

## Complete mitochondrial genome of *Colias sifanica* Grum-Grshimailo, 1891 (Lepidoptera: Pieridae) from montane region in northwestern China

Ying Wang<sup>a</sup>, Zeyu Hu<sup>b</sup>, Liufang Wei<sup>b</sup>, Fengying Wan<sup>c</sup>, Fanyu Wei<sup>b</sup>, Jiasheng Hao<sup>b</sup> and Chengyong Su<sup>b</sup>

<sup>a</sup>College of Physical Education, Anhui Normal University, Wuhu, China; <sup>b</sup>College of Life Sciences, Anhui Normal University, Wuhu, China;

<sup>c</sup>Library of Anhui Normal University, Wuhu, China

### ABSTRACT

*Colias sifanica* Grum-Grshimailo, 1891, is a typical montane butterfly species which occurs on the Qinghai-Tibet Plateau (QTP) and adjacent regions in China. In this study, the complete mitochondrial genome of this species was assembled from data generated by next-generation sequencing. The mitochondrial genome was 15,151 bp in length and comprised 13 protein-coding genes (PCGs), 2 ribosomal RNA genes, 22 transfer RNA genes and a control region. The base composition of the *C. sifanica* mitochondrial genome was 39.7% A, 41.3% T, 11.3% C and 7.7% G, significantly AT biased as commonly found in other Pieridae mitochondrial genomes. Phylogenetic analyses based on all PCGs using both the maximum likelihood and Bayesian inference methods indicated that *C. sifanica* is closely related to *C. fieldii* with high support values, and the phylogenetic relationship of (*Dercas* + ((*Gandaca* + *Gonepteryx*) + (*Phoebis* + (*Anteos* + (*Catopsilia* + (*Zerene* + *Colias*)))))) was shown for the tribe Coliadini. Though both the mitochondrial gene order and overall base composition were found to be conserved, different Ka/Ks ratios for several mitochondrial PCGs were detected between *Colias* and other species in the tribe Coliadini.

### ARTICLE HISTORY

Received 26 December 2023  
Accepted 9 August 2024

### KEYWORDS

*Colias*; mitogenome;  
phylogenetic inference;  
evolution

## Introduction


The butterfly genus *Colias* contains more than 80 species and has long been used for ethological, ecological and evolutionary studies (Remington 1954; Wheat and Watt 2008; Xian et al. 2021). Members of these species have been reported to be the sympatric sister species that hybridize frequently with parallel evolution (Wang and Porter 2004; Porter and Levin 2007; Dwyer et al. 2015). However, nearly half of the *Colias* species only occur in remote mountainous regions of Central-Asia and the Qinghai-Tibet Plateau (QTP) that have been far less studied (Laiho and Ståhls 2013).


The typical Lepidoptera mitochondrial genome is a circular and compact molecule, ranging from 14 to 16 kb in size and consisting of 13 protein-coding genes (PCGs), two ribosomal RNA genes (rRNAs), 22 transfer RNA genes (tRNAs) and a non-coding region (also known as control region), has been widely used in studies of taxonomic resolution, phylogeny and molecular evolution (Hao et al. 2014; Wahlberg et al. 2014; Cao et al. 2016; Yong et al. 2017; Wei et al. 2022). To date, complete mitochondrial genomic sequences have been reported for two *Colias* species (*C. erate* and *C. fieldii*), which were collected from the Anhui and Jiangxi Province of China,

respectively (Wu et al. 2016; Xian et al. 2021). *Colias sifanica* Grum-Grshimailo, 1891, is a typical montane butterfly species, endemic on the QTP and adjacent regions (Laiho and Ståhls 2013; Huang and Song 2022). This species has the dark brown eyes, pink ventral antennae, yellow-green wing surface with radially yellow arrows at the end of the forewing chamber of the frontal submargin area, and the light-black M-vein on the hindwing. The complete mitochondrial genome sequence for *C. sifanica* is not yet available. Deciphering the phylogenetic relationships and molecular evolution pattern underlying the potential high-altitude adaptation mechanisms for species in this genus has been limited because of insufficient genomic data. Herein, we sequenced and characterized the mitochondrial genome of the *C. sifanica*, aiming to provide a complete reference mitochondrial genome, valuable for determining robust phylogenetic relationships and adaptive evolution for this butterfly genus in the future.

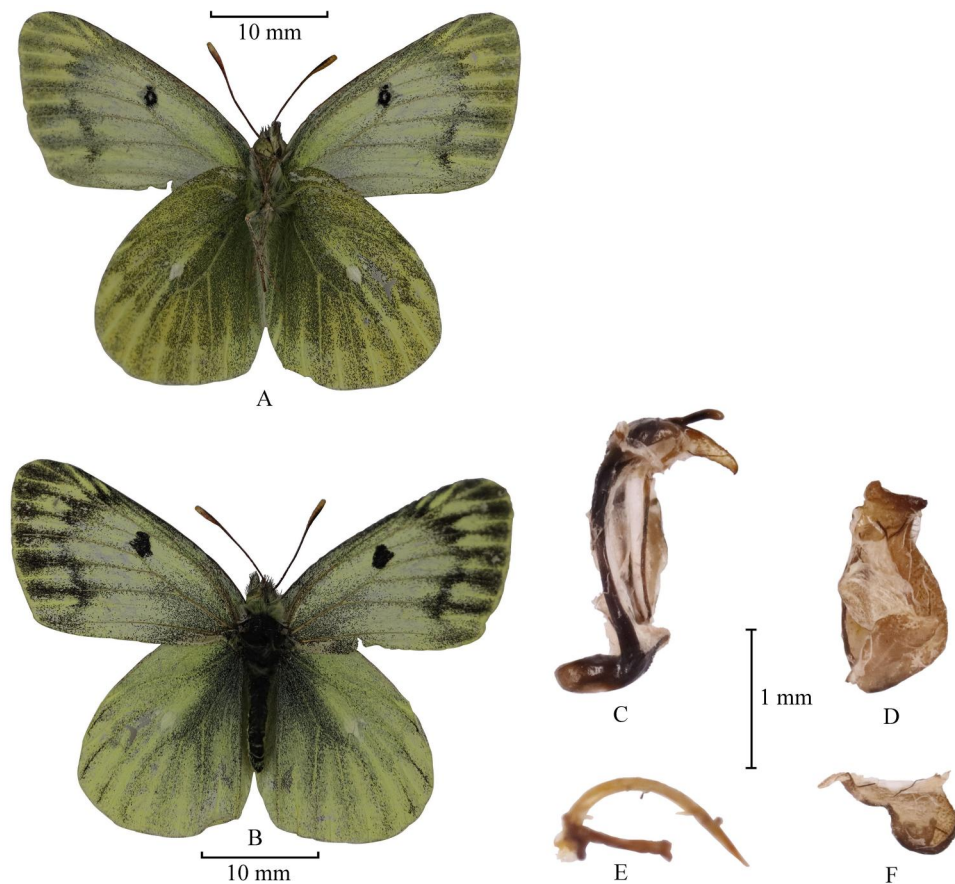
## Materials and methods

The adult individual of *C. sifanica* was sampled from the Qingsha Mountain, Qinghai, China (N 36.268, E 101.978, 3300 m above sea level) in July 2021 (Figure 1), and initially

CONTACT Chengyong Su  [sucy@ahnu.edu.cn](mailto:sucy@ahnu.edu.cn)

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/23802359.2024.2392760>.

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group  
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial 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. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.



**Figure 1.** Species reference image of *Colias sifanica* and the male genitalia. A. Ventral side; B. Dorsal side; C. Ring in lateral view; D. Right-hand valvula; E. Aedeagus; F. Juxta. The image was taken by Zeyu Hu.

preserved in absolute ethyl alcohol in the field and then transferred to  $-80^{\circ}\text{C}$  at the College of Life Sciences, Anhui Normal University ([www.ahnu.edu.cn](http://www.ahnu.edu.cn), Chengyong Su, [sucy@ahnu.edu.cn](mailto:sucy@ahnu.edu.cn)) under the voucher number CS\_QSS.

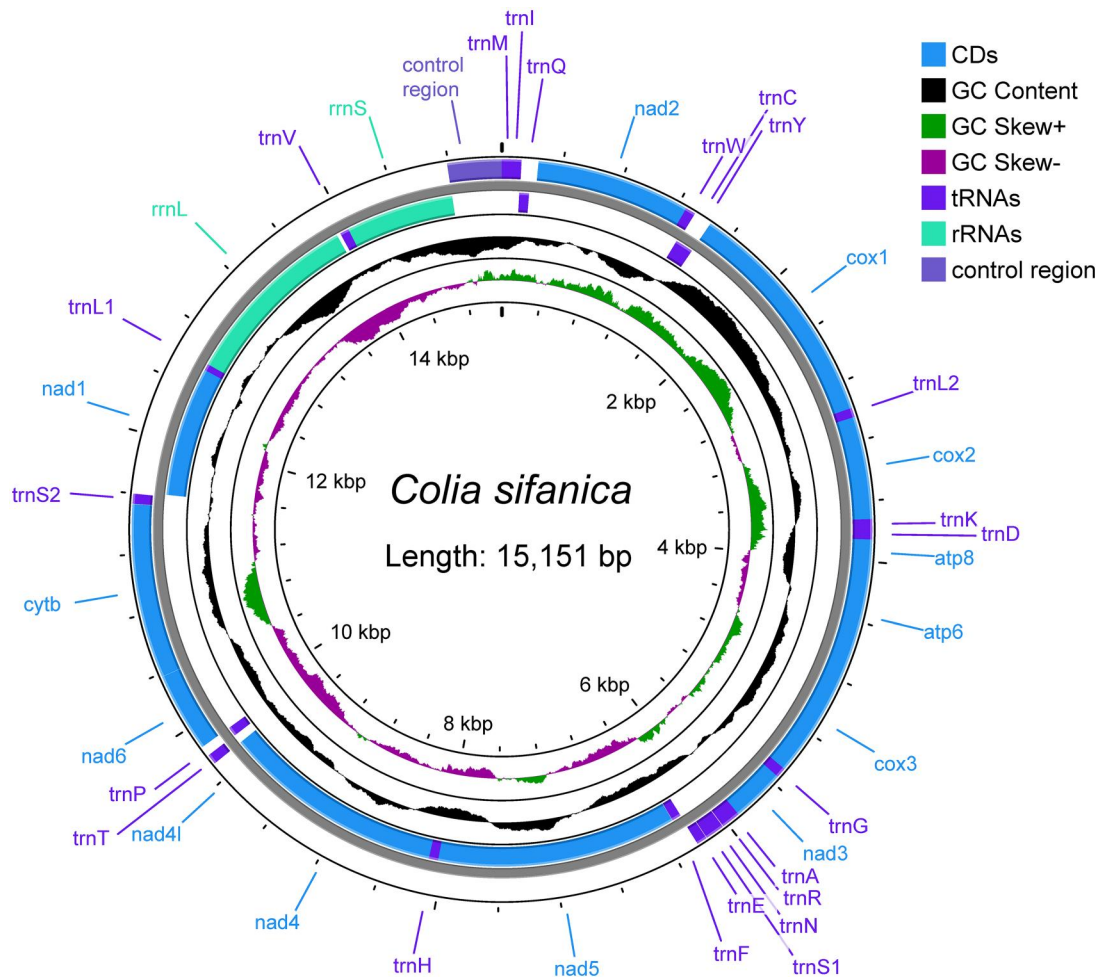
Total genomic DNA extraction, DNA library preparation and genomic DNA sequencing were conducted following the approach in previous study (Wei et al. 2022). The quality of reads was assessed using fastp v0.23.1 (Chen et al. 2018). High-quality clean reads ( $Q20 > 95\%$  and  $Q30 > 90\%$ ) were then assembled using GetOrganelle v1.7.0 with default settings (Jin et al. 2020). The mitogenome was annotated by the MITOS2 (Donath et al. 2019), and visualized via the CGView Web server (<https://cgview.ca>). The coverage depth was calculated using bowtie2 v2.5.2 (Langmead and Salzberg 2012).

Maximum likelihood (ML) and Bayesian inference (BI) methods were employed to reconstruct the phylogenetic trees based on the concatenated amino acid sequences of 13 PCGs from 22 pierid species (*Leptidea morsei* in Dismorphiinae as the outgroup). ML analysis was performed in IQ-TREE v1.6.8 using the best substitution models as determined in ModelFinder (Nguyen et al. 2015; Kalyaanamoorthy et al. 2017), with the branch support values estimated by the ultrafast bootstrap (UFBoot) and SH-like approximate likelihood ratio test (SH-aLRT) both with 10,000 replications (Guindon et al. 2010; Hoang et al. 2018). Bayesian analysis was conducted in MrBayes 3.2.6 (Ronquist et al. 2012), as

performed in previous study (Su et al. 2020). In addition, the non-synonymous rate ( $K_a$ ), synonymous rate ( $K_s$ ), and  $K_a/K_s$  values of 13 PCGs from 22 pierid butterfly species were calculated using KaKs\_Calculator v2.0 with the MYN model (Wang et al. 2010).

## Results

A total of 144.1 million clean reads (clean base = 21.6 Gbp) were generated. The complete mitogenome of *C. sifanica* herein is 15,151 bp in size, with high mean coverage-depth ( $>28,300\times$ , Figure S1), containing 13 PCGs, two rRNA genes, 22 transfer RNAs (tRNA) genes, and one control (AT-rich) region (Figure 2, Table S1). The gene order of mitogenome is conserved when compared to those of other previously sequenced Pieridae butterflies (Hao et al. 2013; Cao et al. 2016; Fang et al. 2016). The overall mitogenomic base composition was estimated to be A 39.7%, T 41.3%, C 11.3%, and G 7.7%, significantly AT biased as commonly found in other butterfly mitogenomes (Tables S2 and S3). Significant differences in relative synonymous codon usage (RSCU), including those of CUU, CCU, GCA, CGU and AGU, were found between the genus *Colias* and other representative genera in Pieridae ( $p < 0.05$ , Wilcoxon rank sum test, Figure S2, Table S4). Moreover, except for *cox1*, all other PCGs use ATN as the



**Figure 2.** Gene arrangement in the mitogenome of *Colia sifanica*.

start codon. Most PCGs use TAA, whereas *cox1*, *cox2*, *nad4* and *nad5* use incomplete T as the stop codon, as also shown in *C. erate* and *C. fieldii* (Wu et al. 2016; Xian et al. 2021). The *rrnL*, *rrnS* and control region in *C. sifanica* are 1326, 771 and 368 bp in size, respectively, and tRNA genes ranged from 60 to 70 bp, totaling 1447 bp in size. In addition, there were no significant differences in amino acids composition of PCGs between the genus *Colias* and other representative genera in Pieridae ( $p > 0.05$ , Wilcoxon rank sum test, Figure S3, Table S5).

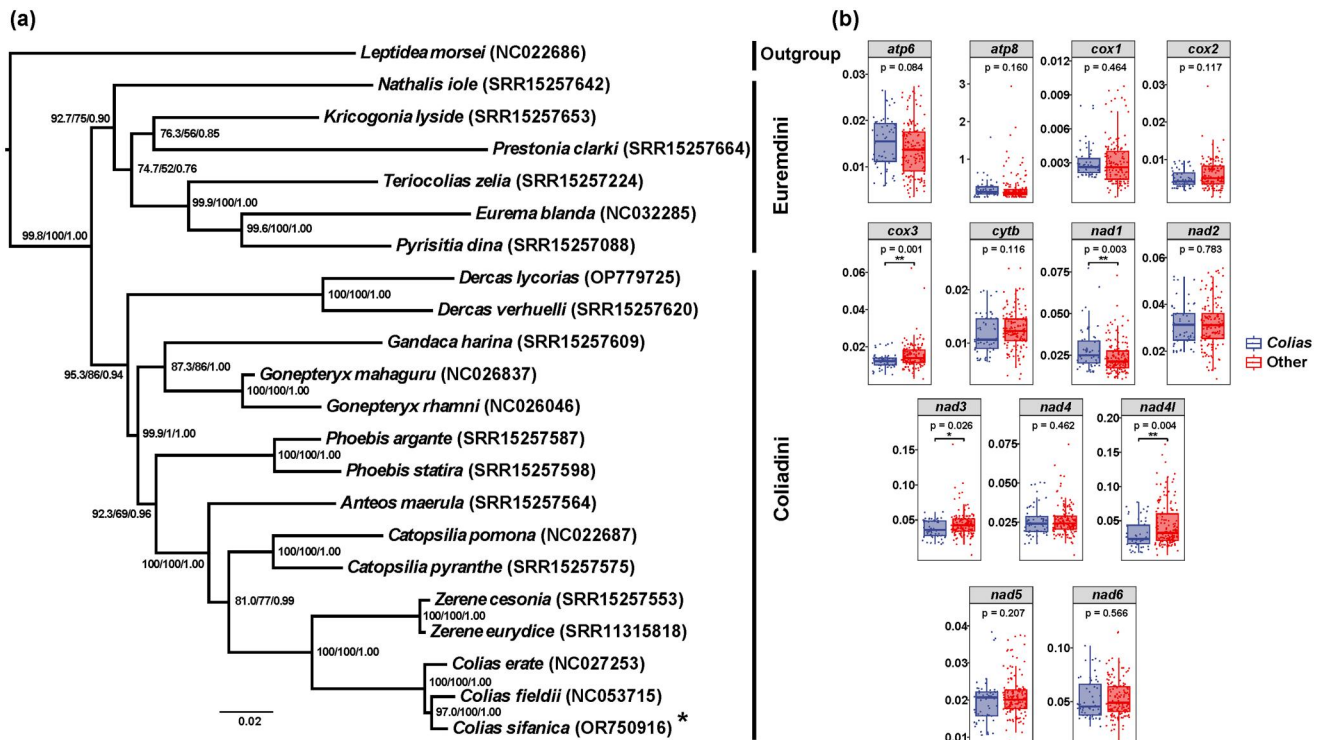
Phylogenetic analyses using the ML and BI methods confirmed that *C. sifanica* belongs to the tribe Coliadini in the subfamily Coliadinae (Figure 3a); among the three *Colia* species of this study, the *C. sifanica*, rather than *C. erate*, was the potential sister species of *C. fieldii* with high support values (SH-aLRT = 97.0, UFboot = 100 and posterior probabilities (PP) = 1.00); and two monophyletic tribes (Euremdini and Coliadini) in the subfamily Coliadinae covering a total of 14 genera were well supported in the phylogenetic tree. Our results also showed that the phylogeny of the tribe Coliadini was as follows: (*Dercas* + ((*Gandaca* + *Gonepteryx*) + (*Phoebis* + (*Anteos* + (*Catopsilia* + (*Zerene* + *Colias*)))))), similar to those in previous studies (Wahlberg et al. 2014; Wei

et al. 2022). Furthermore, the Ka/Ks ratios of four PCGs (*cox3*, *nad1*, *nad3* and *nad4l*) of the three *Colias* species herein were significantly different from those of other species in the tribe Coliadini (Figure 3b,  $p < 0.05$ , Wilcoxon rank sum test), which suggested that these PCGs likely underwent different evolutionary rates between *Colias* and other species.

## Discussion and conclusions

In this work, we first reported the mitogenome of *C. sifanica*, a typical montane butterfly species endemic on the QTP and adjacent regions. Phylogenetic analyses based on the 13 PCGs using both the ML and BI methods indicated that *C. sifanica*, is closely related to *C. fieldii* with high support values. Here, two monophyletic tribes (Euremdini and Coliadini) in the subfamily Coliadinae were recovered, and the phylogenetic relationship of (*Dercas* + ((*Gandaca* + *Gonepteryx*) + (*Phoebis* + (*Anteos* + (*Catopsilia* + (*Zerene* + *Colias*)))))) was shown for the tribe Coliadini. Additionally, though both the mitogenomic gene order and overall base composition were found to be conserved, different Ka/Ks ratios for several mitogenomic PCGs were detected between *Colias* and other species in the tribe Coliadini.





**Figure 3.** Phylogenetic tree and comparison of the Ka/Ks ratio using the amino acid sequences of 13 PCGs. (a) Phylogenetic relationships of Coliadae species based on the ML (model: mtART + F+R4) and BI (model: mixed) analysis. Support values (SH-aLRT, UFboot, and PP) for each node were shown in order from the left to right. Mitogenomic sequences of following species were also used: *L. morsei* NC022686 (Hao et al. 2014), *E. blanda* NC032285 (Yong et al. 2017), *D. lycorias* OP779725 (Wei et al. 2022), *G. mahaguru* NC026837 (Yang et al. 2016), *G. rhamni* NC026046 (Wang et al. 2015), *C. pomona* NC022687 (Hao et al. 2014), *C. erate* NC027253 (Wu et al. 2016), *C. fieldii* NC053715 (Xian et al. 2021), and other sequences deposited in NCBI Sequence Read Archive were assembled by Wei et al. (2022). (b) Comparison result of Ka/Ks ratio for each PCGs between *Colias* and other Coliadini species based on Wilcoxon rank sum test.

## Acknowledgments

The authors thank Prof. Ruie Nie (Anhui Normal University, China) for assistance in assembling the mitogenome.

## Authors' contributions

Conceptualization, C.S. and F.Wei; Data curation, Y.W. and F.Wan; Analysis and interpretation of the data, Y.W., Z.H., and L.W.; Funding acquisition, J.H. and C.S.; Writing – original draft, Y.W., Z.H., F.Wei, and C.S.; Writing – review & editing, F.Wan and J.H. All authors agree to be accountable for all aspects of the work.

## Ethical approval

This research does not involve ethical research. Insects are invertebrates, and there are no ethics involved in using them in experiments.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by the National Science Foundation of Universities of Anhui Province, under Grant No. [KJ2021A0100]; and the National Science Foundation of China, under Grant No. [41972029].

## Data availability statement

The genome sequence data that support the findings of this study are openly available in GenBank of NCBI at <https://www.ncbi.nlm.nih.gov>

under the accession no. OR750916, and the figshare website ([https://figshare.com/articles/dataset/i\\_Colias\\_sifanica\\_i\\_mitochondrion\\_complete\\_genome/24596955](https://figshare.com/articles/dataset/i_Colias_sifanica_i_mitochondrion_complete_genome/24596955)). The associated BioProject, SRA, and Bio-Sample numbers are PRJNA1040342, SRR26837766 and SAMN38256405, respectively.

## References

- Cao Y, Hao JS, Sun XY, Zheng B, Yang Q. 2016. Molecular phylogenetic and dating analysis of pierid butterfly species using complete mitochondrial genomes. *Genet Mol Res.* 15(4):gmr15049196. doi:10.4238/gmr15049196.
- Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 34(17):i884–i890. doi:10.1093/bioinformatics/bty560.
- Donath A, Jühling F, Al-Arab M, Bernhart SH, Reinhardt F, Stadler PF, Middendorf M, Bernt M. 2019. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. *Nucleic Acids Res.* 47(20):10543–10552. doi:10.1093/nar/gkz833.
- Dwyer HE, Jasieniuk M, Okada M, Shapiro AM. 2015. Molecular evidence for hybridization in *Colias* (Lepidoptera: Pieridae): are *Colias* hybrids really hybrids? *Ecol Evol.* 5(14):2865–2877. doi:10.1002/ece3.1574.
- Fang J, Wu Y, Wang H, Sun Z, Han D, Zhang B. 2016. The complete nucleotide sequence of the mitochondrial genome of the Oriental *Pieris*, *Pieris canidia* (Lepidoptera: Pieridae). *Mitochondrial DNA A DNA Mapp Seq Anal.* 27(6):4088–4089. doi:10.3109/19401736.2014.1003857.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59(3):307–321. doi:10.1093/sysbio/syq010.
- Hao JJ, Hao JS, Sun XY, Zhang LL, Yang Q. 2014. The complete mitochondrial genomes of the Fenton's wood white, *Leptidea morsei*, and the lemon emigrant, *Catopsilia pomona*. *J Insect Sci.* 14(130):1–22. doi:10.1093/jis/14.1.130.

- Hao JJ, Wang YL, Sun XY, Zhang LL, Hao JS, Yang Q. 2013. The complete mitochondrial genome of *Hebomoia glaucippe* (Lepidoptera: Pieridae). *Mitochondrial DNA*. 24(6):668–670. doi:10.3109/19401736.2013.773983.
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 35(2):518–522. doi:10.1093/molbev/msx281.
- Huang H, Song K. 2022. A review of the genus *Colias* Fabricius, 1807 (Lepidoptera: Pieridae) from Qinghai with descriptions of a new subspecies of *Colias thrasibulus* Fruhstorfer, 1910. *Zootaxa*. 5105(4):451–489. doi:10.11646/zootaxa.5105.4.1.
- Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol*. 21(1):241. doi:10.1186/s13059-020-02154-5.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 14(6):587–589. doi:10.1038/nmeth.4285.
- Laiho J, Ståhls G. 2013. DNA barcodes identify Central-Asian *Colias* butterflies (Lepidoptera, Pieridae). *Zookeys*. 365(365):175–196. doi:10.3897/zookeys.365.5879.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 9(4):357–359. doi:10.1038/nmeth.1923.
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 32(1):268–274. doi:10.1093/molbev/msu300.
- Porter AH, Levin EJ. 2007. Parallel evolution in sympatric, hybridizing species: performance of *Colias* butterflies on their introduced host plants. *Entomologia Exp Applicata*. 124(1):77–99. doi:10.1111/j.1570-7458.2007.00553.x.
- Remington CL. 1954. The genetics of *Colias* (Lepidoptera). *Adv Genet*. 6: 403–450. doi:10.1016/s0065-2660(08)60133-9.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 61(3):539–542. doi:10.1093/sysbio/sys029.
- Su C, Xie T, Wang Y, Si C, Li L, Ma J, Li C, Sun X, Hao J, Yang Q. 2020. Miocene diversification and high-altitude adaptation of *Parnassius* butterflies (Lepidoptera: Papilionidae) in Qinghai-Tibet Plateau revealed by large-scale transcriptomic data. *Insects*. 11(11):754. doi:10.3390/insects11110754.
- Wahlberg N, Rota J, Braby MF, Pierce NE, Wheat CW. 2014. Revised systematics and higher classification of pierid butterflies (Lepidoptera: Pieridae) based on molecular data. *Zool Scr*. 43(6):641–650. doi:10.1111/zsc.12075.
- Wang BQ, Porter AH. 2004. An AFLP-based interspecific linkage map of sympatric, hybridizing *Colias* butterflies. *Genetics*. 168(1):215–225. doi:10.1534/genetics.104.028118.
- Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010. KaKs\_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genom Proteom Bioinform*. 8(1):77–80. doi:10.1016/S1672-0229(10)60008-3.
- Wang Y, Peng C, Yao Q, Shi Q, Hao J. 2015. The complete mitochondrial genome of *Gonepteryx rhamni* (Lepidoptera: Pieridae: Coliadae). *Mitochondrial DNA*. 26(5):791–792. doi:10.3109/19401736.2013.855755.
- Wei F, Huang W, Fang L, He B, Zhao Y, Zhang Y, Shu Z, Su C, Hao J. 2022. Spatio-temporal evolutionary patterns of the Pieridae butterflies (Lepidoptera: papilionoidea) inferred from mitogenomic data. *Genes (Basel)*. 14(1):72. doi:10.3390/genes14010072.
- Wheat CW, Watt WB. 2008. A mitochondrial-DNA-based phylogeny for some evolutionary-genetic model species of *Colias* butterflies (Lepidoptera, Pieridae). *Mol Phylogenet Evol*. 47(3):893–902. doi:10.1016/j.ympev.2008.03.013.
- Wu YH, Fang J, Li WB, Han DM, Wang H, Zhang BW. 2016. The complete mitochondrial genome of *Colias erate* (Lepidoptera: Pieridae). *Mitochondrial DNA A DNA Mapp Seq Anal*. 27(6):4209–4210. doi:10.3109/19401736.2015.1022743.
- Xian WR, Liu YX, Ma YQ, Zhou H, Song X. 2021. Phylogenetic relationship and characterization of the complete mitochondrial genome of *Colias fieldii* (Lepidoptera: papilionoidea: Pieridae). *Mitochondrial DNA B Resour*. 6(8):2146–2148. doi:10.1080/23802359.2021.1944379.
- Yang J, Xu C, Li J, Lei Y, Fan C, Gao Y, Xu C, Wang R. 2016. The complete mitochondrial genome of *Gonepteryx mahaguru* (Lepidoptera: Pieridae). *Mitochondrial DNA A DNA Mapp Seq Anal*. 27(2):877–878. doi:10.3109/19401736.2014.919485.
- Yong HS, Song SL, Lim PE, Eamsobhana P, Suana IW. 2017. Complete mitochondrial genome of *Eurema blanda* and phylogenetic relationship with its congener *E. hecabe* and other pierid butterflies. *Meta Gene*. 11:141–146. doi:10.1016/j.mgene.2016.08.005.