the dominant species in croplands consumed relatively more insects and appeared as more generalized in their trophic preferences. Such differences may be explained by environmental differences in resource availability between the two habitat types. For instance, a greater use intensity of proteinaceous resources in cropland vs. forests was found in

TABLE 5 | Observed and expected range of trophic niche overlap measured at assemblage and habitat level.

		Trophic experiment				Acquisition strategy			
Habitat	**Plot**	**Obs**	**95% IC limits**	***P*-value**	**SES**	**Obs**	**95% IC limits**	***P*-value**	**SES**
Forest	For1	0.609	0.526–0.603	0.012	2.726	0.590	0.553–0.639	ns	0.143
	For2	0.543	0.434–0.535	0.012	2.7	0.576	0.488–0.631	ns	0.663
	For3	0.611	0.445–0.56	0.002	4.058	0.635	0.566–0.677	ns	0.742
	For4	0.622	0.43–0.513	0.001	7.095	0.735	0.63–0.74	0.033	2.339
	For5	0.401	0.294–0.489	ns	0.558	0.610	0.544–0.712	ns	−0.048
	$For_{Habitatlevel}$	0.54	0.490–0.513	0.001	6.831	0.656	0.590–0.616	0.001	7.978
Croplands	Crop1	0.58	0.497–0.603	ns	1.386	0.733	0.687–0.803	ns	0.15
	Crop2	0.564	0.449–0.587	ns	1.809	0.919	0.88–0.967	ns	0.253
	Crop3	0.594	0.52–0.703	0.028	0.028	0.82	0.727–0.88	ns	0.841
	Crop4	0.688	0.592–0.695	0.035	0.035	0.785	0.766–0.856	ns	−0.458
	$Crop_{Habitatlevel}$	0.654	0.546–0.603	0.001	5.525	0.722	0.709–0.781	ns	0.635

The results correspond to the observed and expected overlap range calculated as the mean Pianka index at 95% confidence interval. The P-values indicate the second tail probability and the standardized effect size (SES) derived from these distributions are indicated.

previous studies (Bihn et al., 2008; Peters et al., 2014). Moreover, different resource limitations between environments may filter different species in relation to their ability to use the available forms of resources (McKane et al., 2002; Kaspari et al., 2012; Fowler et al., 2014). In addition, species in forest assemblages presented an important overlap in their acquisition strategies at the habitat level, but also occupied a large niche space indicating an important diversity of strategies in this habitat. This contrasted with cropland ants, for which niche overlap was not found and the niche space occupied was quite small. These findings may be explained by the denser litter and plant diversity found in forest. Microhabitat heterogeneity provides a wide range of spatio-temporal foraging possibilities not present in croplands, enabling niche partitioning and selection of different foraging strategies by forest ants.

Ants in the studied assemblages were likely structured, at least in part, by competition. This claim is frequently controverted, but we report four findings that together suggest an important effect of biotic filters in ant community structure. In first place, the majority of baits were rapidly exploited, often, by several species, supporting resource limitations, which is the essential driver of competition. Secondly, we confirmed that intraspecific level variations of resource use in different assemblages were unimportant in comparison to interspecific variations within assemblages. This is the most basic assumption of the limiting similarity prediction: species are more limited by conspecifics than by other species (Abrams, 1983). In third place, niche partitioning was the rule in the studied assemblages, in at least one of the two ecological dimensions considered (trophic partitioning and foraging strategies). Finally, the finding of interspecific niche differences is no proof of competition, so it must be evidenced that partitioning indeed relates to community structure (Connell, 1980; Cadotte and Tucker, 2017). Congruently, we found that when significant niche overlap was found at the habitat level, the average values of niche overlap were lower at the assemblage level. This finding suggest that local assemblages more likely contained subsets of species foraging for different resources, among the set of possible species from the habitat pool suitable for that habitat type (Levine and Hille Ris Lambers, 2009; Kraft and Ackerly, 2010). Likewise, resource dominance in the different local assemblages was shared by species belonging to a diversity of genera. This is important, because phylogenetic niche conservatism is an important factor explaining habitat filtering and niche partitioning (Andersen, 1995; Losos, 2008; Salas-López, 2017). In forests, every ant assemblage contained *Ochetomyrmex neopolitus*, *Wasmannia auropunctata*, and a *Solenopsis* species, as well as one or two dominant *Pheidole* species and a dominant *Crematogaster* species. Likewise, three ant species dominated croplands: *Solenopsis saevissima*, *Pheidole* (cf. fallax), and a *Crematogaster* species. Other traits such as the number of workers and soldiers captured were recorded but will be studied elsewhere in addition to trophic and morphological traits.

Another striking finding was that habitat filtering and niche partitioning were more intense for the same combination of habitat and ecological dimension. For example, the greatest overlap was found in acquisition strategies of forest ants at the habitat level, suggesting an important competition for some strategies. This finding contrasted however, with the absence of overlap in the majority of forest assemblages, together with the finding of niche partitioning in acquisition strategies in all of these assemblages. Additionally, by partitioning on acquisition strategies, forest dominant ants may be less constrained by their trophic similarity than cropland ants, as was also reported in a previous study (Jacquemin et al., 2012). Furthermore, neither niche overlap nor partitioning was found in acquisition strategies in croplands, suggesting that in general, species were unspecialized to use a particular resource or another. These findings suggest that interspecific niche partitioning was more important when important fractions of resources were shared among species. Our explanation is that, if adaptation to a specific set of conditions is important for the survival of species within a specific habitat, it is likely that related functional traits, will be selected, promoting species filtering and, therefore, niche overlap. Yet, competition may select for the maximal differences within the range of conditions and/or resources available in that habitat. This is supported by the central place that dominant species held in the PCAs compared to the overall niche space, and yet their important interspecific differences. Another related finding was that, despite important differences in species richness between habitats, the average number of species at baits was similar. We believe that for a given area, even if the number of species is greater in forest habitats, the competitive pressure for limiting resources can only allow a certain number of ant competitors to stand the presence of other species at a given bait. We have no quantitative evidence to confirm this, but we suggest that future studies reporting the number of ant species discovering a bait in the lapse of an hour would greatly help us to understand this pattern.

In combination, the presented results provide a strong support of the relevance of habitat filtering and competition in structuring ant assemblages. Resource availabilities in different habitats may have resulted in the filtering of species with similar ecological traits, but at the assemblage level, limiting similarity may result in non-overlapping or, at least, less overlapping assemblages (Ackerly and Cornwell, 2007; Algar et al., 2011). Finally, we could observe, a functional homogenization in agricultural-forest gradients as recurrent in many studies concerning ants and other organisms (Bihn et al., 2010; Clavel et al., 2011; Peters et al., 2014).

CONCLUSION

We provide evidence of the relevance of habitat and niche differentiation in resource use patterns of species and their interplay in community structure. Our study joins more and more studies providing similar evidence (Ackerly and Cornwell, 2007; Gilbert et al., 2008; Kraft and Ackerly, 2010; Algar et al., 2011). Niche overlap and partitioning are simultaneous covariant forces of community structure, since species must compete to exploit the highest quality resources within their habitats while preventing the depletion of identical sources. For a given habitat, niche overlap is expected in all its extension, but the limits of such overlap depend on the competitive

exclusion of redundant competitors in local assemblages. We thus recommend the separation of overlap and partitioning hypotheses in different ecological dimensions while trying to understand assembly processes, as well as in considering the resource limitations in different environments (Fox and Vasseur, 2008; Correa and Winemiller, 2014).

DATA AVAILABILITY STATEMENT

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

AS-L, JO, and FMe presented the idea. AS-L carried out the experiment, identified the biological material, and wrote the main version of the manuscript. AS-L, CV, and FMu developed the theory and performed the statistical analyses. All authors contributed to the final version of the manuscript.

FUNDING

Financial support for this study was provided by a Ph.D. fellowship from the LabEx CEBA (Centre d'Etude de la Biodiversité Amazonienne) and the Fond Social Européen (FSE) to AS-L, and by an "Investissement d'Avenir" grant managed by the Agence Nationale de la Recherche (CEBA, ref. ANR-10-LABX-25- 01).

ACKNOWLEDGMENTS

We are most grateful to the landowners who generously allowed us to carry out the experiments on their properties, to Julian Frederick, Pierre Freyre, Isabelle Kozon, and Jean-Romain Roussel for field assistance and to Andrea Dejean for proofreading this manuscript.

REFERENCES

Abrams, P. (1983). The theory of limiting similarity. *Annu. Rev. Ecol. Syst.* 14, 359–376. doi: 10.1146/annurev.es.14.110183.002043

Ackerly, D. D., and Cornwell, W. K. (2007). A trait-based approach to community assembly: partitioning of species trait values into within- and among-community components. *Ecol. Lett.* 10, 135–145. doi: 10.1111/j.1461-0248.2006.01006.x doi: 10.1016/j.bcp.2008.05.025

Agarwal, V. M., and Rastogi, N. (2009). Food resource and temporal partitioning amongst a guild of predatory agroecosystem - inhabiting ant species. *Curr. Zool.* 55, 366–375.

Agosti, D., and Alonso, L. E. (2000). "The ALL protocol: a standard protocol for the collection of ground-dwelling ants," in *Ants: Standard Methods for Measuring and Monitoring Biodiversity*, eds T. R. S. D. Agosti and J. D. Majer (Washington DC: Smithsonian Institution Press), 204–206.

Algar, A. C., Kerr, J. T., and Currie, D. J. (2011). Quantifying the importance of regional and local filters for community trait structure in tropical and temperate zones. *Ecology* 92, 903–914. doi: 10.1890/10-0606.1

Andersen, A. N. (1991). "Parallels between ants and plants: implications for community ecology," in *Ant-plant interactions*, eds C. R. Huxley and D. F. Cutler (Oxford: Oxford University Press), 539–553.

Andersen, A. N. (1992). Regulation of "momentary" diversity by dominant species in exceptionally rich ant communities of the Australian seasonal tropics. *Am. Nat.* 140, 401–420. doi: 10.1086/285419

Andersen, A. N. (1995). A classification of australian ant communities, based on functional groups which parallel plant life-forms in relation to stress and disturbance. *J. Biogeogr.* 22, 15–29. doi: 10.2307/2846070

Bihn, J. H., Gebauer, G., and Brandl, R. (2010). Loss of functional diversity of ant assemblages in secondary tropical forests. *Ecology* 91, 782–792. doi: 10.1890/08-1276.1

Bihn, J. H., Verhaagh, M., and Brandl, R. (2008). Ecological stoichiometry along a gradient of forest succession: bait preferences of litter ants. *Biotropica* 40, 597–599. doi: 10.1111/j.1744-7429.2008.00423.x

Blüthgen, N., and Fiedler, K. (2004). Preferences for sugars and amino acids and their conditionality in a diverse nectar-feeding ant community. *J. Anim. Ecol.* 73, 155–166. doi: 10.1111/j.1365-2656.2004.00789.x

Bolton, B. (2003). Synopsis and classification of formicidae. *Memoirs Am. Entomol. Institute* 71, 1–370.

Brandão, C., Silva, R., and Delabie, J. (2012). "Neotropical ants (Hymenoptera) functional groups: nutritional and applied implications," in *Insect Bioecology and Nutrition for Integrated Pest Management*, eds A. R. Panizzi and J. R. P. Parra (Boca Raton, FL: CRC Press, Taylor & Francis group), 213–236. doi: 10.1201/b11713-13

Cadotte, M. W., and Tucker, C. M. (2017). Should environmental filtering be abandoned? *Trends Ecol. Evol.* 32, 429–437. doi: 10.1016/J.TREE.2017.03.004

Cerdá, X., Arnan, X., and Retana, J. (2013). Is competition a significant hallmark of ant (Hymenoptera: Formicidae) ecology? *Myrmecol. News* 18, 131–147.

Cerdá, X., Retana, J., and Cros, S. (1998). Prey size reverses the outcome of interference interactions of scavenger ants. *Oikos* 82, 99–110. doi: 10.2307/3546920

Clavel, J., Julliard, R., and Devictor, V. (2011). Worldwide decline of specialist species: toward a global functional homogenization? *Front. Ecol. Environ.* 9:80216. doi: 10.1890/080216

Connell, J. H. (1980). Diversity and the coevolution of competitors, or the ghost of competition past. *Oikos* 35, 131–138. doi: 10.2307/3544421

Cornwell, W. K., Schwilk, D. W., and Ackerly, D. D. (2006). A trait-based test for habitat filtering: convex hull volume. *Ecology* 87, 1465–1471. doi: 10.1890/0012-9658(2006)87[1465:attfhf]2.0.co;2

Correa, S. B., and Winemiller, K. O. (2014). Niche partitioning among frugivorous fishes in response to fluctuating resources in the Amazonian floodplain forest. *Ecology* 95, 210–224. doi: 10.1890/13-0393.1

Davidson, D. W. (1998). Resource discovery versus resource domination in ants: a functional mechanism for breaking the trade-off. *Ecol. Entomol.* 23, 484–490. doi: 10.1046/j.1365-2311.1998.00145.x

Davidson, D. W., and Patrell-kim, L. (1996). "Tropical arboreal ants: why so abundant?," in *Neotropical Biodiversity and Conservation*, ed. A. C. Gibson (Los Angeles, CA: Mildred E. Mathias Botanical Garden, University of California press), 127–140.

Dejean, A., Corbara, B., Orivel, J., and Leponce, M. (2007). Rainforest canopy ants: the implications of territoriality and predatory behavior. *Funct. Ecosyst. Commun.* 1, 105–120.

Fowler, D., Lessard, J. P., and Sanders, N. J. (2014). Niche filtering rather than partitioning shapes the structure of temperate forest ant communities. *J. Anim. Ecol.* 83, 943–952. doi: 10.1111/1365-2656.12188

Fox, J. W., and Vasseur, D. A. (2008). Character convergence under competition for nutritionally essential resources. *Am. Nat.* 172, 667–680. doi: 10.1086/591689

Gibb, H., and Parr, C. L. (2013). Does structural complexity determine the morphology of assemblages? an experimental test on three continents. *PLoS One* 8:e64005. doi: 10.1371/journal.pone.0064005

Gibb, H., Stoklosa, J., Warton, D. I., Brown, A. M., Andrew, N. R., and Cunningham, S. A. (2015). Does morphology predict trophic position and

habitat use of ant species and assemblages? *Oecologia* 177, 519–531. doi: 10.1007/s00442-014-3101-3109

Gilbert, B., Srivastava, D. S., and Kirby, K. R. (2008). Niche partitioning at multiple scales facilitates coexistence among mosquito larvae. *Oikos* 117, 944–950. doi: 10.1111/j.0030-1299.2008.16300.x

Gotelli, N. J., and Entsminger, G. L. (2001). *EcoSim: Null Models Software for Ecology. Version 7.0*.

Gotelli, N. J., and Graves, G. R. (1996). *Null Models in Ecology*. Washington D.C: Smithsonian Institution Press, doi: 10.1007/BF01199989

Götzenberger, L., de Bello, F., Bråthen, K. A., Davison, J., Dubuis, A., Guisan, A., et al. (2012). Ecological assembly rules in plant communities-approaches, patterns and prospects. *Biol. Rev.* 87, 111–127. doi: 10.1111/j.1469-185X.2011.00187.x

Hahn, D. A., and Wheeler, D. E. (2002). Seasonal foraging activity and bait preferences of ants on Barro Colorado island, Panama. *Biotropica* 34, 348–356. doi: 10.1111/j.1744-7429.2002.tb00548.x

Hammer, Ø, Harper, D. A. T., and Ryan, P. D. (2001). Paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* 4, 9–18.

Hille Ris, Lambers, J., Adler, P. B., Harpole, W. S., Levine, J. M., and Mayfield, M. M. (2012). Rethinking community assembly through the lens of coexistence theory. *Annu. Rev. Ecol. Evol. Syst.* 43, 227–248. doi: 10.1146/annurev-ecolsys-110411-160411

Hölldobler, B., and Wilson, E. O. (1990). *The Ants*. Cambridge: Harvard University Press.

Holway, D. A. (1999). Competitive mechanisms underlying the displacement of native ants by the invasive argentine ant. *Ecology* 80, 238–251.

Houadria, M., Salas-López, A., Bluthgen, N., Orivel, J., and Menzel, F. (2015). Dietary and temporal niche differentiation in in tropical ants - can they explain local tropical ant coexistence? *Biotropica* 47, 208–217.

Ipser, R. M., and Gardner, W. A. (2020). Abundance, activity, diversity, and species interactions of ground-dwelling ants (Hymenoptera: Formicidae) in open versus canopied habitats in Central Georgia (USA). *J. Entomol. Sci.* 55, 469–486. doi: 10.18474/0749-8004-55.4.469

Jacquemin, J., Maraun, M., Roisin, Y., and Leponce, M. (2012). Differential response of ants to nutrient addition in a tropical Brown Food Web. *Soil Biol. Biochem.* 46, 10–17. doi: 10.1016/j.soilbio.2011.11.007

Kaspari, M., and Yanoviak, S. P. (2001). Bait use in tropical litter and canopy ants - evidence of differences in nutrient limitation. *Biotropica* 33, 207–211.

Kaspari, M., Donoso, D., Lucas, J. A., Zumbusch, T., and Kay, A. D. (2012). Using nutritional ecology to predict community structure: a field test in neotropical ants. *Ecosphere* 3, 1–15. doi: 10.1890/ES12-00136.1

Keddy, P. A. (1992). Assembly and response rules: two goals for predictive community ecology. *J. Veg. Sci.* 3, 157–164. doi: 10.2307/3235676

Kraft, N. J. B., and Ackerly, D. D. (2010). Functional trait and phylogenetic tests of community assembly across spatial scales in an Amazonian forest. *Ecol. Monogr.* 80, 401–422. doi: 10.1890/09-1672.1

Lanan, M. (2014). Spatiotemporal resource distribution and foraging strategies of ants (Hymenoptera: Formicidae). *Myrmecol. News* 20, 53–70.

Levine, J. M., and Hille Ris, Lambers, J. (2009). The importance of niches for the maintenance of species diversity. *Nature* 461, 254–257. doi: 10.1038/nature08251

Losos, J. B. (2008). Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecol. Lett.* 11, 995–1003. doi: 10.1111/j.1461-0248.2008.01229.x

Luque, G. M., and Reyes López, J. (2007). Effect of experimental small-scale spatial heterogeneity on resource use of a Mediterranean ground-ant community. *Acta Oecol.* 32, 42–49. doi: 10.1016/j.actao.2007.03.003

MacArthur, R., and Levins, R. (1967). The limiting similarity, convergence, and divergence of coexisting species. *Am. Nat.* 101, 377–385. doi: 10.1086/282505

McCrum-Gardner, E. (2008). Which is the correct statistical test to use? *Br. J. Oral Maxillofac. Surg.* 46, 38–41. doi: 10.1016/J.BJOMS.2007.09.002

McGlynn, T. P., and Kirksey, S. E. (2000). The effects of food presentation and microhabitat upon resource monopoly in a ground-foraging ant (Hymenoptera: Formicidae) community. *Rev. Biol. Trop.* 48, 629–642.

McKane, R. B., Johnson, L. C., Shaver, G. R., Nadelhoffer, K. J., Rastetter, E. B., Fry, B., et al. (2002). Resource-based niches provide a basis for plant species diversity and dominance in arctic tundra. *Nature* 415, 68–71. doi: 10.1038/415068a

Nooten, S. S., Schultheiss, P., Rowe, R. C., Facey, S. L., and Cook, J. M. (2019). Habitat complexity affects functional traits and diversity of ant assemblages in urban green spaces (Hymenoptera: Formicidae). *Myrmecol. News* 29, 67–77. doi: 10.25849/myrmecol.news_029067

Parr, C. L. (2008). Dominant ants can control assemblage species richness in a South African savanna. *J. Anim. Ecol.* 77, 1191–1198. doi: 10.1111/j.1365-2656.2008.01450.x

Parr, C. L., and Gibb, H. (2010). "Competion and the role of dominant ants," in *Ant Ecology*, eds L. Lach, C. Parr, and K. L. Abbott (New York, NY: Oxford University Pess), 77–96. doi: 10.1093/acprof:oso/9780199544639.003.0005

Pearce-Duvet, J. M. C., Elemans, C. P. H., and Feener, D. H. (2011). Walking the line: search behavior and foraging success in ant species. *Behav. Ecol.* 22, 501–509. doi: 10.1093/beheco/arr001

Peters, M. K., Mayr, A., Röder, J., Sanders, N. J., and Steffan-Dewenter, I. (2014). Variation in nutrient use in ant assemblages along an extensive elevational gradient on Mt Kilimanjaro. *J. Biogeogr.* 41, 2245–2255. doi: 10.1111/jbi.12384

Pfeiffer, M., Mezger, D., and Dyckmans, J. (2014). Trophic ecology of tropical leaf litter ants (Hymenoptera: Formicidae) - a stable isotope study in four types of Bornean rain forest. *Myrmecol. News* 19, 31–41.

Pike, N. (2011). Using false discovery rates for multiple comparisons in ecology and evolution. *Methods Ecol. Evol.* 2, 278–282. doi: 10.1111/j.2041-210X.2010.00061.x

Pyke, G. H., Pulliam, H. R., and Charnov, E. L. (1977). Optimal foraging: a selective review of theory and tests. *Q. Rev. Biol.* 52, 137–154.

Salas-López, A. (2017). Predicting resource use in ant species and entire communities by studying their morphological traits: Influence of habitat and subfamily. *Ecol. Indic.* 78, 183–191. doi: 10.1016/j.ecolind.2017.02.034

Salas-López, A., Mickal, H., Menzel, F., and Orivel, J. (2017). Ant-mediated ecosystem processes are driven by trophic community structure but mainly by the environment. *Oecologia* 183, 249–261. doi: 10.1007/s00442-016-3741-z

Salas-López, A., Violle, C., Mallia, L., and Orivel, J. (2018). Land-use change effects on the taxonomic and morphological trait composition of ant communities in French Guiana. *Insect Conserv. Divers.* 11, 162–173. doi: 10.1111/icad.12248

Sanders, N. J., and Gordon, D. M. (2003). Resource-dependent interactions and the organization of desert ant communities. *Ecology* 84, 1024–1031.

Sarty, M., Abbott, K. L., and Lester, P. J. (2006). Habitat complexity facilitates coexistence in a tropical ant community. *Oecologia* 149, 465–473. doi: 10.1007/s00442-006-0453-459

Savolainen, R., and Vepsäläinen, K. (1988). A competition hierarchy among boreal ants: impact on resource partitioning and community structure. *Oikos* 51, 135–155. doi: 10.2307/3565636

Schoener, T. (1983). Field experiments on interspecific competition. *Am. Nat.* 122, 240–285.

Weiher, E., Freund, D., Bunton, T., Stefanski, A., Lee, T., and Bentivenga, S. (2011). Advances, challenges and a developing synthesis of ecological community assembly theory. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 366, 2403–2413.

Wiens, J. A. (1977). On competition and variable environments. *Am. Sci.* 65, 590–597.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership