


















DATA NOTE

Chromosomal-level reference genome of Chinese peacock butterfly (*Papilio bianor*) based on third-generation DNA sequencing and Hi-C analysis

Sihan Lu ^{1,2,†}, Jie Yang ^{1,†}, Xuelei Dai ^{3,†}, Feiang Xie ⁴, Jinwu He ¹, Zhiwei Dong ², Junlai Mao ⁴, Guichun Liu ^{1,2}, Zhou Chang ², Ruoping Zhao ², Wenting Wan ¹, Ru Zhang ¹, Yuan Li ⁵, Wen Wang ^{1,2,6,*} and Xueyan Li ^{2,*}

¹Center for Ecological and Environmental Sciences, Northwestern Polytechnical University, No.1 Dongxiang Road, Chang'an District, Xi'an, Shaanxi 710129, China; ²State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, No.32 Jiaochang Raod, Kunming, Yunnan 650223, China; ³Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, No.22 Xinong Road, Yangling, Shaanxi 712100, China; ⁴School of Marine Science and Technology, Zhejiang Ocean University, No.1 Haida South Road, Lincheng Changzhi Island, Zhoushan, Zhejiang 316022, China; ⁵Nextomics Biosciences Institute, No.666 Gaoxin Road, Wuhan, Hubei 430000, China and ⁶Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, No.32 Jiaochang Raod, Kunming, Yunnan 650223, China

*Correspondence address. Xueyan Li, State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, No.32 Jiaochang Raod, Kunming, Yunnan 650223, China. E-mail: lxy@mail.kiz.ac.cn  <http://orcid.org/0000-0002-7801-2066>; Wen Wang, Center for Ecological and Environmental Sciences, Northwestern Polytechnical University, No.1 Dongxiang Road, Chang'an District, Xi'an, Shaanxi 710129, China. E-mail: wwang@mail.kiz.ac.cn  <http://orcid.org/0000-0003-0457-7846>

[†]These authors contributed equally to this work.

Abstract

Background: *Papilio bianor* Cramer, 1777 (commonly known as the Chinese peacock butterfly) (Insecta, Lepidoptera, Papilionidae) is a widely distributed swallowtail butterfly with a wide number of geographic populations ranging from the southeast of Russia to China, Japan, India, Vietnam, Myanmar, and Thailand. Its wing color consists of both pigmentary colored scales (black, reddish) and structural colored scales (iridescent blue or green dust). A high-quality reference genome of *P. bianor* is an important foundation for investigating iridescent color evolution, phylogeography, and the evolution of swallowtail butterflies. **Findings:** We obtained a chromosome-level *de novo* genome assembly of the highly heterozygous *P. bianor* using long Pacific Biosciences sequencing reads and high-throughput chromosome conformation capture technology. The final assembly is 421.52 Mb on 30 chromosomes (29 autosomes and 1 Z sex chromosome) with 13.12 Mb scaffold N50. In total, 15,375 protein-coding genes and 233.09 Mb of repetitive sequences were identified. Phylogenetic analyses indicated that *P. bianor* separated from a common ancestor of swallowtails ~23.69–36.04 million years ago. Demographic history

Received: 11 April 2019; Revised: 18 August 2019; Accepted: 4 October 2019

© The Author(s) 2019. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

suggested that the population expansion of this species from the last interglacial period to the last glacial maximum possibly resulted from its decreased natural enemies and its adaptation to climate change during the glacial period.

Conclusions: We present a high-quality chromosome-level reference genome of *P. bianor* using long-read single-molecule sequencing and Hi-C-based chromatin interaction maps. Our results lay the foundation for exploring the genetic basis of special biological features of *P. bianor* and also provide a useful data source for comparative genomics and phylogenomics among butterflies and moths.

Keywords: *Papilio bianor*; single-molecule real-time (SMRT) sequencing; high-throughput chromosome conformation capture map; chromosome-level reference genome; butterfly

Background

Butterflies are widely considered one of the most aesthetically appealing and popular animals owing to their extraordinarily diverse wing patterns among species, populations, sexes, and seasonal forms [1–3]. They also have many other intriguing traits such as complex life cycles, diverse larval morphology and habits, and high species diversity [4]. In light of this interest, butterflies have been regarded as important model organisms in such fields as morphology, physiology, ecology, development, genetics, and evolutionary biology [4–6] since Darwin proposed his theory of natural selection in 1859 [7]. Back in 1864, Bates, the famous pioneer of mimicry theory, predicted that “the study of butterflies... will someday be valued as one of the most important branches of Biological science” [8]. With recent technological advances, it is possible to conduct direct analysis (and even manipulation) of the genomes of individuals sampled from natural habitats without the need of inbreeding to reduce heterozygosity or to develop laboratory lines [9–11]. Thus, butterflies are becoming a promising system to explore the genetics and evolution of morphological diversification and speciation.

Compared with the extensive butterfly diversity of >18,000 described species [12], only 37 butterfly species in 6 families including 5 swallowtails (Papilionidae) have had their reference genomes dissected (as of 1 May 2019) [9, 13–31]. Among them, chromosomal-level reference genomes have been assembled only for 2 nymphalids (*Heliconius melpomene* and *Melitaea cinxia*) and 1 swallowtail (*Papilio xuthus*) [9, 24, 25] using linkage map methods. Chromosomal-level reference genomes for more butterflies are not only indispensable to identify subtle genetic variations underpinning morphological traits that may often result from small mutations in regulatory elements [32, 33] but also will provide a unique opportunity to promote evolutionary biological studies on butterflies as an important model system.

The development of third-generation single-molecule technology has paved the way for the dissection of complex genomes of different kinds of wild organisms including butterflies [25, 28, 30, 34, 35]. Combined with high-throughput chromosome conformation capture (Hi-C) technology, which was developed to identify chromatin interactions across the entire genome and is now also used as a powerful tool to assist genome assembly [36], chromosomal-level reference genomes have been obtained for many organisms including such insects as fruit flies [37], mosquitoes [38], and moths [39, 40]. Despite this, up to now there have been no such examples combining single-molecule sequencing and Hi-C technologies to assemble chromosomal-level reference genomes reported for butterflies.

Papilio bianor Cramer, 1777 (NCBI:txid76199) (Papilionidae, Papilioninae, Papilionini) (Fig. 1a), also known as the Chinese peacock black swallowtail emerald or the Chinese peacock, is a widely distributed swallowtail butterfly with a large range of ge-

ographic populations ranging from the southeast of Russia to China, Japan, India, Vietnam, Myanmar, and Thailand [41–43]. Its larvae mainly feed on plants of the family Rutaceae, such as *Citrus reticulata*, *Euodia meliifolia*, and *Zanthoxylum bungeanum* [41, 44, 45], and its complete life cycle lasts 40–50 days. Its wing colors consist of both pigmentary colored scales (black, reddish) and structurally colored scales (iridescent blue or green dust) [45], which makes it a promising model to explore the origin and evolution of combined colors in insects. Scientific interest in *P. bianor* has long existed, e.g., in its prothoracicotropic hormones [46], oviposition behavior [44, 47, 48], phylogenetic position and species delimitation [49–53], chromosome numbers [54], or mitochondrial genome [50, 55]. Here, combining Pacific Biosciences (PacBio) single-molecule real-time (SMRT) and Hi-C technologies, we constructed the chromosome-level reference genome of *P. bianor* (30 chromosomes).

Data Description

Insect collection and breeding

Wild eggs of *P. bianor* were collected in a northern suburb of Kunming City, Yunnan, China and then reared under conditions of 26°C, 80% relative humidity with 16 h/8 h light/darkness. The hatched larvae were fed with the Rutaceous plant *Zanthoxylum piperitum* under the same conditions. Two fifth instar larvae were collected for Hi-C sequencing. Pupae were reared under the same conditions as the eggs until their eclosion. Adults were collected for a genome survey using the Illumina sequencing platform and for *de novo* genome sequencing using the PacBio platform.

Genome survey using Illumina sequencing technology

Genomic DNA was isolated from the thorax and abdomen of a single male adult using a Genra Puregene Blood kit (Qiagen, Germany) following manual instructions. Paired-end (PE) libraries of 2 different insertion sizes (150 and 500 bp) were constructed and sequenced on an Illumina HiSeq2000 platform at BGI (Shenzhen, China). The total number of sequencing reads was ~16.45 Gb for PE150 and 28.42 Gb for PE500 (Table S1). We estimated genome size using Illumina short reads (PE150 and PE500), by *k*-mer distribution analysis with *k* = 17, using the formula: $G = k\text{-mer_number}/k\text{-mer_depth}$ [56]. Our data indicate that *P. bianor* has an estimated genome size of 496.05 Mb and a high heterozygosity of 1.81% (Fig. S1 and Table S2).

Library construction and sequencing using SMRT and Hi-C technologies

Genomic DNA was extracted from the thorax and abdomen of another male adult and used to construct one 20-kb library for the PacBio platform according to the manufacturers' protocols

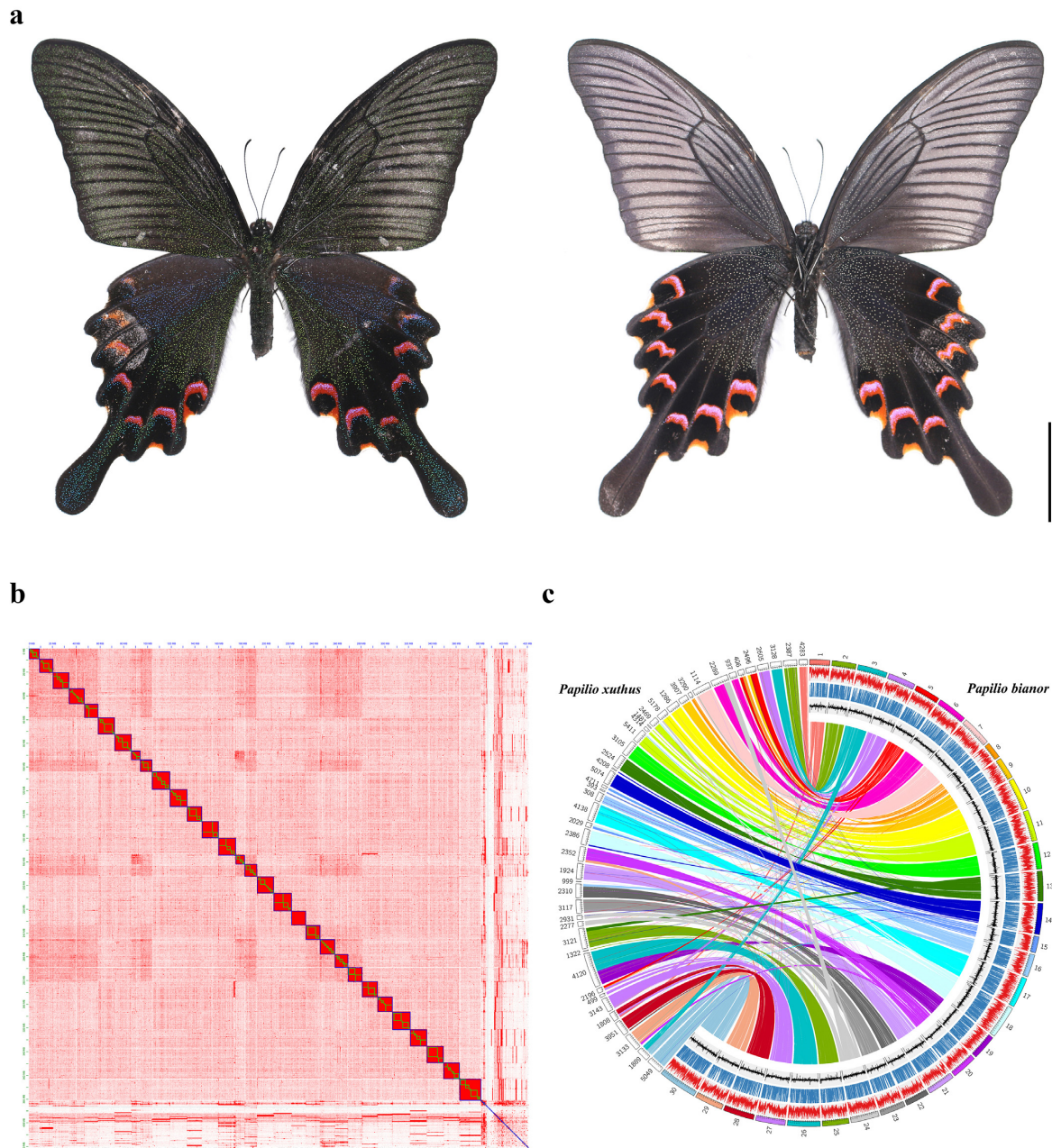


Figure 1: Characterization of *Papilio bianor*. (a) Female adult. Left, dorsal view; right, ventral view (scale bar = 20.0 mm; photo by Zhiwei Dong). (b) Heat map of chromosomal interactions. Each chromosome is framed with a blue block, and each scaffold is framed with a green block. (c) Circos plot of *P. bianor* chromosome-level reference genome with the previously released *Papilio xuthus* genome (obtained from a Chinese group) [9]. Shown from outermost to innermost are (1) gene density, (2) repeat element density, (3) GC content, and (4) syntenic regions with *P. xuthus* (left).

(NextOmics, China). With 10 SMRT cells in the PacBio RSII platform, we generated 43.19 Gb subreads with a mean read length of 16.4 kb after removing adaptor sequences within sequences (Table S1). The long subreads were used for *de novo* genome assembly of *P. bianor*.

A sample mixed from the whole bodies of 2 male larval individuals (fifth instar) was used for library construction for Hi-C sequencing according to the methods reported in the previous study [36]. A 400–700 bp library was sequenced on the Illumina HiSeq X Ten platform with 150 PE mode, resulting in ~75.11 Gb raw reads (Table S1).

Chromosomal-level genome assembly

Considering the high heterozygosity of *P. bianor* (1.81%: Fig. S1 and Table S2), we first performed a PacBio-only assembly using Wtdbg (version 1.2.8; Wtdbg, [RRID:SCR_017225](https://doi.org/10.26434/chemrxiv-2019-07-01-11111); with `-tidy-reads 5000 -k 0 -p 17 -S 1`) [57], which is a *de novo* sequence assembler for noisy long reads produced by PacBio or Oxford Nanopore Technologies and is based on the fuzzy Buijn graph algorithm. Second, to eliminate the high error rate of the PacBio long reads, we further polished the PacBio-only assembled sequences using Illumina reads as follows: all the Illumina reads were mapped to the PacBio-only assembly with BWA (version

0.7.12-r1039; BWA, [RRID:SCR.010910](#)) [58], which was further corrected with 2-round Pilon (version 1.21; Pilon, [RRID:SCR.014731](#)) correction [59, 60]. Third, because the polished assembly still contained a number of shorter contigs with significantly lower coverage, which perhaps represents the regions of high heterozygosity that were not merged to equivalent segments in the homologous chromosomes, we used a looser cut-off for identity (>90%) to merge the contigs with lower coverage and smaller size (size < 1,000 bp and coverage < 50 or size < 10,000 bp and coverage < 35) into the longer contigs as previously reported [14]. Fourth, the raw reads generated from the Hi-C sequencing were mapped to the polished assembled genome using Juicer (version 1.5; Juicer, [RRID:SCR.017226](#)) [61] and 3D *de novo* assembly (version 180114; 3D *de novo* assembly, [RRID:SCR.017227](#)) [38] to improve the assembly. Approximately 90.50% of contigs were anchored onto 30 super-scaffolds (Fig. 1b and Table S3; for more details see Fig. S2), which likely correspond to the 30 chromosomes as reported by cytogenetic karyotype [54]. Finally, we obtained the chromosomal-level high-quality assembly of *P. bianor* with a total length of ~421.52 Mb and the longest scaffold N50 (13.12 Mb) of any published butterfly genome to date (Tables 1 and S4). The assembled genome accounts for 85% of the estimated genome size (496.05 Mb) by the *k*-mer distribution analysis (Table S2).

Quality evaluation of assembled genome

The assembled genome quality was evaluated using 3 methods. First, the completeness of the assembly was evaluated by BUSCO (version 2.0; BUSCO, [RRID:SCR.015008](#)) [62] with the *insecta_odb9* BUSCO set. The BUSCO data showed that the *P. bianor* assembly covered 96.90% of the core genes with 96.30% covered genes complete (Table S5), which is similar to those of other published high-quality butterfly genomes (Table 1). We also checked the mapping rates of Illumina and PacBio reads to the *P. bianor* assembly by BWA (version 0.7.12-r1039; BWA, [RRID:SCR.010910](#)) [58] and BLASR (BLASR, [RRID:SCR.000764](#)) [63]. Our results indicate that 96.31% of Illumina reads mapped to the assembled genome with few heterozygous regions (Fig. S3 and Table S6); 96.86% of PacBio reads also mapped to the assembled genome with few heterozygous regions (Fig. S4 and Table S7). Third, we compared the syntenic relationships of the *P. bianor* genome with that of *P. xuthus*, which is the only chromosomal-level assembly (by linkage map methods) [9] among all *Papilio* reference genomes released to date and thus was considered to be the best-assembled one (Fig. 1c). We found that 61,082,412 bp of the *P. bianor* assembled genome could be aligned (1:1) with high confidence (-m 0.01) to the *P. xuthus* reference genome. All these results suggest that the *P. bianor* genome, which is assembled on the basis of PacBio reads, Illumina reads, and Hi-C data sequenced from different wild individuals, is of high quality (including completeness, base level contiguity, and accuracy) (Table 1).

Genome annotation

Repetitive sequences including tandem repeats and transposable elements (TEs) were searched for in the *P. bianor* assembled genome. To do this, we first used Tandem Repeats Finder (version 4.07b; Tandem Repeats Database, [RRID:SCR.005659](#); with 2 7 7 80 10 50 2000 -d -h parameters) [64] to annotate the tandem repeats. Then, TEs were identified using a combination of *de novo* and homology-based approaches at both the DNA and protein levels. At the DNA level, we used RepeatModeler (ver-

sion 1.0.4; RepeatModeler, [RRID:SCR.015027](#)) [65] to construct a *de novo* repeat library, which built a repeat consensus database with classification information, and then we adopted RepeatMasker (version 4.0.5; RepeatMasker, [RRID:SCR.012954](#)) [66] to search similar TEs against the known Repbase TE library (version 16.02) [67] and *de novo* repeat library. We also used LTR.FINDER (LTR Finder, [RRID:SCR.015247](#)) [68] to find long terminal repeats (LTRs). At the protein level, software RepeatProteinMask (version 3.3.0, a package in RepeatMasker) [66] was used to search the assembled *P. bianor* genome against the TE protein database using the WU-BLASTX engine. Finally, we identified and masked 55.3% of the *P. bianor* assembly as repeat regions (Table S8), which is the highest in published butterfly genomes (Table 1). Among all TEs, the most abundant class of repetitive elements are long interspersed nuclear elements (LINEs, 14.22%), and the next are DNA transposons (8.81%) (Table S9). Compared with the reference genomes of other swallowtail butterflies, LINEs, DNA transposons, and LTRs have expanded in the *P. bianor* genome (Fig. 2a). To confirm the reliability of the high repetitive sequences seen in *P. bianor*, which is much higher than those (40%) of other butterflies (Table 1), we also used other *de novo* annotation methods reported by Lavoie et al. [69] and Platt II et al. [70] to annotate the repetitive sequences of the *P. bianor* genome. On the basis of these methods, the *P. bianor* genome possesses 53% repeat elements, similar to the previous annotated results (Tables 1 and S8), thus confirming a high proportion of repetitive sequences in the *P. bianor* genome.

To annotate protein-coding genes of *P. bianor*, we used both *de novo* and homology-based gene prediction approaches. For *de novo* gene prediction, the repeat-masked genome was analyzed by SNAP (version 2006-07-28; SNAP, [RRID:SCR.002127](#)) [71], GENSCAN (version 1.0; GENSCAN, [RRID:SCR.012902](#)) [72], glimmerHMM (version 3.0.3; glimmerHMM, [RRID:SCR.002654](#)) [73], and AUGUSTUS (version 2.5.5; Augustus, [RRID:SCR.008417](#)) [74]. For homology-based predictions, the protein sequences from 8 insect species including the beetle *Tribolium castaneum* [75], fruit fly *Drosophila melanogaster* [76], silkworm *Bombyx mori* [77], moth *Helicoverpa armigera* [78], and 4 butterfly species *Papilio polytes* [23], *P. xuthus* [9], *Heliconius melpomene* [24], and *Danaus plexippus* [20] were used as templates for homology-based gene prediction. Then we used TBLASTN (version 2.2.26; TBLASTN, [RRID:SCR.011822](#)) [79] with an E-value cut-off of $1e^{-5}$ to align the protein sequences of the reference gene set to the *P. bianor* genome, and GeneWise (v2.2.0; GeneWise, [RRID:SCR.015054](#)) [80] to perform more precise alignment. Gene sequences with length < 150 bp or percent identity < 25% were removed. Evidence-Modeler software (EVM, version 1.1.1; [RRID:SCR.014659](#)) [81] was used to integrate the genes predicted by the homology and *de novo* approaches and generate a comprehensive, non-redundant gene set. Finally, 15,375 protein-coding genes were annotated in the assembled *P. bianor* genome (Table S10), which is similar to the published reference genomes of other swallowtail butterflies (Fig. S3).

The KEGG, TrEMBL, SwissProt, and COG databases were searched for best matches to *P. bianor* for the protein sequences yielded by EVM software, using BLASTP (version 2.2.26; BLASTP, [RRID:SCR.001010](#)) with an E-value cutoff of $1e^{-5}$, and Pfam, PRINTS, ProDom, and SMART databases were searched for known motifs and domains in our sequences using InterProScan software (version 5.18-57.0; InterProScan, [RRID:SCR.005829](#)) [82]. We also searched all predicted gene sequences against the GenBank nonredundant protein (nr) database using BLASTN (BLASTN, [RRID:SCR.001598](#)) with a maximal e-value of $1e^{-5}$. In sum, 13,343 genes were annotated with ≥ 1 related function,

Table 1: Comparison of quality and composition of different butterfly genomes

Family	Species	Genome size (Mb)	Genome size		Heterozygosity ^a (%)	Scaffold N50 (kb)	BUSCO ^b (%)	De novo assembled transcripts ^c (%)	GC content (%)	Repeat (%)	Exon (%)	Intron (%)	Number of proteins (k)
			without gap (Mb)	with gap (Mb)									
Papilionidae	<i>Papilio bianor</i>	421	421		1.8	13,120	96.3	NA	36.6	55.3	5.05	27.44	15.4
	<i>Papilio xuthus</i> [9]	244	238		1.0	6,199	97.6	NA	33.8	22.4	8.59	45.50	13.1
	<i>Papilio machaon</i> [9]	281	266		1.2	1,150	95.5	98	32.3	22.3	7.37	30.36	15.5
	<i>Papilio polytes</i> [17]	227	218		NA	3,672	91.8	NA	34.0	23.8	12.97	48.58	12.2
Hesperiidae	<i>Papilio memnon</i> [23]	233	219		NA	5,457	96.6	NA	32.8	22.5	11.31	43.17	12.4
	<i>Papilio glaucus</i> [14]	375	361		2.3	231	95.5	98	35.4	22.0	5.07	25.60	15.7
	<i>Achalarus lyciades</i> [31]	567	536		1.5	558	97.3	98	35.3	25.0	3.57	28.40	15.9
	<i>Lerema accius</i> [13]	298	290		1.5	525	95.1	98	34.4	15.5	6.96	31.60	17.4
Pieridae	<i>Megathymus ursus violae</i> [16]	429	427		0.1	4,153	98.3	99	34.7	25.8	4.59	30.90	14.1
	<i>Pieris rapae</i> [15]	246	243		1.5	617	98.0	99	32.7	22.7	7.91	33.30	13.2
	<i>Phoebis sennae</i> [21]	406	347		1.2	257	97.7	97	39.0	17.2	6.20	25.50	16.5
	<i>Danaus plexippus</i> [18]	249	242		0.6	716	98.0	96	31.6	16.3	8.40	28.10	15.1
Nymphalidae	<i>Heliconius melpomene</i> [24]	274	270		NA	194	95.6	NA	32.8	24.9	6.38	25.40	12.8
	<i>Melitaea cinxia</i> [25]	390	361		NA	119	83.0	97	32.6	27.5	4.34	31.20	16.7
	<i>Bicyclus anynana</i> [28]	475	470		NA	638	97.6	NA	36.5	25.8	4.73	38.36	22.6
	<i>Calephelis nemesis</i> [27]	809	783		0.5	206	95.6	99	34.9	34.8	2.25	19.60	15.4
Lycaenidae	<i>Calephelis virginensis</i> [27]	855	824		1.3	175	93.9	99	35.0	38.8	2.17	20.50	15.6
	<i>Calycoptis cecropis</i> [26]	729	689		1.2	233	95.5	96	37.1	34.0	3.11	24.00	16.5

^aThe heterozygosity of *P. bianor*, *P. machaon*, and *P. xuthus* was calculated on the basis of *k*-mer distribution analysis. The heterozygosity values of others (*P. glaucus*, *A. lyciades*, *L. accius*, *M. ursus violae*, *P. rapae*, *P. sennae*, *D. plexippus*, *C. nemesis*, *C. virginensis*, *C. cecropis*) were estimated using the Genome Analysis Toolkit (GATK).

^bBUSCO is calculated in this study.

NA: not available in the referenced citation.

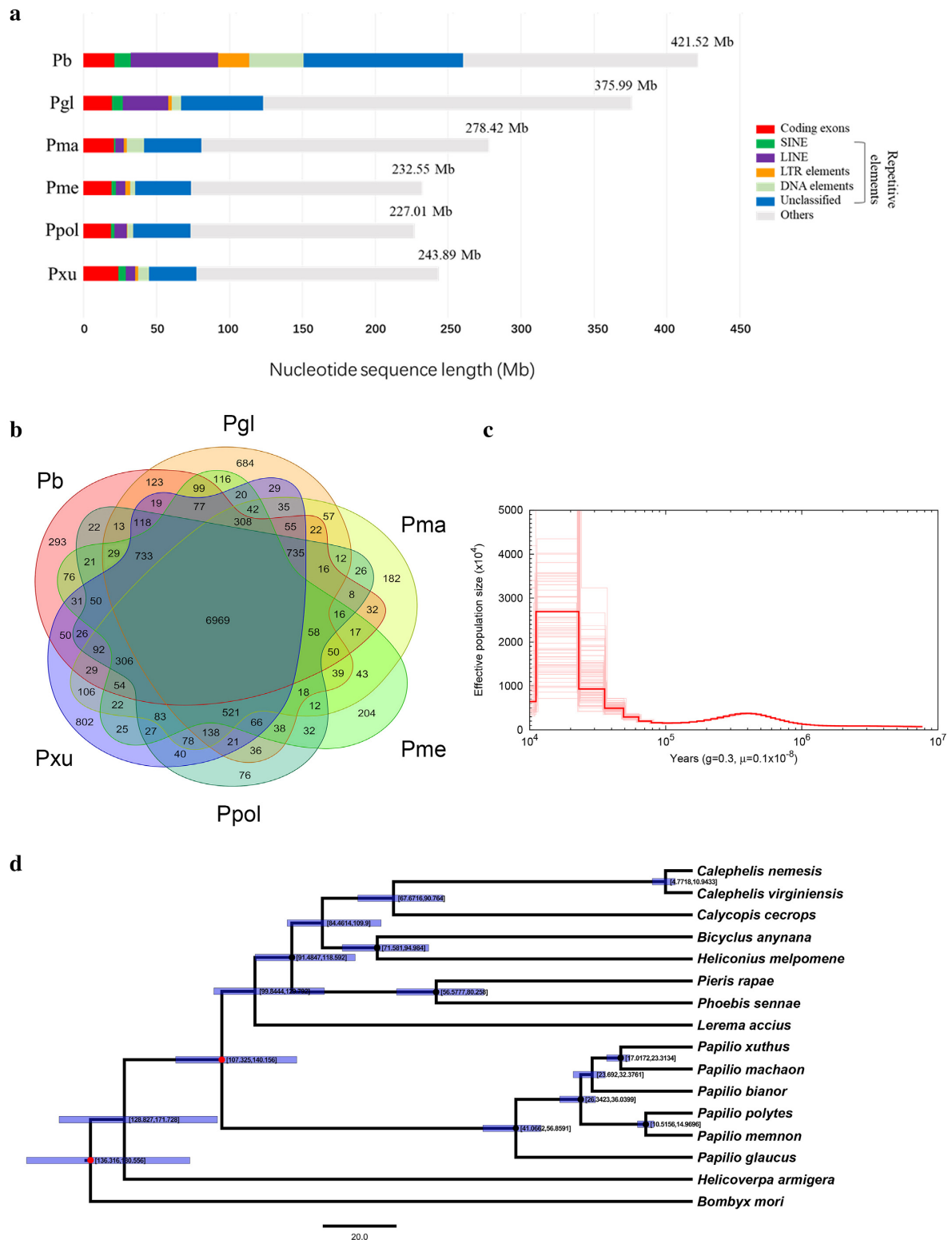


Figure 2: Genomic analysis of *Papilio bianor*. (a) Breakdown of the whole-genome assemblies into different functional classes in *Papilio*. (b) Venn diagram of the shared gene families of *Papilio*. (c) The dynamic changes of the effective population size were plotted using PSMC software, with 100 bootstrap replicates to test the robust variations. The parameter “g” represents the generation time in years, and the parameter “ μ ” means the per generation mutation rate. Pb: *Papilio bianor*; Pgl: *Papilio glaucus*; Pma: *Papilio machaon*; Pme: *Papilio memnon*; Ppol: *Papilio polytes*; Pxu: *Papilio xuthus*. (d) Maximum likelihood phylogenetic tree of Papilionoidea constructed by the concatenated alignment of 1,378 1-to-1 single-copy ortholog genes. The numbers in the square brackets on the nodes are the 95% confidence intervals of divergence time. The red dots are fossil evidence downloaded from the TimeTree website [89], and the black dots are inferred time obtained from the TimeTree website. Both were used to calibrate divergent time.

which accounts for ~86.78% of the *P. bianor* annotated genes (Table S11).

Gene family identification and phylogenetic analysis

We used OrthoMCL (version 2.0.9; OrthoMCL DB: Ortholog Groups of Protein Sequences, [RRID:SCR.007839](#)) [83] to cluster the *P. bianor* annotated genes with an E-value cutoff of $1e-5$, and Markov Chain Clustering with default inflation parameter in an all-to-all BLASTP analysis of entries for the reference genomes of 6 swallowtail butterflies including *P. bianor* in this study and the other 5 published so far (*P. polytes*, *P. xuthus*, *Papilio machaon*, *Papilio glaucus*, and *Papilio memnon*). The result showed that 293 gene families were specific to *P. bianor* (Fig. 2b). Using CAFE, version 4.0.1 [84], we also identified 375 expanded gene families and 1,863 contracted gene families in *P. bianor*. The *P. bianor* expanded gene families were enriched in 17 GO categories and the contracted gene families were enriched in 14 GO categories, most of which are related to oxygen metabolism (Tables S12 and S13).

To reveal the phylogenetic position of *P. bianor* among Papilionoidea, we selected 14 butterfly species in 5 families (Papilionidae [6]: *P. xuthus*, *P. polytes*, *P. machaon*, *P. glaucus*, *P. memnon*; Hesperidae [1]: *Lerema accius*; Pieridae [2]: *Phoebis sennae*, *Pieris rapae*; Nymphalidae [2]: *Bicyclus anynana*, *Heliconius melpomene*; Riodinidae [2]: *Calephelis nemesi*, *Calephelis virginensis*; Lycaenidae [1]: *Calycopis cecrops*) [9, 13–15, 17, 21, 23, 24, 26–28] with 2 moths (*B. mori* [77], *H. armigera* [78]) as outgroups for phylogenetic analysis. A total of 1,378 one-to-one single-copy orthologs that contain only 1 protein for each species were collected and clustered by OrthoMCL (version 2.0.9; OrthoMCL DB: Ortholog Groups of Protein Sequences, [RRID:SCR.007839](#)) [83] from these 16 species and their nucleic acid sequences were aligned using PRANK (version 3.8.31; PRANK, [RRID:SCR.017228](#)) [85]. Gene alignments were concatenated and phylogenetic trees were constructed using RAxML (version 7.2.8; RAxML, [RRID:SCR.006086](#)) [86] with the GTR+G+I model. Furthermore, to clarify our results, we also have constructed the gene trees for each of the orthologs with RAxML software (version 7.2.8; RAxML, [RRID:SCR.006086](#)) [84] by choosing the GTR+G+I model and inferred the species tree from these with ASTRAL software (version 5.6.3) [87] (Fig. S4). As expected, the results are consistent with each other. To further investigate the divergence time of these species, the phylogeny was further analyzed by MCMCtree in PAML (version 4.5; PAML, [RRID:SCR.014932](#)) software [88] using default parameters, and calibrated with published divergent times of some nodes estimated from fossil evidence or obtained from the TimeTree website [89]. Our phylogenetic tree showed that *P. bianor* clusters at the base of *P. machaon* and *P. xuthus* and diverged from them 23 million years ago (mya); the *Papilio* genus was monophyletic with a crown node age of ~41.07–56.86 mya (Fig. 2d). This tree is largely consistent with those constructed from cytochrome oxidases I, cytochrome oxidases II, and elongation factor 1 α [90, 91], and from 425 loci from 2 outgroups and 173 species of butterflies [92].

We also inferred the demographic histories of *P. bianor* applying the Pairwise Sequentially Markovian Coalescence (PSMC; PSMC, [RRID:SCR.017229](#); with $-p$ 64*1 parameters) analysis [93] (3.56×10^{-3} mutations per site per generation calculated by r8s [94]; 3 or 4 generations per year [48]), which was carried out by mapping Illumina short reads to the assembled genome with BWA (version 0.7.12-r1039; BWA, [RRID:SCR.010910](#)) [58] and calling variants with SAMtools (version 1.3.1; SAMTOOLS, [RRID:SCR.002105](#); with samtools mepileup -C50 -uf parameters) [95]. Our result suggested that the effective population size increased

significantly corresponding to the transition phase from the last interglacial period (~0.14–0.12 mya) to the last glacial maximum (~0.021–0.018 mya) (Fig. 2c), which is in good agreement with the other 5 published *Papilio* species [96]. We hypothesize that the population expansion of this species possibly results from the decrease of its natural enemies (e.g., birds or lizards) and from its adaptation to climate change during the last interglacial period and last glacial maximum

Conclusion

We present the chromosomal-level genome assembly of *P. bianor* with a contig and scaffold N50 of 5.50 and 12.51 Mb, respectively. The assembled genome included 15,375 protein-coding genes, 293 species-specific gene families, 375 expanded gene families, and 1,863 contracted gene families. *P. bianor* diverged from other *Papilio* ~23.69–36.04 mya. Our results also show that the effective population size of *P. bianor* increased significantly during the glacial period. Our results lay the foundation for exploring the special biological features of the Chinese peacock butterfly, and also provide a useful data source for comparative genomics and phylogenomics among butterflies and lepidopterans.

Availability of Supporting Data and Materials

The raw reads have been deposited at NCBI in the SRA under BioProject Number: PRJNA530186. The chromosome-level genome, annotation, and other supporting data are also available via the GigaScience database, GigaDB [97].

Additional Files

Figure S1: *k*-mer ($k = 17$) distribution in *Papilio bianor* genome. The first peak (depth = 26) is a heterozygous peak, which is higher than the main peak (depth = 53), suggesting that the *P. bianor* genome is highly heterozygous. The x-axis is depth (\times); the y-axis is the proportion that represents the frequency at that depth divided by the total frequency of all the depth.

Figure S2: Heat map of per-chromosomal interactions. Each scaffold is framed with a green block.

Figure S3: The coverage distribution of Illumina reads mapping to *Papilio bianor* genome. The histogram follows a normal distribution, indicating few heterozygous regions in the assembled genome.

Figure S4: The coverage distribution of PacBio reads mapping to *Papilio bianor* genome. The histogram follows a normal distribution, indicating few heterozygous regions in the assembled genome.

Figure S5: The statistics of annotated protein-coding genes of *Papilio*. (a) Messenger RNA length, (b) coding sequence (CDS) length, (c) exon length, (d) intron length, (e) exon number. The x-axis represents length or number and the y-axis represents the density of genes. Pb: *Papilio bianor*; Pgl: *Papilio glaucus*; Pma: *Papilio machaon*; Pme: *Papilio memnon*; Ppol: *Papilio polytes*; Pxu: *Papilio xuthus*.

Figure S6: Maximum Likelihood phylogenetic tree of Papilionoidea constructed by merging each of the single-copy orthologs.

Table S1: The statistics of sequencing data generated for *Papilio bianor* genome. The sequencing depth was calculated by the assembled genome size.

Table S2: Genome size estimation of *Papilio bianor* with *k*-mer distribution analysis using $k = 17$.

Table S3: The statistics of assembled chromosome-level genome of *Papilio bianor*. The Hi-C data were filtered by HiC-Pro software. In total, 6,690,421 pairs of reads, accounting for 68.04% of the total Hi-C data, were used in downstream analysis.

Table S4: The contiguity assessment of genome assembly of *Papilio bianor*.

Table S5: The quality evaluation of assembled genome of *Papilio bianor* by BUSCO software with insecta_odb9.

Table S6: The statistics of mapping ratio of Illumina reads to *Papilio bianor* assembled genome.

Table S7: The statistics of mapping ratio of PacBio reads to *Papilio bianor* assembled genome.

Table S8: The statistics of the annotated repeat sequences in *Papilio bianor* genome.

Table S9: The statistics of the TE contents in *Papilio bianor* genome.

Table S10: The statistics of predicted protein-coding genes in *Papilio bianor* genome.

Table S11: The statistics of gene function annotation in *Papilio bianor* genome.

Table S12: The GO term enrichment of expanded gene families in *Papilio bianor* genome.

Table S13: The GO term enrichment of contracted gene families in *Papilio bianor* genome.

Abbreviations

BLASR: Basic Local Alignment with Successive Refinement; bp: base pair; BUSCO: Benchmarking Universal Single-Copy Orthologs; BWA: Burrows-Wheeler Aligner; CAFE: Computational Analysis of gene Family Evolution; COG: Clusters of Orthologous Groups; EVM: EvidenceModeler; Gb: gigabase pairs; GC: guanine-cytosine; GO: gene ontology; Hi-C: high-throughput chromosome conformation capture; kb: kilobase pairs; KEGG: Kyoto Encyclopedia of Genes and Genomes; LINE: long interspersed nuclear element; LTR: long terminal repeat; Mb: megabase pairs; mya: million years ago; NCBI: National Center for Biotechnology Information; PacBio: Pacific Biosciences; PAML: Phylogenetic Analysis by Maximum Likelihood; PE: paired-end; PSMC: Pairwise Sequentially Markovian Coalescence; RAXML: Randomized Axelerated Maximum Likelihood; SMRT: single-molecule real-time; SNAP: Semi-HMM-based Nucleic Acid Parser; SRA: Sequence Read Archive; TE: transposable element; TrEMBL: Translation of European Molecular Biology Laboratory.

Competing interests

The authors declare that they have no competing interests.

Authors' Contributions

X.L. and W.Wang conceived and supervised the study. J.H., Z.D., Z.C., G.L., R.Zhao, and W.Wan fed and collected the samples. G.L., J.H., R.Zhao, and W.Wan extracted the genomic DNA. Y.L. took charge of Hi-C sequencing. S.L. and X.D. assembled the genome. S.L., J.Y., F.X., and R.Zhang carried out the quality assessment, repeat annotation, and gene annotation. J.Y., F.X., J.M., and R.Zhang carried out evolutionary analyses. S.L. uploaded the raw read data, genome assembly, and annotation in the GenBank and GigaScience (GigaDB) databases. S.L., X.L., and W.Wang wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (No. 31621062) (to W.Wang), the Chinese Academy of Sciences (XDB13000000 (to W.Wang), and CAS "Light of West China" (to X.L.).

References

- Boggs CL, Watt WB, Ehrlich PR. Butterflies: Ecology and Evolution Taking Flight. Chicago, IL: University of Chicago Press; 2003.
- Joron M, Mallet JLB. Diversity in mimicry: paradox or paradigm? Trends Ecol Evol 1998;13(11):461–6.
- Nijhout HF. The Development and Evolution of Butterfly Wing Patterns. Smithsonian Inst; 1991:293.
- Heikkilä M, Kaila L, Mutanen M, et al. Cretaceous origin and repeated tertiary diversification of the redefined butterflies. Proc Biol Sci 2012;279(1731):1093–9.
- Kawahara AY, Breinholt JW. Phylogenomics provides strong evidence for relationships of butterflies and moths. Proc Biol Sci 2014;281(1788):20140970.
- Mitter C, Davis DR, Cummings MP. Phylogeny and evolution of Lepidoptera. Annu Rev Entomol 2017;62:265–83.
- Darwin C. The Origin of Species; And, the Descent of Man. London, UK: John Murray; 1859.
- Bates H. New species of butterflies from Guatemala and Panama, collected by Osbert Salvin and F. du Cane Godman, Esqs. Entomol Mon Mag 1864;1(1/7):1–164.
- Li X, Fan D, Zhang W, et al. Outbred genome sequencing and CRISPR/Cas9 gene editing in butterflies. Nat Commun 2015;6:8212.
- Zhang LL, Reed RD. Genome editing in butterflies reveals that spalt promotes and Distal-less represses eyespot colour patterns. Nat Commun 2016;7:11769.
- Markert MJ, Zhang Y, Enuameh MS, et al. Genomic access to monarch migration using TALEN and CRISPR/Cas9-mediated targeted mutagenesis. G3 (Bethesda) 2016;6(4):905–15.
- van Nieukerken EJ, Kaila L, Kitching IJ, et al. Order Lepidoptera Linnaeus, 1758. Zootaxa 2011;3148:212–21.
- Cong Q, Borek D, Otwinowski Z, et al. Skipper genome sheds light on unique phenotypic traits and phylogeny. BMC Genomics 2015;16:639.
- Cong Q, Borek D, Otwinowski Z, et al. Tiger swallowtail genome reveals mechanisms for speciation and caterpillar chemical defense. Cell Rep 2015;10(6):910–9.
- Shen J, Cong Q, Kinch LN, et al. Complete genome of *Pieris rapae*, a resilient alien, a cabbage pest, and a source of anti-cancer proteins. F1000Res 2016;5:2631.
- Cong Q, Li W, Borek D, et al. The Bear Giant-Skipper genome suggests genetic adaptations to living inside yucca roots. Mol Genet Genomics 2019;294(1):211–26.
- Iijima T, Kajitani R, Komata S, et al. Parallel evolution of Batesian mimicry supergene in two *Papilio* butterflies, *P. polytes* and *P. memnon*. Sci Adv 2018;4(4):eaao5416.
- Zhan S, Merlin C, Boore JL, et al. The monarch butterfly genome yields insights into long-distance migration. Cell 2011;147(5):1171–85.
- Hill JA, Neethiraj R, Rastas P, et al. Unprecedented reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. Sci Adv 2019;5(6):eaau3648.

20. Zhan S, Zhang W, Niitepold K, et al. The genetics of monarch butterfly migration and warning colouration. *Nature* 2014;**514**(7522):317–21.
21. Cong Q, Shen JH, Warren AD, et al. Speciation in cloudless sulphurs gleaned from complete genomes. *Genome Biol Evol* 2016;**8**(3):915–31.
22. Talla V, Suh A, Kalsoom F, et al. Rapid increase in genome size as a consequence of transposable element hyperactivity in wood-white (Leptidea) butterflies. *Genome Biol Evol* 2017;**9**(10):2491–505.
23. Nishikawa H, Iijima T, Kajitani R, et al. A genetic mechanism for female-limited Batesian mimicry in *Papilio* butterfly. *Nat Genet* 2015;**47**(4):405–U169.
24. Dasmahapatra KK, Walters JR, Briscoe AD, et al. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 2012;**487**(7405):94–8.
25. Ahola V, Lehtonen R, Somervuo P, et al. The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. *Nat Commun* 2014;**5**:4737.
26. Cong Q, Shen JH, Borek D, et al. Complete genomes of hairstreak butterflies, their speciation, and nucleomitochondrial incongruence. *Sci Rep* 2016;**6**:24863.
27. Cong Q, Shen JH, Li WL, et al. The first complete genomes of metalmarks and the classification of butterfly families. *Genomics* 2017;**109**(5–6):485–93.
28. Nowell RW, Elsworth B, Oostra V, et al. A high-coverage draft genome of the mycalesine butterfly *Bicyclus anynana*. *Giga-science* 2017;**6**(7):1–7.
29. Mallet J. New genomes clarify mimicry evolution. *Nat Genet* 2015;**47**(4):306–7.
30. Davey JW, Chouteau M, Barker SL, et al. Major improvements to the *Heliconius melpomene* genome assembly used to confirm 10 chromosome fusion events in 6 million years of butterfly evolution. *G3 (Bethesda)* 2016;**6**(3):695–708.
31. Shen J, Cong Q, Borek D, et al. Complete genome of *Achalarus lyciades*, the first representative of the Eudaminae subfamily of skippers. *Curr Genomics* 2017;**18**(4):366–74.
32. Loehlin DW, Carroll SB. Evolutionary biology sex, lies and butterflies. *Nature* 2014;**507**(7491):172–3.
33. Brunetti CR, Selegue JE, Monteiro A, et al. The generation and diversification of butterfly eyespot color patterns. *Curr Biol* 2001;**11**(20):1578–85.
34. VanBuren R, Bryant D, Edger PP, et al. Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum*. *Nature* 2015;**527**(7579):508–U209.
35. Andere AA, Ii RNP, Ray DA, et al. Genome sequence of *Phormia regina* Meigen (Diptera: Calliphoridae): Implications for medical, veterinary and forensic research. *BMC Genomics* 2016;**17**:842.
36. Belaghzal H, Dekker J, Gibcus JH. Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. *Methods* 2017;**123**:56–65.
37. Chakraborty M, VanKuren NW, et al. Hidden genetic variation shapes the structure of functional elements in *Drosophila*. *Nat Genet* 2018;**50**(1):20–5.
38. Dudchenko O, Batra SS, Omer AD, et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 2017;**356**(6333):92–5.
39. Chen WB, Yang XW, Tetreau G, et al. A high-quality chromosome-level genome assembly of a generalist herbivore, *Trichoplusia ni*. *Mol Ecol Resour* 2019;**19**(2):485–96.
40. Xiang H, Liu XJ, Li MW, et al. The evolutionary road from wild moth to domestic silkworm. *Nat Ecol Evol* 2018;**2**(8):1268–79.
41. Wu C. *Fauna Sinica Insect Vol. 25 Lepidoptera Papilionidae*. Beijing: Science Press; 2001.
42. Sinev SY, ed. *Catalogue of the Lepidoptera of Russia*. Saint Petersburg, Moscow: KMK; 2008.
43. Chou I. *Monograph of Chinese Butterflies*. Zhengzhou: Henan Scientific and Technological Publishing House; 1994:1–854.
44. Ono H, Nishida R, Kuwahara Y. Oviposition stimulant for a Rutaceae-feeding swallowtail butterfly, *Papilio bianor* (Lepidoptera: Papilionidae): Hydroxycinnamic acid derivative from *Orixa japonica*. *Appl Entomol Zool* 2000;**35**(1):119–23.
45. Perveen F, Khan A, Sikander. Characteristics of butterfly (Lepidoptera) fauna from Kabal, Swat, Pakistan. *J Entomol Zool Stud* 2014;**2**(1):56–69.
46. Yokoyama I, Endo K, Yamanaka A, et al. Species-specificity in the action of big and small prothoracicotropic hormones (PT-THs) of the swallowtail butterflies, *Papilio xuthus*, *P. machaon*, *P. bianor* and *P. helenus*. *Zool Sci* 1996;**13**(3):449–54.
47. Ono H, Nishida R, Kuwahara Y. A dihydroxy-gamma-lactone as an oviposition stimulant for the swallowtail butterfly, *Papilio bianor*, from the Rutaceous plant, *Orixa japonica*. *Biosci Biotech Biochem* 2000;**64**(9):1970–3.
48. Dongsheng L. A preliminary observation on the artificial rearing of Xinyang *Papilio bianor*. *J Xinyang Teach Coll Nat Sci Ed* 1997;**2**:67–9.
49. Lixin Z, Xiaobing W, Chunsheng W, et al. Phylogenetic evaluation of *Papilio bianor* and *P. polyctor* (Lepidoptera: Papilionidae). *Orient Insect* 2009;**43**(1):25–32.
50. Hou LX, Ying S, Yang XW, et al. The complete mitochondrial genome of *Papilio bianor* (Lepidoptera: Papilionidae), and its phylogenetic position within Papilionidae. *Mitochondrial DNA A DNA Mapp Seq Anal* 2016;**27**(1):102–3.
51. Ae S. Some problems in hybrids between *Papilio bianor* and *P. maackii*. *Academia (Nanzan Univ)* 1962;**33**:21–8.
52. Chang Y-J. A study on hybridization of two subspecies of *Papilio bianor* (Lepidoptera, Papilionidae) in Taiwan. *Lepid Sci* 1990;**41**(1):1–6.
53. Yamada A. A study of interspecific hybrids between *Papilio bianor* and *P. maackii*. *Nat Insect* 1977;**12**:27–8.
54. Maeki K, Makino S. Chromosome numbers of some Japanese Rhopalocera. *Lepid News* 1953;**7**:36–8.
55. Dong Y, Zhu L-X, Wu Y-F, et al. The complete mitochondrial genome of the Chinese peacock, *Papilio bianor* (Insecta: Lepidoptera: Papilionidae). *Mitochondrial DNA* 2013;**24**(6):636–8.
56. Li R, Fan W, Tian G, et al. The sequence and de novo assembly of the giant panda genome. *Nature* 2010;**463**(7279):311–7.
57. Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. *bioRxiv* 2019:530972.
58. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 2013:13033997.
59. Walker BJ, Abeel T, Shea T, et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 2014;**9**(11):e112963.
60. Vaser R, Sovic I, Nagarajan N, et al. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* 2017;**27**(5):737–46.
61. Durand NC, Shamim MS, Machol I, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst* 2016;**3**(1):95–8.
62. Simao FA, Waterhouse RM, Ioannidis P, et al. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015;**31**(19):3210–2.
63. Chaisson MJ, Tesler G. Mapping single molecule sequencing reads using Basic Local Alignment with Successive Refine-

- ment (BLASR): Application and theory. *BMC Bioinformatics* 2012;13:238.
64. Benson G. Tandem Repeats Finder: A program to analyze DNA sequences. *Nucleic Acids Res* 1999;27(2):573–80.
 65. Smith A, Hubley R, Green P. RepeatMasker Open-4.0 (2013–2015). 2016. www.repeatmasker.org
 66. Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* 2004;5(1):4.10.1–14.
 67. Bao WD, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* 2015;6:11.
 68. Xu Z, Wang H. LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* 2007;35:W265–W8.
 69. Lavoie CA, Platt RN, Novick PA, et al. Transposable element evolution in *Heliconius* suggests genome diversity within Lepidoptera. *Mob DNA* 2013;4(1):21.
 70. Platt RN, Blanco-Berdugo L, Ray DA. Accurate transposable element annotation is vital when analyzing new genome assemblies. *Genome Biol Evol* 2016;8(2):403–10.
 71. Korf I. Gene finding in novel genomes. *BMC Bioinformatics* 2004;5:59.
 72. Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. *J Mol Biol* 1997;268(1):78–94.
 73. Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: Two open source ab initio eukaryotic gene-finders. *Bioinformatics* 2004;20(16):2878–9.
 74. Stanke M, Keller O, Gunduz I, et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res* 2006;34:W435–W9.
 75. Richards S, Gibbs RA, Weinstock GM, et al. The genome of the model beetle and pest *Tribolium castaneum*. *Nature* 2008;452(7190):949–55.
 76. Adams MD, Celniker SE, Holt RA, et al. The genome sequence of *Drosophila melanogaster*. *Science* 2000;287(5461):2185–95.
 77. Duan J, Li R, Cheng D, et al. SilkDB v2.0: A platform for silkworm (*Bombyx mori*) genome biology. *Nucleic Acids Res* 2010;38(Database issue):D453–6.
 78. Pearce SL, Clarke DF, East PD, et al. Genomic innovations, transcriptional plasticity and gene loss underlying the evolution and divergence of two highly polyphagous and invasive *Helicoverpa* pest species. *BMC Biol* 2017;15(1):63.
 79. Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 1997;25(17):3389–402.
 80. Birney E, Clamp M, Durbin R. GeneWise and genomewise. *Genome Res* 2004;14(5):988–95.
 81. Haas BJ, Salzberg SL, Zhu W, et al. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the program to assemble spliced alignments. *Genome Biol* 2008;9(1):R7.
 82. Jones P, Binns D, Chang H-Y, et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* 2014;30(9):1236–40.
 83. Li L, Stoeckert CJ, Roos DS. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res* 2003;13(9):2178–89.
 84. De Bie T, Cristianini N, Demuth JP, et al. CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* 2006;22(10):1269–71.
 85. Loytynoja A, Goldman N. An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci U S A* 2005;102(30):10557–62.
 86. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30(9):1312–3.
 87. Mirarab S, Reaz R, Bayzid MS, et al. ASTRAL: Genome-scale coalescent-based species tree estimation. *Bioinformatics* 2014;30(17):i541–8.
 88. Yang ZH. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Biol Evol* 2007;24(8):1586–91.
 89. Kumar S, Stecher G, Suleski M, et al. TimeTree: A resource for timelines, timetrees, and divergence times. *Mol Biol Evol* 2017;34(7):1812–9.
 90. Zakharov EV, Caterino MS, Sperling FA. Molecular phylogeny, historical biogeography, and divergence time estimates for swallowtail butterflies of the genus *Papilio* (Lepidoptera: Papilionidae). *Syst Biol* 2004;53(2):278–98.
 91. Dupuis JR, Sperling FA. Repeated reticulate evolution in North American *Papilio machaon* group swallowtail butterflies. *PLoS One* 2015;10(10):e0141882.
 92. Espeland M, Breinholt J, Willmott KR, et al. A comprehensive and dated phylogenomic analysis of butterflies. *Curr Biol* 2018;28(5):770–8.e5.
 93. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature* 2011;475(7357):493–U84.
 94. Sanderson MJ. r8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 2003;19(2):301–2.
 95. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25(16):2078–9.
 96. Zhang W, Westerman E, Nitzany E, et al. Tracing the origin and evolution of supergene mimicry in butterflies. *Nat Commun* 2017;8(1):1269.
 97. Lu S, Yang J, Dai X, et al. Supporting data for “Chromosomal-level reference genome of Chinese peacock butterfly (*Papilio bianor*) based on third-generation DNA sequencing and Hi-C analysis.” *GigaScience Database* 2019. <http://dx.doi.org/10.5524/100653>.