

Article

Open Access

High-quality reference genomes of swallowtail butterflies provide insights into their coloration evolution

Jin-Wu He^{1,2,#}, Ru Zhang^{2,#}, Jie Yang^{2,#}, Zhou Chang^{1,#}, Li-Xin Zhu^{3,#}, Si-Han Lu^{2,#}, Fei-Ang Xie⁴, Jun-Lai Mao⁴, Zhi-Wei Dong¹, Gui-Chun Liu¹, Ping Hu¹, Yan Dong³, Wen-Ting Wan^{1,2}, Ruo-Ping Zhao¹, Tian-Zhu Xiong⁵, Jorge L. León-Cortés⁶, Chu-Yang Mao¹, Wei Zhang⁷, Shuai Zhan⁸, Jun Li^{1,9}, Lei Chen^{2,*}, Wen Wang^{1,2,10,*}, Xue-Yan Li^{1,*}

¹ State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China

² School of Ecology and Environment, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China

³ Department of Biology and Food Engineering, Chuzhou University, Chuzhou, Anhui 239000, China

⁴ School of Marine Science and Technology, Zhejiang Ocean University, Zhoushan, Zhejiang 316022, China

⁵ Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

⁶ Ecología para la Conservación de la Fauna Silvestre, El Colegio de la Frontera Sur (ECOSUR), Carretera Panamericana y Periférico Sur s/n Barrio María Auxiliadora, San Cristóbal de Las Casas, Chiapas CP 29290, Mexico

⁷ State Key Laboratory of Protein and Plant Gene Research, Peking University, Beijing 100871, China

⁸ CAS Key Laboratory of Insect Developmental and Evolutionary Biology, CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200032, China

⁹ Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, Yunnan 650223, China

¹⁰ Center for Excellence in Animal Evolution and Genetics, Kunming, Yunnan 650223, China

ABSTRACT

Swallowtail butterflies (Papilionidae) are a historically significant butterfly group due to their colorful wing patterns, extensive morphological diversity, and phylogenetically important position as a sister group to all other butterflies and have been widely studied regarding ecological adaption, phylogeny, genetics, and evolution. Notably, they contain a unique class of pigments, i.e., papiliochromes, which contribute to their color diversity and various biological functions such as predator avoidance and mate preference. To date, however, the genomic and genetic basis of

their color diversity and papiliochrome origin in a phylogenetic and evolutionary context remain largely unknown. Here, we obtained high-quality reference genomes of 11 swallowtail butterfly species covering all tribes of Papilioninae and Parnassiinae using long-read sequencing technology. Combined with

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2022 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Received: 09 January 2022; Accepted: 30 March 2022; Online: 31 March 2022

Foundation items: This work was supported by the National Natural Science Foundation of China (31621062 to W.W., 32070482 to X.Y.L.), Chinese Academy of Sciences ("Light of West China" to X.Y.L., XDB13000000 to W.W.), Yunnan Provincial Science and Technology Department (Talent Project of Yunnan: 202105AC160039), and Biodiversity Conservation Program of the Ministry of Ecology and Environment, China (China BON-Butterflies)

#Authors contributed equally to this work

*Corresponding authors, E-mail: chen_lei@nwpu.edu.cn; wwang@mail.kiz.ac.cn; lixu@mail.kiz.ac.cn

previously published butterfly genomes, we obtained robust phylogenetic relationships among tribes, overcoming the challenges of incomplete lineage sorting (ILS) and gene flow. Comprehensive genomic analyses indicated that the evolution of Papilionidae-specific conserved non-exonic elements (PSCNEs) and transcription factor binding sites (TFBSs) of patterning and transporter/cofactor genes, together with the rapid evolution of transporters/cofactors, likely promoted the origin and evolution of papiliochromes. These findings not only provide novel insights into the genomic basis of color diversity, especially papiliochrome origin in swallowtail butterflies, but also provide important data resources for exploring the evolution, ecology, and conservation of butterflies.

Keywords: High-quality reference genome; Swallowtail butterfly tribe; Color evolution; Papiliochromes

INTRODUCTION

Swallowtail butterflies (Papilionidae Latreille, 1802) account for less than 5% (570 species) of all butterfly species (18 000+ described species) (Van Nieuwerkerken et al., 2011a) but are the most historically significant group due to their extensive morphological diversity (especially wing color patterns) (Wallace, 1865) and their phylogenetically important sister position to all other butterflies (Espeland et al., 2018). Studies on Papilionidae have provided fundamental insights into the phylogeny, biogeography, conservation biology, speciation, insect-plant interactions, evolution, and ecology of butterflies (Bonebrake et al., 2010; Condamine et al., 2018; Espeland et al., 2018). The family is distributed worldwide and currently includes 32 genera, seven tribes, and three subfamilies (Häuser et al., 2005; Van Nieuwerkerken et al., 2011b). As the largest subfamily, Papilioninae includes about 490 species worldwide, but with the greatest diversity found in the tropics (Wallace, 1865; Zakharov et al., 2004). The subfamily Parnassiinae includes 80 species, which are mainly distributed in the Palearctic region (Nazari et al., 2007). The monotypic subfamily Baroniinae accommodates one species (*Baronia brevicornis*) endemic to Mexico (Scriber et al., 1995; Tyler et al., 1994). Of note, Batesian mimicry in Papilionidae represents an empirical model of natural selection (Nijhout, 1991). Like all butterflies, their color patterns consist of finely tiled scale mosaics, in which the color of each scale is determined by both pigment and structure (Nijhout, 1990). Scales develop from epidermal cells, with pigments produced through a developmental process that includes patterning (i.e., spatiotemporal positioning of pigments determined by patterning genes) and biochemical synthesis (determined by effector genes) (Matsuoka & Monteiro, 2018; Wittkopp & Beldade, 2009). Patterning genes regulate the distribution of pigments by directly and indirectly activating the expression of effector genes that encode the enzymes and cofactors required for pigment biosynthesis (Wittkopp & Beldade, 2009).

Papiliochromes are a class of pigments uniquely synthesized by Papilionidae (Umebachi, 1985), which produce yellow, orange, and red scales that may be advantageous in avoiding predators and facilitating intraspecific communication (Koch et al., 2000; Umebachi, 1985; Wilts et al., 2012). Although papiliochromes are specific to Papilionidae, their precursors, i.e., intermediates derived from the melanin (tyrosine metabolism) and ommochrome (tryptophan metabolism) pathways and from β -alanine (Li et al., 2015), are ubiquitous in insects. However, the origin and evolution of papiliochromes in this phylogenetically important lineage remain unknown.

High-quality reference genomes not only provide comprehensive evidence for inferring robust phylogeny, but also provide a genomic background to investigate the genetic basis of phenotypic trait innovations (Chen et al., 2019). However, previous studies using different datasets and methods have revealed phylogenetic incongruence among swallowtail butterflies at the species, genus, and tribe level (Allio et al., 2020; Condamine et al., 2012, 2018; Igarashi, 1984; Yagi et al., 1999). Furthermore, all currently available reference genomes of swallowtail butterflies are limited to *Papilio* species and one *Parnassius* butterfly (Iijima et al., 2018; Li et al., 2015; Lu et al., 2019; Nishikawa et al., 2015; Palmer & Kronforst, 2020; Podsiadlowski et al., 2021). Additional reference genomes from representative swallowtail butterfly tribes are required to help clarify the genomic and genetic basis of their adaptive phenotypic traits and diversity in a well-supported phylogenetic context.

Here, we generated high-quality genome assemblies for 11 species representing all tribes in Papilioninae and Parnassiinae (except the monotypic subfamily Baroniinae) using long-read sequencing technology. Combined with published butterfly genomes, we performed comprehensive phylogenetic analyses at the tribe level. Comparative genomic analysis indicated that variation in patterning and transporter/cofactor genes and their regulatory elements likely contributed to papiliochrome origin and wing color diversity in swallowtail butterflies. These genomic resources will promote further studies on the molecular basis underlying rapid diversification in butterflies and contribute to butterfly diversity conservation.

MATERIALS AND METHODS

Butterflies

The 11 swallowtail butterfly species covered all tribes from the two main subfamilies (Parnassiinae and Papilioninae) of Papilionidae. The monotypic subfamily Baroniinae was excluded due to specimen inaccessibility. Live adults were collected from natural populations and returned to the laboratory in plastic bags. For each individual, wings were dissected and kept as voucher specimens. The body samples were flash-frozen in liquid nitrogen and stored at -80°C until use. For each species, one individual was used for Illumina sequencing and another individual for long-read sequencing. Individuals of the same sex were used for both Illumina and long-reads sequencing, except for two cases due to difficulty in collecting the same sex in the field. Details on all species are listed in Supplementary Tables S1, S2. In addition, at least

one dried specimen was retained as a voucher specimen for each sequenced species. All voucher specimens were deposited in the Kunming Institute of Zoology, Chinese Academy of Sciences, China.

Genome sequencing, assembly, and annotation

All species were sequenced to obtain long reads using PacBio SMRT or Nanopore PromethION technology and to obtain short reads using Illumina next-generation methods. Assembly was performed with WTDBG v1.2.8 (Ruan & Li, 2020), and evaluated using three methods, i.e., comparison of estimated and actual assembled genome sizes, BUSCO (Benchmarking Universal Single-Copy Orthologs) analysis, and mapping ratio calculations of both Illumina and long reads. Repeat elements were annotated by combining the results of Tandem Repeat Finder v4.07b (Benson, 1999), RepeatMasker v4.0.5, and RepeatModeler v1.0.4 (Smit et al., 2015) based on Repbase TE library v16.02 (Bao et al., 2015) and LTR_FINDER v1.0.6 (Xu & Wang, 2007). Genes were *de novo* annotated using SNAP (Korf, 2004), GENSCAN v1.0 (Burge & Karlin, 1997), GlimmerHMM v3.0.3 (Majoros et al., 2004), and AUGUSTUS v2.5.5 (Stanke et al., 2006), combined with homology-based analysis using TBLASTN v2.2.26 (Altschul et al., 1997). EvidenceModeler v1.1.1 (Haas et al., 2008) was used to merge *de novo* and homology gene sets into comprehensive, non-redundant gene sets.

Phylogenetic analyses

To investigate Papilionidae phylogeny, especially at the tribe level, 17 Papilionidae species covering all tribes of Parnassiinae and Papilioninae, including the 11 genomes sequenced in this study and six previously published genomes, were included in phylogenetic analysis (Supplementary Table S2). *Kallima inachus* (Nymphalidae) (Yang et al., 2020a) was used as an outgroup. The phylogenetic trees were constructed using RAxML v8.2.10 (Stamatakis, 2014) based on different datasets, including multiple whole-genome alignments (WGAs), orthologous genes and their codon partitioned datasets, and conserved non-exonic elements (CNEs). Multiple WGAs were performed using the Cactus package with default parameters (<https://github.com/ComparativeGenomicsToolkit/cactus>) (Armstrong et al., 2020). ASTRAL-III v5.6.3 (Zhang et al., 2018) was used to perform coalescent species tree estimation. DiscoVista v1.0 (Sayyari et al., 2018) was used to quantify and visualize gene tree discordance for alternative topologies.

Hybridization inference and incomplete lineage sorting (ILS) simulation

To investigate hybridization inference and the effect of ILS on incongruent phylogeny, a representative species was selected in each Papilionidae tribe, with *K. inachus* used as an outgroup for analysis. Hybridization inference among tribes was examined using PhyloNet v3.6.8 (Wen et al., 2018), PhyloNetworks (Solís-Lemus et al., 2017), and ABBA-BABA test (*D*-statistics) using the “qpDstat” command in the AdmixTools v1.01 (Zou & Zhang, 2015). The ILS simulation was performed following Wang et al. (2018) under the multispecies coalescent model.

Genomic features related to swallowtail butterfly evolution

PhyloFit v1.4 (Hubisz et al., 2011) and phastCons v1.4 (Siepel et al., 2005) were used to infer CNEs in the swallowtail butterflies. Positively selected genes (PSGs) and rapidly evolving genes (REGs) were identified using branch and branch site models in the PAML package (Yang, 2007). Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of the above genes/gene families was performed using the KEGG Orthology-Based Annotation System (KOBAS) v3.0 (Wu et al., 2006) or the Omicshare platform (<http://www.omicshare.com/tools/>).

Genomic variations related to swallowtail butterfly pigmentation

We investigated genomic variations in those genes related to pigment position in space and time (patterning genes) and to biochemical synthesis of pigments (effector genes) in swallowtail butterflies. BLASTP homologous searches (E -value < 1e-5) and conserved domain scanning in HMMER v3.2.1 (Finn et al., 2015) were performed against the reference genomes of multiple butterfly species (17 swallowtail butterflies, dead-leaf butterfly (*K. inachus*), and three moths (*Trichoplusia ni*, *Hyposmocoma kahamanoa*, and *Cydia pomonella*) to identify candidate patterning and effector genes. In addition, gene trees were constructed using RAxML v8.2.10 (Stamatakis, 2014) and homologous searches in the NCBI protein database were used to exclude those candidate genes still in doubt. Finally, the evolutionary features of the obtained candidate genes/gene families were associated with the above results, including Papilionidae-specific CNEs (PSCNEs), PSGs, and REGs. Specifically, Papilionidae-specific transcription factor binding sites (TFBSs) in the upstream regulatory regions of the genes were predicted using JASPAR 2020 (http://jaspar.genereg.net/search?q=&collection=CORE&tax_group=insects) (Fornes et al., 2020). Cytoscape v3.8.0 (Shannon et al., 2003) was used to integrate gene relationships (i.e., patterning genes, effector genes, and predicted TFs) into a network.

Details on the methods mentioned above are described in the Supplementary Information.

RESULTS

Genome sequencing, assembly, and annotation

We *de novo* assembled the genomes of 11 species representing all tribes in Papilioninae and Parnassiinae (excluding the monotypic subfamily Baroniinae) (Supplementary Table S2) using long-reads (Nanopore/PacBio: ~56–174× coverage) and short-reads (Illumina: ~75–203× coverage) sequencing technologies (Supplementary Table S1). The sizes of the assembled genomes varied from 240 Mb (Papilioninae: Papilionini: *Papilio demoleus*) to 1176 Mb (Parnassiinae: Parnassini: *Parnassius orleans*), consistent with their estimated genome sizes using *k*-mer (Table 1) and C-value analysis (Liu et al., 2020). The long scaffolds (N50: 2.3–14.8 Mb) and high BUSCO scores (89.9%–98.1%) suggested high contiguity and completeness of all assemblies. After masking repeats

Table 1 Genome assembly and quality estimation of swallowtail butterflies

Subfamily	Tribe	Species	K-mer	Genome assembly			Reads mapping ratio (%)	
			Estimated size (Mb)	Assembly size (Mb)	Scaffold N50 (Mb)	Complete BUSCO ratio (%)	Illumina reads	Long reads
Parnassiinae	Parnassini	<i>Parnassius orleans</i>	1 115	1 176	3.2	92.0	98.3	96
		<i>Luehdorfia chinensis</i>	678	656	2.3	90.4	98.9	93
	Zerynthini	<i>Sericinus montelus</i>	571	567	5.5	95.8	97.5	88.4
		<i>Bhutanitis thaidina</i>	449	449	14.8	98	99.1	99
Papilioninae	Leptocircini	<i>Lamproptera curius</i>	550	595	2.5	89.9	95.8	85.8
	Teinopalpini	<i>Teinopalpus imperialis</i>	535	504	12.5	95.8	97.1	93.0
	Troidini	<i>Troides helena</i>	336	330	10.5	96.6	97.4	78.1
		<i>Byasa hedistus</i>	271	288	9.2	98.1	93.4	98.5
		<i>Papilio demoleus</i>	240	240	9.1	97.5	97.8	99.0
	Papilionini	<i>Meandrusa payeni</i>	406	392	12.5	96.3	98.8	97.5
		<i>Papilio protenor</i>	214	246	5.4	97.4	97.7	98.7

(Supplementary Table S3: 30.5%–64.8%), we used both *de novo*- and homology-based gene prediction to annotate the genomes. The final annotated genes among swallowtail butterflies varied from 12956 to 19416 (Supplementary Table S3). The gene structures were similar to each other (Supplementary Figure S1 and Table S3), indicating that our annotations were fairly comprehensive. In addition, our results indicated that *Parnassius orleans*, which lives in high altitude habitat, possessed the largest genome (1176 Mb) in Papilionidae, which is consistent with that of another *Parnassius* butterfly (*P. apollo*) (1392 Mb) (Podsiadlowski et al., 2021).

High-quality genomic data revealed a well-supported swallowtail butterfly phylogeny

Phylogeny is fundamental for understanding the evolution of the diverse phenotypes of swallowtail butterflies. High-quality genomic data have unmatched advantages in phylogenetic analysis (Chen et al., 2019; Lü et al., 2021; Yuan et al., 2021). In this study, we reconstructed the phylogenetic trees for 17 swallowtail butterflies covering all tribes of Papilionidae, except for the monotypic subfamily Baroniinae. We first established a WGA tree with RAxML v8.2.10 (Stamatakis, 2014) using *K. inachus* as an outgroup. In total, ~45 Mb of orthologous syntenic sequences were obtained from the WGA, yielding a whole-genome tree with 100% bootstrap support for all nodes (Figure 1A). We then performed phylogenetic analyses with different genomic partitions, including the amino acid sequences of 12 274 one-to-one orthologous protein-coding genes and their derived sequences (four-fold degenerate (4d) sites, all concatenated codons, third position only (3rd), first-second positions only (1st-2nd), third and first-second positions (1st-2nd+3rd)), and CNEs. Aside from codon position (1st-2nd and 1st-2nd+3rd) (Supplementary Figure S2B), all other tree topologies (Supplementary Figure S2A) were identical to that of the WGA tree (Figure 1A), consistent with the tree based on 6 621 orthologs (Allio et al., 2020). The discordant tree based on codon site partitions (Supplementary Figure S2B) was observed in a previous butterfly phylogenetic tree based on 350 genes (Espeland et al., 2018), which may result from high-base compositional heterogeneity of codon-positions in

protein-coding genes (Jarvis et al., 2014).

We further assessed genome-wide tree heterogeneity by randomly extracting 2 000 genomic 5 kb windows from WGAs to generate window-based gene trees (WGTs). Phylogenetic discordance at the tribe level was also pervasive across genomic regions (Figure 1B). Accounting for WGT discordance, we performed a coalescent-based method for inferring species trees from WGTs with ASTRAL-III v5.6.3. The results (Supplementary Figure S2A) were identical to the WGA tree, thus supporting the robustness of the inferred WGA phylogeny of the swallowtail butterflies. Nevertheless, the significant differences in quartet frequencies of branch 12 (Leptocircini+all other Papilioninae) in the two alternative topologies (Supplementary Figure S3) suggested that the incongruent positioning of *L. curius* (Leptocircini) may be caused by ILS and hybridization (Sayari et al., 2018). We performed ILS simulation using a multispecies coalescent-based method (Wang et al., 2018). The *F*-statistic ($P < 2.2e-16$), adjusted R^2 (0.791), and DensiTree plot of the topology results (Figure 1B) between the simulated and observed gene trees indicated that ILS contributed to the incongruent placement among tribes. We also performed hybridization inference using two methods. PhyloNet and PhyloNetworks analyses based on the WGTs (Supplementary Figure S4A–E) suggested the occurrence of ancient gene flow between Leptocircini (*L. curius*) and other tribes. The classic ABBA-BABA test (Supplementary Figure S4F, G) also suggested that phylogenetic discordance may be caused by ancient gene flow. Taken together, both ILS and gene flow may have jointly contributed to the observed topological discordances in the swallowtail butterfly phylogeny, similar to the findings reported for *Heliconius* (Edelman et al., 2019).

PSCNEs promote evolution of swallowtail butterflies

Using nine outgroups (Supplementary Table S2), we identified 230 528 CNEs (≥ 20 bp) (15.55 Mb) in Papilionidae (Figure 2A). Among them, 65% (150 029 CNEs, 10.33 Mb) were PSCNEs, including 46.53% Type I PSCNEs (107 265 CNEs, 5.58 Mb), in which sequences were not found in outgroups, and 18.55% Type II PSCNEs (42 764 CNEs, 4.75 Mb), which had orthologous sequences in at least three outgroup species but were only conserved in Papilionidae.

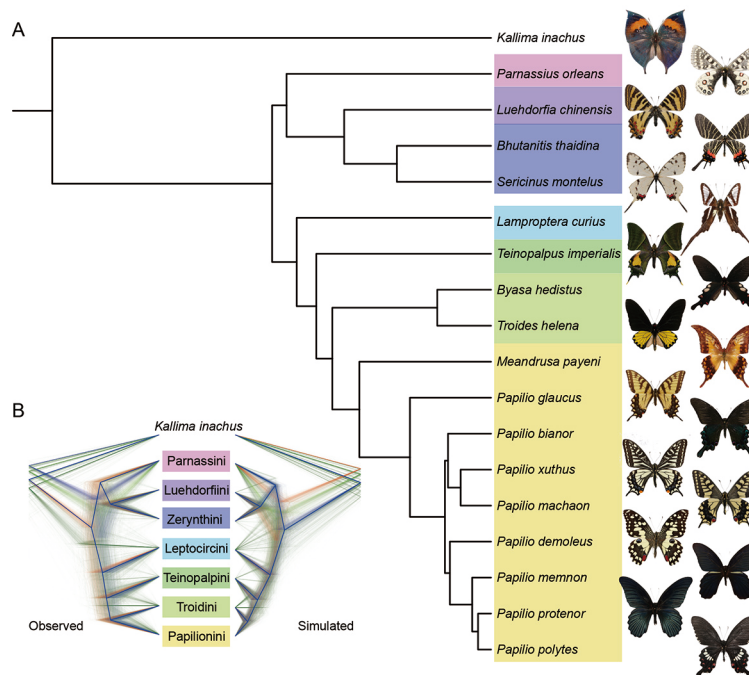


Figure 1 Phylogeny of swallowtail butterflies

A: Maximum-likelihood (ML) phylogenetic tree from ~45 Mb of syntenic whole-genome alignment (WGA) sequences of 17 swallowtail butterfly species (Papilionidae) and *Kallima inachus* (outgroup). One hundred bootstraps were used to compute node support, with all nodes showing 100% support. B: Superimposed ultrametric gene trees in a consensus DensiTree plot. Window-based gene trees (WGTs) and multispecies coalescent simulated gene trees are shown as observed and simulated, respectively.

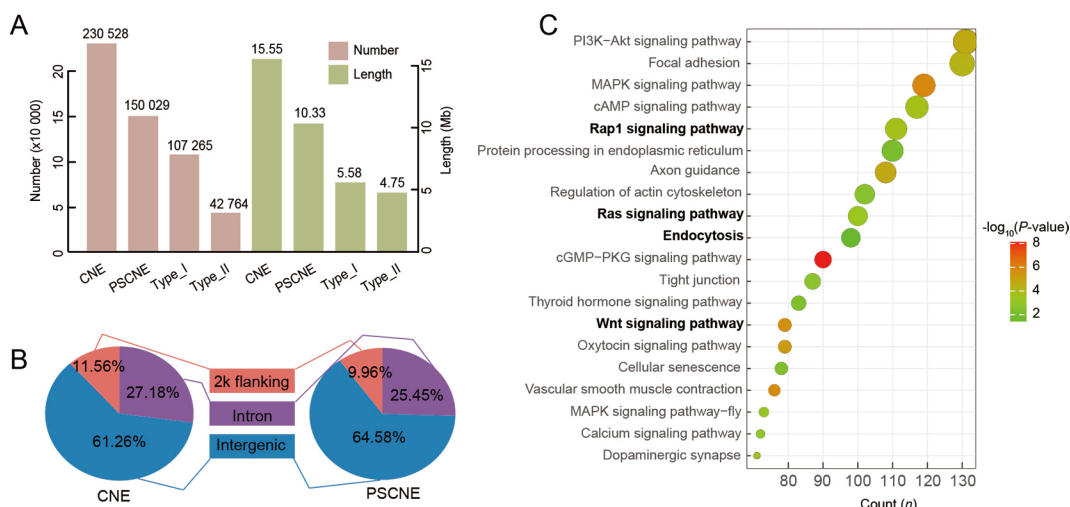


Figure 2 Evolution of conserved non-exonic elements (CNEs) among swallowtail butterfly genomes

A: Length and number of CNEs and Papilionidae-specific CNEs (PSCNEs). B: Location distribution of CNEs and PSCNEs in different genomic regions. C: KEGG enrichment ($P < 0.05$) of neighboring genes physically associated with transcription factor binding sites (TFBSs) in PSCNEs. We selected top 20 KEGG terms ranked by count in PSCNEs. Pathways possibly related to color diversity in butterflies are highlighted in bold.

Both CNEs and PSCNEs were mainly located in the intergenic regions, with ~10% of them observed 2 kb upstream of genes (Figure 2B). Furthermore, 95.26% (142,924 CNEs) of PSCNEs contained the TFBSs (10,248,879) of 585 known transcription factors (TFs) (Supplementary Table S4). The neighboring genes associated with these TFBSs in PSCNEs were enriched in several important signaling pathways (e.g., *Wnt*, *Ras*, and *Rap1* signaling pathways), transport and

catabolism (e.g., endocytosis), and endocrine system (melanogenesis) (Figure 2C; Supplementary Table S5). *Wnt* signaling underlies the evolution and development of butterfly wing pattern diversity (Martin et al., 2012; Martin & Reed, 2014). Both *Rap* and *Ras* are members of the small GTPase family and play important roles in the morphogenesis of fruit flies (Mishra et al., 2005; Reiner & Lundquist, 2018). Taken together, these results suggest that PSCNEs and associated

genes may play important roles in trait evolution of swallowtail butterflies, especially morphological evolution.

Evolution of genes and gene families related to unique and diverse traits of swallowtail butterflies

We acquired a high-confidence gene set (7 126 single-copy orthologs) for the 17 swallowtail butterflies and four outgroups (Supplementary Table S2) based on conserved genome synteny (Chen et al., 2019). The lineage-specific evolutionary rates for each branch were run for each ortholog using the CODEML program in the PAML package (Yang, 2007). Results demonstrated that the average evolutionary rate (Ka/Ks) within the swallowtail butterfly species (0.056) was twice as fast as that of the outgroups (0.025) (Supplementary Figure S5A). The optimized branch-site model and branch model in the PAML package were used to detect PSGs and REGs along specific lineages, respectively. In total, 264 PSGs and 402 REGs were identified in the ancestral branch of Papilionidae (Supplementary Figure S5B) and were functionally enriched in several important signal transduction pathways (e.g., Ras, Rap1, and Notch signaling pathways) and membrane transport (e.g., adenosine triphosphate (ATP)-binding cassette (ABC) transporter) (Supplementary Table S6). Members of the small GTPase family, including Rap, Ras, and Rab, and the Notch signaling pathway play important roles during morphological development of the fruit fly eye (Mishra et al., 2005; Sundaram, 2005). ABC transporters (especially ABCG members) and Rab are crucial proteins for

the transport the pigment precursors or related metabolites from the cytoplasm to subcellular pigment granules (Bhuin & Roy, 2014; Lloyd et al., 1998). Combined with the PSCNE and TFBS results mentioned above, these findings suggest that the pathways related to transportation and signaling may be related to the evolution of the unique and diverse traits of swallowtail butterflies.

Lineage-specific evolution of patterning and effector genes may contribute to papiliochrome origin in swallowtail butterflies

We explored the evolution of both *cis*-regulatory and protein-coding regions in patterning genes/gene families and effector genes (enzymes and transportation-related proteins) that may contribute to papiliochrome biosynthesis (Figure 3). Our results (Supplementary Figures S6–S8 and Tables S7–S9) indicated that most patterning genes (e.g., *wingless* (*wg*), *wnt2*, *wnt6*, *wnt10*, *abdominal A* (*abd-A*), *cortex*, *optix*, *aristalless 1 and 2* (*art1/art2*), *engrailed/invented* (*en/inv*), *hairless*) contained PSCNEs in their upstream *cis*-regulatory regions (0–5 kb). Furthermore, Papilionidae-specific TFBSs were also predicted in the PSCNE regions of many genes (e.g., Figure 3: *art1/art2*, *abd-A*, *en/inv*, *optix*, *wg/wnt10*), with each TFBS having a binding site for 1–44 TFs. It is worth mentioning that *en/inv* are highly conserved in Lepidoptera (Figure 4A). Taking *en* as an example, we identified four PSCNEs in its upstream region (0–5 kb) and six Papilionidae-specific TFBSs (TTATTGAA (*hematopoietically expressed*

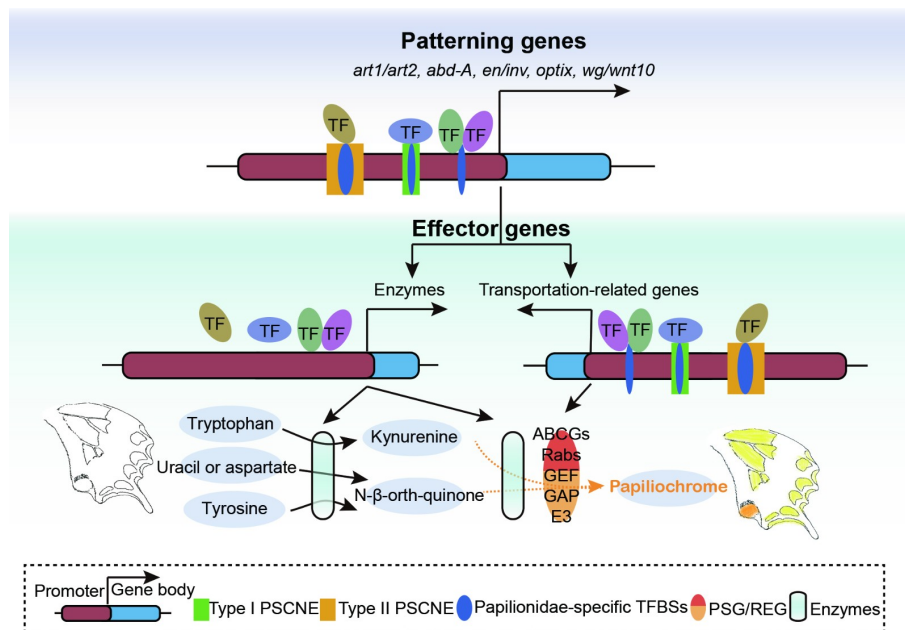


Figure 3 Schematic of papiliochrome biosynthesis

Firstly, spatiotemporal expression of papiliochromes is mediated by different transcription factors (TFs) and patterning genes (above). Effector genes include two major categories. Enzymes (bottom left) that control biochemical synthesis of papiliochrome precursors. Transporters (such as ATP-binding cassette (ABC) transmembrane transporters and small G-proteins (Rab GTPases)) and their cofactors (bottom right) possibly contribute to the transport of pigment precursors to pigment granules, where different pigment precursors meet to combine into papiliochromes. *art1/art2*, two copies of *aristalless*; *abd-A*, *abdominal A*; *en/inv* paralogous genes *engrailed* and *invected*; *wg*, *wingless*; ABCGs, subfamily of ABC; GEF, guanine nucleotide exchange factor; GAP, GTPase-activating protein; E3, E3 ubiquitin ligase.

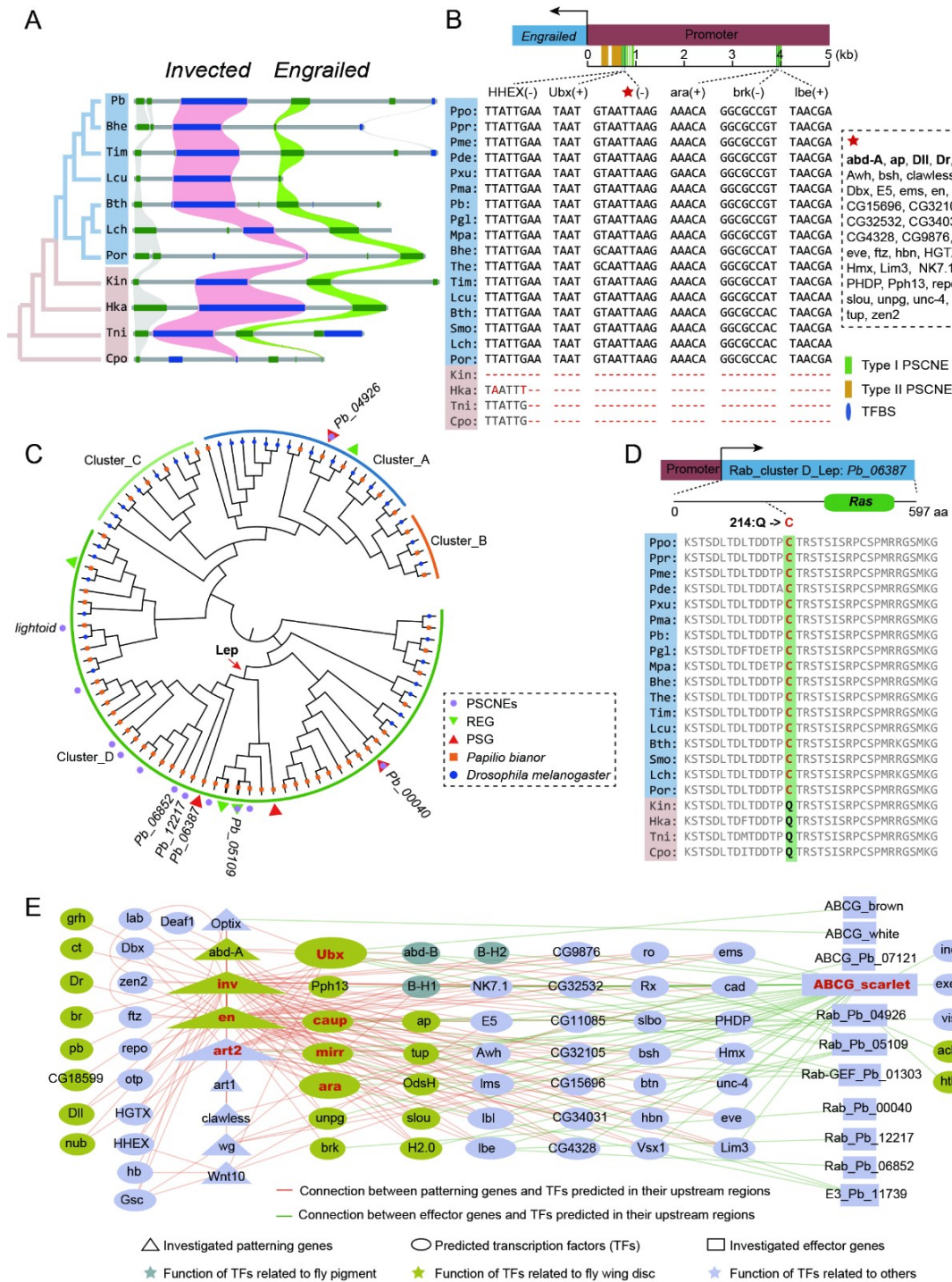


Figure 4 Evolution of papiliochrome biosynthesis-related genes

A: Syntenic relationships of orthologous patterning genes (*en/inv*: *engrailed* and *invected*) among Lepidoptera were identified and visualized using MCScan. Species abbreviations with light blue and light brown backgrounds in tree represent species in Papilionidae (ingroups) and outgroups, respectively (same in panels B and D). B: Evolution of PSCNEs and Papilionidae-conserved TFBSs in upstream region (0–5 kb) of patterning gene (*en*). C: Phylogenetic tree of effector genes (small GTPases: Rab family). *Papilio bianor* and *Drosophila melanogaster* represent whole ingroup species (Papilionidae) and outgroup species, respectively. Genes labeled with gene ID evolved into Papilionidae-specific TFBSs in their upstream regions, or evolved Papilionidae-specific positively selected sites in their coding regions. Red arrow indicates Lepidoptera-specific (Lep) clade. REG, rapidly evolving gene; PSG, positively selected gene. D: Evolution of positively selected effector gene (Rab: *Pb_06387*) in Papilionidae. E: Network of patterning and effector genes in pigmentation and predicted TFs in their upstream regulatory regions. Size of different nodes is related to number of connections among nodes. Hub genes (number of connection nodes exceeding 10) in gene regulatory network are highlighted in red.

homeobox (HHEX)), TAAT (*ultrabithorax* (*Ubx*)), G[T/C]AATTAAG, [G/A]AACA (*araucan* (*ara*)), [A/G][C/T]GGCGCC (*brinker* (*brk*)), and TAAC[G/A]A (*ladybird early* (*lbe*)) in its PSCNEs (Figure 4B; Supplementary Tables S8, S9). Notably, one TFBS (G[T/C]AATTAAG) could bind to 31 TFs, including *abd-A*, *Ubx*, *Distal-less* (*Dll*), *en*, *apterous* (*ap*), and *Drop* (*DrImsh*), which are related to wing disc development in *Drosophila* and butterflies (Gorfinkiel et al., 1997; Halfon, 2017; Milán et al., 2001; Tong et al., 2014; Villa-Cuesta & Modolell, 2005). As a paralogous gene, *inv* also evolved many PSCNEs and Papilionidae-specific TFBSs (Supplementary Figure S6).

We next investigated the effector genes/gene families encoding the enzymes in papiliochrome biosynthesis (Figure 3; Supplementary Tables S7–S9) (Li et al., 2015). Results revealed no positive selection or rapid evolution for any gene encoding the enzymes involved in papiliochrome biosynthesis. Furthermore, except for *black* and *laccase* (*Pb_03719/CG32554* and *Pb_14699*), no PSCNEs or Papilionidae-specific TFBSs were identified in the upstream regulatory region (0–5 kb) of the genes. These findings suggest conserved evolution in both the protein-coding and *cis*-regulatory regions of these enzyme-coding genes. We then investigated those effector genes/gene families that potentially function as transporters/cofactors of pigment precursors (Figure 3; Supplementary Tables S7–S9). Results demonstrated that six ABC genes (Supplementary Figure S9A), four Rab genes (Figure 4C), three guanine nucleotide exchange factor (GEF) genes, one GTPase activating protein (GAP) encoding gene, and eight E3 ubiquitin ligase (E3) encoding genes showed positive selection in Papilionidae (Supplementary Table S8). Among them, Papilionidae-specific positively selected sites were observed in five transporter/cofactor genes (ABCG: *Pb_03311*; Rab: *Pb_06387*; GEF: *Pb_09814*; and E3 E3: *Pb_08975*, and *Pb_09937*) (Figure 4D; Supplementary Figures S9B, S10 and Table S8). In addition, several PSCNEs and/or Papilionidae-specific TFBSs were identified in the upstream regulatory regions of certain ABCG, Rab, GEF, GAP, and E3 genes (Supplementary Figures S9, S11, S12, and Table S8).

We finally explored the relationships among patterning genes, effector genes, and all predicted TFs in their upstream regulatory regions through network analysis. Results suggested that both patterning genes and transporter/cofactor genes formed a complex regulatory network connected via predicted TFs. Three patterning genes (*inv*, *en*, and *art2*) and four predicted TFs (*Ubx* and *iroquois* gene complex (*Iro-C*: *ara*, *mirror* (*mirr*) and *caupolican* (*caup*)) were important hubs in the gene regulatory network (Figure 4E). *Ubx* alters the embryonic body plan and development of eyespots and hindwing patterns in *Bicyclus anynana* butterflies by acquiring different sets of target genes (Matsuoka & Monteiro, 2021; Tong et al., 2014). In addition, *Iro-C* is an essential component of the boundary between the body wall and wings in *Drosophila* (Villa-Cuesta & Modolell, 2005). Four predicted TFs were also connected to many patterning genes (e.g., *inv*, *en*, *art2*, *wnt10*, *optix*, *abd-A*, and *doublesex* (*dsx*)) and some transporters (e.g., ABCGs: *Pb_11403_brown* and *Pb_07121*; Rabs: *Pb_05109* and *Pb_04926*; GEF: *Pb_01303_DENN*; and

E3: *Pb_11739*). Among them, several Rabs (*Pb_05109*, *Pb_12217*, and *Pb_06852*) were located in the Lepidoptera-specific clade (Figure 4C) and all contained PSCNEs and the same TF (*lbe*) in their 0–5 kb upstream regions (Supplementary Figures S11D, S12A, B and Table S8). Intriguingly, *lbe* encodes a TF involved in regulating the embryonic expression of *wg* (Jagla et al., 1994). The latter is associated with color pattern induction in butterfly wing pattern evolution (Martin & Reed, 2010, 2014). Double staining experiments revealed that *ladybird late* (*lbl*) and *lbe* were expressed in cells just anterior to those expressing *en*. We also predicted that *lbl/lbe* was located in the 0–5 kb upstream regions of *en* (Figure 4B). In addition, another Lepidoptera-specific Rab (*Pb_06387*) evolved a Papilionidae-specific positively selected site (Figure 4D: Q>C). Network analysis, together with CNE, REG, and PSG, suggested that the divergent evolution of these patterning and transporter/cofactor effector genes may have contributed to Papilionidae-specific traits such as papiliochrome biosynthesis via involvement in a complex gene regulatory network.

DISCUSSION

The prerequisite for genetic and evolutionary studies is to clarify the historical relationships among organisms. Although there have been many systematic studies of Papilionidae, debate over historical relationships still exists among swallowtail butterflies at the tribe, genus, and species level due to the quality of datasets (morphological, genetic, mitogenomic, transcriptomic, ultra-conserved element, and genomic data) or limited key taxon sampling (Allio et al., 2020; Condamine et al., 2012, 2018; Igarashi, 1984; Yagi et al., 1999). In this study, we dissected the reference genomes of 11 swallowtail butterfly species representing all tribes in Papilioninae and Parnassiinae, except subfamily Baroniinae, using long-reads sequencing technology. Including six previously published *Papilio* genomes (Cong et al., 2015; Iijima et al., 2018; Li et al., 2015; Lu et al., 2019), 17 reference genomes were used to infer the phylogeny of Papilionidae and investigate the genomic and genetic basis of their phenotypic trait innovations. Based on comprehensive datasets of the high-quality reference genomes, our results supported Leptocircini as a sister clade to the remaining Papilioninae, confirming that Papilioninae is monophyletic (Allio et al., 2020). To exclude base-bias of codon positions (1st-2nd and 1st-2nd+3rd) (Supplementary Figure S2B), we generated 2 000 WGT trees. Among them, the true topology ratio (species tree) accounted for only 17% and the inconsistent Leptocircini relationship ratio accounted for 13%. Apart from phylogenetic estimation error, the many anomalous gene trees revealed that swallowtail butterflies have likely undergone complicated evolutionary processes. The topology frequencies based on DiscoVista (Supplementary Figure S3), ILS simulation (Figure 1B), PhyloNet and PhyloNetworks (Supplementary Figure S4A–E), and classic ABBA-BABA analysis results (Figure 4F, G) suggested that both ILS and gene flow may have jointly contributed to the observed topological discordances in the phylogeny of swallowtail butterflies. Insufficient sampling may also have impacted the discordant

phylogenetic relationships of Leptocircini (Allio et al., 2020). Moreover, deep phylogenetic incongruences between gene and species trees are pervasive in both plants and animals (Wang et al., 2018; Yang et al., 2020b). The significance of such events (ILS and gene flow) in the evolution of swallowtail butterflies is still unknown, but similar events have been reported as ecologically adaptive (Jiao et al., 2020; Mallet et al., 2016). For example, closely related *Heliconius* species promiscuously exchange protective color-pattern genes to mimic adaptive radiation and speciation (Edelman et al., 2019).

Butterfly wing patterns are an ideal model in the field of evolutionary developmental biology (Beldade & Brakefield, 2002; Oxford & Gillespie, 1998). An important tenet of evolutionary developmental biology is that adaptive mutations affecting morphology are more likely to occur in the *cis*-regulatory regions of the pleiotropic developmental regulatory loci and in the target genes within the vast networks they control than in the protein-coding regions of genes (Carroll, 2008; Hoekstra & Coyne, 2007), and conserved non-exonic genomic sequences usually function in *cis* regulation of neighboring genes (Nelson & Wardle, 2013; Polychronopoulos et al., 2017; Vavouri et al., 2007). Our high-quality reference genomes provide feasibility and reliability for studying genome-wide CNEs in a comparative genomic context. Many PSCNEs that contain Papilionidae-specific TFBSs were observed in the swallowtail butterfly genomes, suggesting that these CNEs have contributed to adaptive evolution. Based on our findings in Papilionidae, we anticipate that more comprehensive investigations on genome-wide CNEs of more representative butterfly taxa in other families will contribute to our understanding of butterfly diversity evolution.

Papiliochromes are a unique class of pigments in swallowtail butterflies (Umebachi, 1985), despite the ubiquity of their precursors, i.e., dopamine, β -alanine, and kynurenine (Li et al., 2015). Transportation of pigment precursors or related metabolites from the cytoplasm to subcellular pigment granules is a key step in insect pigmentation and is usually executed via transporter proteins, such as ABC transmembrane transporters (especially ABCG members) and small G-proteins (e.g., Rabs) (Bhuin & Roy, 2014; Lloyd et al., 1998) and their cofactors (GEFs and E3s). The activation and inactivation of Rab is tightly controlled by specific GEFs, which activate proteins by promoting guanosine diphosphate for guanosine triphosphate exchange, and by GAPs, which cause inactivation by enhancing low intrinsic GTPase activity (Bernards, 2003; Ishida et al., 2016). Abundant E3s guarantee specific substrate recognition and activate GEFs as coordinators of vesicle traffic (Hochrainer et al., 2005; Lamber et al., 2019). Thus, we hypothesize that the evolution of transporter proteins and/or their regulatory network may play key roles in papiliochrome origin. Interestingly, our comparative genomic analyses identified lineage-specific evolution of some transportation-related proteins in both their protein-coding regions and upstream regulatory regions (Supplementary Figures S9–S12 and Tables S7–S9). With this hypothesis in mind and encouraged by our comparative genomic analysis results, we focused on the evolution of certain representative patterning genes, enzyme-coding genes

related to Papiliochrome biosynthesis, and transporter/cofactor genes. Our results indicated that most patterning genes contained PSCNEs in their upstream regulatory regions (0–5 kb), and Papilionidae-specific TFBSs were also predicted in these PSCNE regions, e.g., *wg*, *wnt10*, *abd-A*, *optix*, *art1*, *art2*, *en*, and *inv*. *Wg* and *wnt10* are the main members of the *wg* signaling pathway, which plays multiple roles in the regulation of tissue growth, polarity, and patterning (Swarup & Verheyen, 2012). *Abd-A*, together with two other homeobox genes (*abd-B* and *clawless*), plays an important role in camouflage color pre-patterning in *Papilio xuthus* (Jin et al., 2019). *Optix* and *aristaleless* (*art1* and *art2*), two major-effect Mendelian loci, are reused in butterfly mimicry polymorphisms (Vankuren et al., 2019). *En/inv* also participates in mimicry diversification in *Papilio dardanus* (Timmermans et al., 2020). Notably, in the current study, three patterning genes (i.e., *inv*, *en*, and *art2*) occupied important hubs in the gene regulatory network (Figure 4E). Thus, we speculate that diverse patterning genes contribute to the diverse color evolution in swallowtail butterflies. The PSCNEs and Papilionidae-specific TFBSs in their upstream regulatory regions revealed that Papilionidae-specific regulation in patterning genes may contribute to the convergent evolution of color, such as papiliochromes.

We also identified PSCNEs and Papilionidae-specific TFBSs in the upstream *cis*-regulatory regions (0–5 kb) of some ABCG transporters (e.g., ABCGs: *Pb_11403_brown* and *Pb_07121*; Rab transporters (Lepidoptera-specific clade): *Pb_05109* and *Pb_04926*; a well-known pigment transporter in the ommochrome pathway: *lightoid*) and their cofactor genes (GEF: *Pb_01303_DENN*; E3: *Pb_11739*) (Supplementary Table S8). Furthermore, our results indicated that one ABCG transporter (*Pb_03311_white*), one Rab transporter (*Pb_06387*), and their cofactors (GEF: *Pb_09814*; E3: *Pb_08975* and *Pb_09937*) have evolved to be Papilionidae-specific positively selected sites (Figure 4D; Supplementary Figures S9B, S10 and Table S8). Among these transporters/cofactors, *white*, *brown*, and *scarlet* are well-studied ABC members and are involved in the uptake of pigment precursors in the ommochrome and pteridine pathways in the development cells of Malpighian tubules and compound eyes (Tatematsu et al., 2011). In addition, *lightoid* plays an important role in eye color via participating in the biogenesis and degradation of pigment granules (Ma et al., 2004). Although our previous gene-editing experiments suggest that these four transporters do not contribute to papiliochromes in *P. xuthus*, they play a key role in body color mimicry in swallowtail butterfly larvae (Liu et al., 2021). These findings also suggest that other transporters (e.g., Lepidoptera-specific Rab transporter) and their cofactors may play key roles in papiliochrome origin and evolution. Thus, we deduce that the PSCNEs and Papilionidae-specific TFBSs of patterning genes may have ensured the spatiotemporal localization of papiliochromes, while PSCNEs, Papilionidae-specific TFBSs, rapid evolution, and positive selection of transporters/cofactors such as Rabs, GEFs, and E3 may have promoted the transportation of papiliochrome precursors into pigment granules, thus promoting the origin and evolution of papiliochromes.

CONCLUSIONS

We obtained high-quality reference genomes for 11 swallowtail butterflies covering all tribes of Papilioninae and Parnassiinae, thus providing important data resources for exploring the genetic basis of morphological diversity in a comprehensive phylogenetic context. We reconstructed a robust phylogeny among tribes in the presence of ILS and gene flow. We also observed that specific mutations in both protein-coding regions and regulatory elements along the Papilionidae lineage may have contributed to morphological diversity in swallowtail butterflies, especially their wing color patterns. Furthermore, we found that PSCNEs and Papilionidae-specific TFBSs of patterning and transporter/cofactor genes as well as the rapid evolution and positive selection of certain transporters/cofactors probably promoted the origin and evolution of papiliochromes in swallowtail butterflies. These findings suggest that both the *cis*-regulatory and protein-coding regions play important roles in morphological evolution, with changes in the former more often observed in the patterning genes of the upstream regulatory network and changes in the latter more often found in the effector genes of the downstream regulatory network. These results provide a reference for future functional studies on coloration genetics and evolution in swallowtail butterflies.

DATA AVAILABILITY

All data on the 11 swallowtail butterflies produced from our Butterfly Genome Project were submitted to NCBI (BioProjectID PRJNA714807). Illumina short reads were submitted to SRA with accession Nos.: SRR14205975–SRR14205981, SRR14205983, SRR14205993–SRR14205995. Nanopore/PacBio long reads were submitted to SRA with accession Nos.: SRR14205982, SRR14205984–SRR14205992, SRR14209577. Genome assemblies were deposited in NCBI under accession Nos. JAGSM[P-Z]000000000 and Science Data Bank (<http://www.scidb.cn/cstr/31253.11.sciencedb.o00023.00001>) and GSA under accession Nos.: GWHBHRU000000000, GWHBHRV000000000, GWHBHSA000000000, GWHBHSB000000000, GWHBHSC000000000, GWHBHSD000000000, GWHBHSE000000000, GWHBHSF000000000, GWHBHSG000000000, GWHBHS000000000, GWHBHSI000000000.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

X.Y.L. and W.W. conceived the study, designed the scientific objectives, and led the project and manuscript preparation. L.C. supervised the data analyses and helped with the manuscript. J.L.M., F.A.X., and J.Y. conducted genome assembly and annotation. J.W.H., R.Z., and J.Y. performed

phylogenetic analyses. R.Z., J.Y., and J.W.H. performed evolutionary analyses. R.Z. and J.W.H. performed CNE analyses. J.W.H., X.Y.L., P.H., and J.L. performed analyses on color-related genomic features. Z.C., Z.W.D., L.X.Z., Y.D., G.C.L., W.T.W., and Z.R.Z. collected and raised butterflies and prepared DNA samples for genomic sequencing. W.Z., L.Z., C.Y.M., and S.Z. helped in data analyses and manuscript improvement. T.Z.X., J.L., and L.C. helped in designing study materials and manuscript improvement. X.Y.L., J.W.H., W.W., R.Z., J.Y., S.H.L., Z.C., and L.X.Z. wrote the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We would like to thank Ze-Shan Lin and Chun-Yan Chen for help with data analyses, Gao Chen for discussions on swallowtail butterfly biology, and Yan-Qiong Peng for helping in sample collection.

REFERENCES

- Allio R, Scornavacca C, Nabholz B, Clamens AL, Sperling FA, Condamine FL. 2020. Whole genome shotgun phylogenomics resolves the pattern and timing of swallowtail butterfly evolution. *Systematic Biology*, **69**(1): 38–60.
- Altschul SF, Madden TL, Schäffer AA, Zhang JH, Zhang Z, Miller W, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**(17): 3389–3402.
- Armstrong J, Hickey G, Diekhans M, Fiddes IT, Novak AM, Deran A, et al. 2020. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature*, **587**(7833): 246–251.
- Bao WD, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, **6**(1): 11.
- Beldade P, Brakefield PM. 2002. The genetics and evo-devo of butterfly wing patterns. *Nature Reviews Genetics*, **3**(6): 442–452.
- Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, **27**(2): 573–580.
- Bernards A. 2003. GAPs galore! A survey of putative Ras superfamily GTPase activating proteins in man and *Drosophila*. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, **1603**(2): 47–82.
- Bhuin T, Roy JK. 2014. Rab proteins: the key regulators of intracellular vesicle transport. *Experimental Cell Research*, **328**(1): 1–19.
- Bonebrake TC, Ponisio LC, Boggs CL, Ehrlich PR. 2010. More than just indicators: a review of tropical butterfly ecology and conservation. *Biological Conservation*, **143**(8): 1831–1841.
- Burge C, Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology*, **268**(1): 78–94.
- Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell*, **134**(1): 25–36.
- Chen L, Qiu Q, Jiang Y, Wang K, Lin ZS, Li ZP, et al. 2019. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science*, **364**(6446): eaav6202.
- Condamine FL, Nabholz B, Clamens AL, Dupuis JR, Sperling FAH. 2018. Mitochondrial phylogenomics, the origin of swallowtail butterflies, and the impact of the number of clocks in Bayesian molecular dating. *Systematic Entomology*, **43**(3): 460–480.
- Condamine FL, Sperling FAH, Wahlberg N, Rasplus JY, Kergoat GJ. 2012. What causes latitudinal gradients in species diversity? Evolutionary processes and ecological constraints on swallowtail biodiversity. *Ecology*

Letters, **15**(3): 267–277.

- Cong Q, Borek D, Otwinowski Z, Grishin NV. 2015. Tiger swallowtail genome reveals mechanisms for speciation and caterpillar chemical defense. *Cell Reports*, **10**(6): 910–919.
- Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, et al. 2019. Genomic architecture and introgression shape a butterfly radiation. *Science*, **366**(6465): 594–599.
- Espeland M, Breinholt J, Willmott KR, Warren AD, Vila R, Toussaint EFA, et al. 2018. A comprehensive and dated phylogenomic analysis of butterflies. *Current Biology*, **28**(5): 770–778.
- Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, et al. 2015. HMMER web server: 2015 update. *Nucleic Acids Research*, **43**(W1): W30–W38.
- Fomes O, Castro-Mondragon JA, Khan A, Van Der Lee R, Zhang X, Richmond PA, et al. 2020. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*, **48**(D1): D87–D92.
- Gorfinkiel N, Morata G, Guerrero I. 1997. The homeobox gene *Distal-less* induces ventral appendage development in *Drosophila*. *Genes & Development*, **11**(17): 2259–2271.
- Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. 2008. Automated eukaryotic gene structure annotation using EvidenceModeler and the program to assemble spliced alignments. *Genome Biology*, **9**(1): R7.
- Halfon MS. 2017. Perspectives on gene regulatory network evolution. *Trends in Genetics*, **33**(7): 436–447.
- Häuser CL, De Jong R, Lamas G, Robbins RK, Smith C, Vane-Wright RI. 2005. Papilionidae—revised GloBIS/GART species checklist (2nd draft). <http://www.insectsonline.de/frames/papilio.htm>.
- Hochrainer K, Mayer H, Baranyi U, Binder B, Lipp J, Kroismayr R. 2005. The human HERC family of ubiquitin ligases: novel members, genomic organization, expression profiling, and evolutionary aspects. *Genomics*, **85**(2): 153–164.
- Hoekstra HE, Coyne JA. 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evolution*, **61**(5): 995–1016.
- Hubisz MJ, Pollard KS, Siepel A. 2011. PHAST and RPHAST: phylogenetic analysis with space/time models. *Briefings in Bioinformatics*, **12**(1): 41–51.
- Igarashi S. 1984. The classification of the Papilionidae mainly based on the morphology of their immature stages. *Lepidoptera Science*, **34**(2): 41–96.
- Iijima T, Kajitani R, Komata S, Lin CP, Sota T, Itoh T, et al. 2018. Parallel evolution of Batesian mimicry supergene in two *Papilio* butterflies, *P. polytes* and *P. memnon*. *Science Advances*, **4**(4): eaao5416.
- Ishida M, Oguchi ME, Fukuda M. 2016. Multiple types of guanine nucleotide exchange factors (GEFs) for Rab small GTPases. *Cell Structure and Function*, **41**(2): 61–79.
- Jagla K, Stanceva I, Dretzen G, Bellard F, Bellard M. 1994. A distinct class of homeodomain proteins is encoded by two sequentially expressed *Drosophila* genes from the 93D/E cluster. *Nucleic Acids Research*, **22**(7): 1202–1207.
- Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, **346**(6215): 1320–1331.
- Jiao XY, Flouri T, Rannala B, Yang ZH. 2020. The impact of cross-species gene flow on species tree estimation. *Systematic Biology*, **69**(5): 830–847.
- Jin H, Seki T, Yamaguchi J, Fujiwara H. 2019. Prepatterning of *Papilio xuthus* caterpillar camouflage is controlled by three homeobox genes: *clawless*, *abdominal-A*, and *Abdominal-B*. *Science Advances*, **5**(4): eaav7569.
- Koch PB, Behnecke B, French-Constant RH. 2000. The molecular basis of melanism and mimicry in a swallowtail butterfly. *Current Biology*, **10**(10): 591–594.
- Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics*, **5**(1): 59.
- Lamber EP, Siedenburg AC, Barr FA. 2019. Rab regulation by GEFs and GAPs during membrane traffic. *Current Opinion in Cell Biology*, **59**: 34–39.
- Li XY, Fan DD, Zhang W, Liu GC, Zhang L, Zhao L, et al. 2015. Outbred genome sequencing and CRISPR/Cas9 gene editing in butterflies. *Nature Communications*, **6**(1): 8212.
- Liu GC, Chang Z, Chen L, He JW, Dong ZW, Yang J, et al. 2020. Genome size variation in butterflies (Insecta, Lepidoptera, Papilionoidea): a thorough phylogenetic comparison. *Systematic Entomology*, **45**(3): 571–582.
- Liu GC, Liu W, Zhao RP, He JW, Dong ZW, Chen L, et al. 2021. Genome-wide identification and gene-editing of pigment transporter genes in the swallowtail butterfly *Papilio xuthus*. *BMC Genomics*, **22**(1): 120.
- Lloyd V, Ramaswami M, Krämer H. 1998. Not just pretty eyes: *Drosophila* eye-colour mutations and lysosomal delivery. *Trends in Cell Biology*, **8**(7): 257–259.
- Lu SH, Yang J, Dai XL, Xie FA, He JW, Dong ZW, et al. 2019. Chromosomal-level reference genome of Chinese peacock butterfly (*Papilio bianor*) based on third-generation DNA sequencing and Hi-C analysis. *Gigascience*, **8**(11): giz128.
- Lü ZM, Gong L, Ren YD, Chen YJ, Wang ZK, Liu LQ, et al. 2021. Large-scale sequencing of flatfish genomes provides insights into the polyphyletic origin of their specialized body plan. *Nature Genetics*, **53**(5): 742–751.
- Ma JP, Plesken H, Treisman JE, Edelman-Novemsky I, Ren MD. 2004. Lightoid and claret: a rab GTPase and its putative guanine nucleotide exchange factor in biogenesis of *Drosophila* eye pigment granules. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(32): 11652–11657.
- Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. *Bioinformatics*, **20**(16): 2878–2879.
- Mallet J, Besansky N, Hahn MW. 2016. How reticulated are species?. *Bioessays*, **38**(2): 140–149.
- Martin A, Papa R, Nadeau NJ, Hill RI, Counterman BA, Halder G, et al. 2012. Diversification of complex butterfly wing patterns by repeated regulatory evolution of a *Wnt* ligand. *Proceedings of the National Academy of Sciences of the United States of America*, **109**(31): 12632–12637.
- Martin A, Reed RD. 2010. *Wingless* and *aristaless2* define a developmental ground plan for moth and butterfly wing pattern evolution. *Molecular Biology and Evolution*, **27**(12): 2864–2878.
- Martin A, Reed RD. 2014. *Wnt* signaling underlies evolution and development of the butterfly wing pattern symmetry systems. *Developmental Biology*, **395**(2): 367–378.
- Matsuoka Y, Monteiro A. 2018. Melanin pathway genes regulate color and morphology of butterfly wing scales. *Cell Reports*, **24**(1): 56–65.
- Matsuoka Y, Monteiro A. 2021. Hox genes are essential for the development of eyespots in *Bicyclus anynana* butterflies. *Genetics*, **217**(1): iyaa005.
- Milán M, Weihe U, Tiong S, Bender W, Cohen SM. 2001. *msh* specifies dorsal cell fate in the *Drosophila* wing. *Development*, **128**(17): 3263–3268.
- Mishra S, Smolik SM, Forte MA, Stork PJS. 2005. Ras-independent activation of ERK signaling via the Torso receptor tyrosine kinase is

- mediated by Rap1. *Current Biology*, **15**(4): 366–370.
- Nazari V, Zakharov EV, Sperling FAH. 2007. Phylogeny, historical biogeography, and taxonomic ranking of Parnassiinae (Lepidoptera, Papilionidae) based on morphology and seven genes. *Molecular Phylogenetics and Evolution*, **42**(1): 131–156.
- Nelson AC, Wardle FC. 2013. Conserved non-coding elements and cis regulation: actions speak louder than words. *Development*, **140**(7): 1385–1395.
- Nijhout HF. 1990. A comprehensive model for colour pattern formation in butterflies. *Proceedings of the Royal Society B Biological Sciences*, **239**(1294): 81–113.
- Nijhout HF. 1991. The Development and Evolution of Butterfly Wing Patterns. Washington: Smithsonian Institution Press.
- Nishikawa H, Iijima T, Kajitani R, Yamaguchi J, Ando T, Suzuki Y, et al. 2015. A genetic mechanism for female-limited Batesian mimicry in *Papilio* butterfly. *Nature Genetics*, **47**(4): 405–409.
- Oxford GS, Gillespie RG. 1998. Evolution and ecology of spider coloration. *Annual Review of Entomology*, **43**: 619–643.
- Palmer DH, Kronforst MR. 2020. A shared genetic basis of mimicry across swallowtail butterflies points to ancestral co-option of *doublesex*. *Nature Communications*, **11**(1): 6.
- Podsiadlowski L, Tunström K, Espeland M, Wheat CW. 2021. The genome assembly and annotation of the Apollo butterfly *Parnassius apollo*, a flagship species for conservation biology. *Genome Biology and Evolution*, **13**(8): evab122.
- Polychronopoulos D, King JWD, Nash AJ, Tan G, Lenhard B. 2017. Conserved non-coding elements: developmental gene regulation meets genome organization. *Nucleic Acids Research*, **45**(22): 12611–12624.
- Reiner DJ, Lundquist EA. 2018. Small GTPases. WormBook: The Online Review of *C. elegans* Biology [Internet]: 1–65. http://www.wormbook.org/chapters/www_smallGTPases.2/smallGTPases.2.pdf.
- Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. *Nature Methods*, **17**(2): 155–158.
- Sayyari E, Whitfield JB, Mirarab S. 2018. DiscoVista: interpretable visualizations of gene tree discordance. *Molecular Phylogenetics and Evolution*, **122**: 110–115.
- Scriber JM, Tsubaki Y, Lederhouse RC. 1995. Swallowtail Butterflies: Their Ecology and Evolutionary Biology. Gainesville: Scientific Publishers.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, **13**(11): 2498–2504.
- Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou MM, Rosenbloom K, et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, **15**(8): 1034–1050.
- Smit AFA, Hubley R, Green P. 2015. RepeatMasker Open-4.0 (2015). <http://www.repeatmasker.org/RMDownload.html>.
- Solis-Lemus C, Bastide P, Ané C. 2017. PhyloNetworks: a package for phylogenetic networks. *Molecular Biology and Evolution*, **34**(12): 3292–3298.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**(9): 1312–1313.
- Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Research*, **34**(S2): W435–W439.
- Sundaram MV. 2005. The love-hate relationship between Ras and Notch. *Genes & Development*, **19**(16): 1825–1839.
- Swarup S, Verheyen EM. 2012. Wnt/Wingless signaling in *Drosophila*. *Cold Spring Harbor Perspectives in Biology*, **4**(6): a007930.
- Tatematsu KI, Yamamoto K, Uchino K, Narukawa J, Iizuka T, Banno Y, et al. 2011. Positional cloning of silkworm *white egg 2 (w-2)* locus shows functional conservation and diversification of ABC transporters for pigmentation in insects. *Genes to Cells*, **16**(4): 331–342.
- Timmermans MJTN, Srivathsan A, Collins S, Meier R, Vogler AP. 2020. Mimicry diversification in *Papilio dardanus* via a genomic inversion in the regulatory region of *engrailed-invented*. *Proceedings of the Royal Society B-Biological Sciences*, **287**(1926): 20200443.
- Tong XL, Hrycaj S, Podlaha O, Popadic A, Monteiro A. 2014. Over-expression of *Ultrabithorax* alters embryonic body plan and wing patterns in the butterfly *Bicyclus anynana*. *Developmental Biology*, **394**(2): 357–366.
- Tyler HA, Brown Jr KS, Wilson KH. 1994. Swallowtail Butterflies of the Americas: a Study in Biological Dynamics, Ecological Diversity, Biosystematics, and Conservation. Gainesville: Scientific Publishers.
- Umebachi Y. 1985. Papiliochrome, a new pigment group of butterfly. *Zoological Science*, **2**(2): 163–174.
- Van Nieuwerkerken EJ, Kaila L, Kitching IJ, Kristensen NP, Lees DJ, Minet J, et al. 2011a. Order Lepidoptera. Zhang ZQ. In: Animal Biodiversity: An Outline of Higher-Level Classification and Survey of Taxonomic Richness. Auckland, New Zealand: Zootaxa.
- Van Nieuwerkerken EJ, Kaila L, Kitching IJ, Kristensen NP, Lees DJ, Minet J, et al. 2011b. Order Lepidoptera Linnaeus, 1758. Zhang ZQ. In: Animal Biodiversity: An outline of Higher-Level Classification and Survey of Taxonomic Richness (Vol. 3148, pp. 212–221). Auckland, New Zealand: Magnolia Press (Zootaxa).
- Vankuren NW, Massardo D, Nallu S, Kronforst MR. 2019. Butterfly mimicry polymorphisms highlight phylogenetic limits of gene reuse in the evolution of diverse adaptations. *Molecular Biology and Evolution*, **36**(12): 2842–2853.
- Vavouri T, Walter K, Gilks WR, Lehner B, Elgar G. 2007. Parallel evolution of conserved non-coding elements that target a common set of developmental regulatory genes from worms to humans. *Genome Biology*, **8**(2): R15.
- Villa-Cuesta E, Modolell J. 2005. Mutual repression between *msh* and *Iro-C* is an essential component of the boundary between body wall and wing in *Drosophila*. *Development*, **132**(18): 4087–4096.
- Wallace AR. 1865. On the phenomena of variation and geographical distribution as illustrated by the Papilionidae of the Malayan region. *Transactions of the Linnean Society of London*, **25**(1): 1–71.
- Wang K, Lenstra JA, Liu L, Hu QJ, Ma T, Qiu Q, et al. 2018. Incomplete lineage sorting rather than hybridization explains the inconsistent phylogeny of the wisent. *Communications Biology*, **1**(1): 169.
- Wen DQ, Yu Y, Zhu JF, Nakhleh L. 2018. Inferring phylogenetic networks using PhyloNet. *Systematic Biology*, **67**(4): 735–740.
- Wilts BD, Trzeciak TM, Vukusic P, Stavenga DG. 2012. Papiliochrome II pigment reduces the angle dependency of structural wing colouration in *nireus* group papilionids. *Journal of Experimental Biology*, **215**(5): 796–805.
- Wittkopp PJ, Beldade P. 2009. Development and evolution of insect pigmentation: genetic mechanisms and the potential consequences of pleiotropy. *Seminars in Cell & Developmental Biology*, **20**(1): 65–71.
- Wu JM, Mao XZ, Cai T, Luo JC, Wei LP. 2006. KOBAS server: a web-based platform for automated annotation and pathway identification. *Nucleic Acids Research*, **34**(S2): W720–W724.

- Xu Z, Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research*, **35**(S2): W265–W268.
- Yagi T, Sasaki G, Takebe H. 1999. Phylogeny of Japanese papilionid butterflies inferred from nucleotide sequences of the mitochondrial *ND5* gene. *Journal of Molecular Evolution*, **48**(1): 42–48.
- Yang J, Wan WT, Xie M, Mao JL, Dong ZW, Lu SH, et al. 2020a. Chromosome-level reference genome assembly and gene editing of the dead-leaf butterfly *Kallima inachus*. *Molecular Ecology Resources*, **20**(4): 1080–1092.
- Yang YZ, Sun PC, Lv LK, Wang DL, Ru DF, Li Y, et al. 2020b. Prickly waterlily and rigid hornwort genomes shed light on early angiosperm evolution. *Nature Plants*, **6**(3): 215–222.
- Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**(8): 1586–1591.
- Yuan Y, Zhang YL, Zhang PJ, Liu C, Wang JH, Gao HY, et al. 2021. Comparative genomics provides insights into the aquatic adaptations of mammals. *Proceedings of the National Academy of Sciences of the United States of America*, **118**(37): e2106080118.
- Zakharov EV, Caterino MS, Sperling FAH. 2004. Molecular phylogeny, historical biogeography, and divergence time estimates for swallowtail butterflies of the genus *Papilio* (Lepidoptera: Papilionidae). *Systematic Biology*, **53**(2): 278–298.
- Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, **19**(S6): 153.
- Zou ZT, Zhang JZ. 2015. Are convergent and parallel amino acid substitutions in protein evolution more prevalent than neutral expectations?. *Molecular Biology and Evolution*, **32**(8): 2085–2096.