

## The analysis of genetic variation in the mitochondrial genome and its application for the identification of *Papilio* species

Zuo Ruihua<sup>a,b</sup>, Jiang Ping<sup>a,b</sup>, Sun Chuanbo<sup>a</sup>, She Deyong<sup>a,b</sup>, Zhang Feng<sup>c</sup> and Hu Chaochao<sup>d</sup>

<sup>a</sup>College of Biological and Pharmaceutical Engineering, West Anhui University, Lu'an, China; <sup>b</sup>Animal Healthy Breeding and Animal Epidemic Monitoring and Warning Enter, West Anhui University, Lu'an, China; <sup>c</sup>College of Fisheries and Life Science, Shanghai Ocean University, Shanghai, China; <sup>d</sup>Analytical and Testing Center, Nanjing Normal University, Nanjing, China

### ABSTRACT

Mitochondrial DNA (mtDNA) markers are ideal for evolutionary studies, including phylogeography, population genetics, phylogeny, etc. However, different mitochondrial genes always own different evolutionary rate. In this study, we analysed the genetic variation across the 16 complete mtDNA from 13 species in the genus *Papilio* and recognized the best DNA barcoding for *Papilio* species. The mitochondrial gene arrangement for each species shares a similar order, similar to the typical Papilionidae species, which indicated the relatively conservative state of gene arrangement in *Papilio*. The sliding window of genetic diversity showed that there was a significant difference in the genetic diversity of each gene in the mitochondrial genome of *Papilio*. The relatively mean clock rate of the *ND1* was broadly lower than the other genes in mitochondrial genome of *Papilio*; while the *ATP8* owns the largest values of mean clock rate. Those results suggested that the rate of evolution of each gene is not balanced and all mitochondrial genes except *ND1* and *ATP8* could act as barcoding for the identification of *Papilio* species. The phylogenetic analyses of complete mtDNA data for 13 *Papilio* species divided them into five major branches, which keep the same topological structure with previous studies.

### ARTICLE HISTORY

Received 18 March 2018  
Accepted 9 May 2018

### KEYWORDS

*Papilio*; mitochondrial genome; genetic variation; mean clock rate; phylogeny

### Introduction

The *Papilio* is a genus in the swallow tail butterfly family Papilionidae (Vattikonda et al. 2014). As a widespread species all around the world, it includes a number of well-known species, such as *P. polytes*, *P. polymnestor*, *P. memnon*, and *P. deiphobus* (Sbordoni and Forestiero 1998; Condamine et al. 2013; Cong et al. 2015). Nowadays, about 210 *Papilio* species have been documented worldwide, with some 27 species recorded in China (Wu 2001). As they are highly sensitive to habitat loss and fragmentation, pollution and climate change, the genus face the great crisis of survival (Settele et al. 2009; Li et al. 2011). In IUCN, about 12 species in this genus are considered as vulnerable or endangered species, which make more urgent to investigate the genus. In addition, the swallowtail butterflies were also acted as model organisms for studies in evolutionary biology and conservation biology (Maes and Dyck 2005; Zhu et al. 2011).

The typical metazoan mitochondrial DNA (mtDNA) is a double-stranded circular DNA molecule, which encodes a conserved set of 37 genes, including 13 protein-coding genes (PCGs) plus the two ribosomal RNA (rRNA) genes and 22 transfer RNA (tRNA) genes (Shadel and Clayton 1993; Boore 1999). This, together with other special characteristics such as faster evolutionary rates than nuclear genes, presumed





maternal inheritance, and absence of recombination makes mtDNAs one of the most popular targets for population genetics studies and accurate identification and differentiation of a cryptic species (Zakharov et al. 2004; Kim et al. 2005).

Herein, in this study, we clarified the 16 complete mitochondrial genome sequences about 13 *Papilio* species in China, collected from NCBI, and compared these sequences and genetic variation with each other. In addition, with the help of online software OGDRAW (<http://ogdraw.mpimp-golm.mpg.de/>) to create the mtDNA ring structure, calculate the relative mean clock rate of each gene and also the phylogenetic relationships of this genus using mitochondrion sequence datasets.

### Results

#### Genome organization

The complete mitochondrial genomes of the species in the genus *Papilio* were from 15,239 bp to 16,094 bp (Table S1). The organization of the sequence encodes 13 protein genes (*ATP6*, *ATP8*, *COI-III*, *ND1-6*, *ND4L*, and *Cyt b*), 22 tRNA gene and a putative control region (*D-loop*) (Figure 1). Among these PCGs, the longest one is *ND5* genes (1719–1740 bp) which is located between the *tRNA<sup>Phe</sup>* (64–70 bp) and *tRNA<sup>His</sup>*

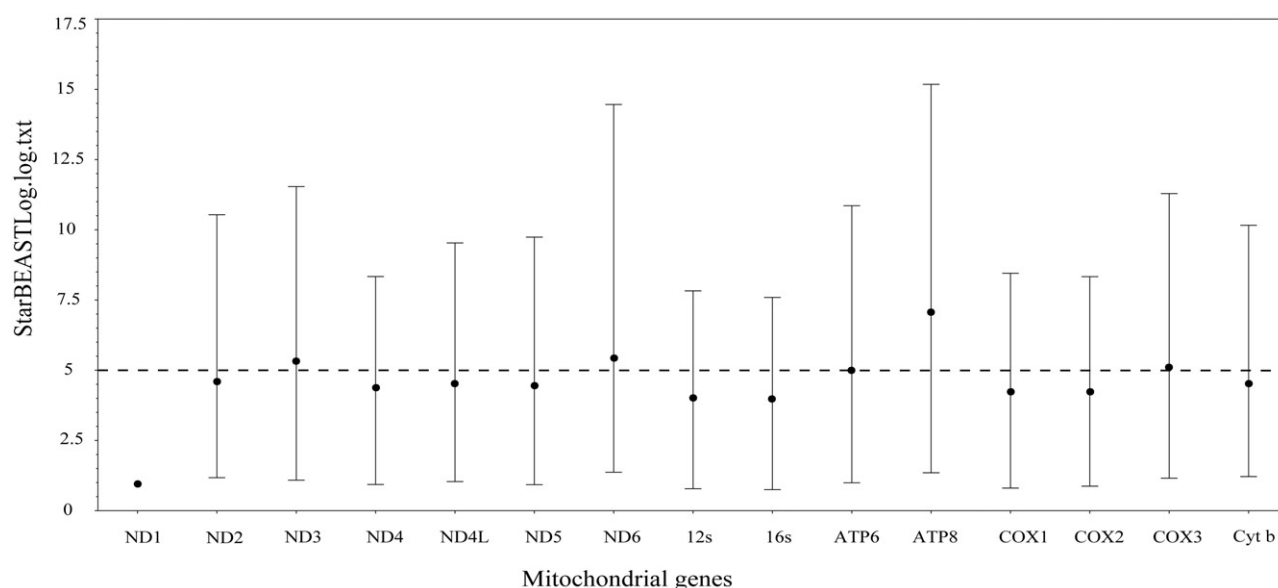
**CONTACT** Zhang Feng  zhangfeng122710@163.com  College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306, China; Chao-Chao Hu  huweichen@126.com  Nanjing Normal University, Nanjing, Jiangsu 210023, China

 Supplemental data for this article can be accessed [here](#).

© 2018 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 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.





**Figure 2.** The comparison of relatively mean clock rate of 15 mito genes. Bars indicated the 95% CI.

endangered species which is recorded in IUCN. For most species in the order of Lepidoptera, the mitochondrial genomes appear as the same topological structure and arrangement, which has no introns, no long intergenic spacers with only a few overlapping sequences. In this study, we compared the length of the 13 species in *Papilio*, and found the length of them is similar with each other. The longest of one from them is *P. maraho* (16,082 bp) and the shortest one is *P. machaon* (15,185 bp). The overall base composition is: A, 38.9–40.5%; T, 40.0–41.6%; C, 11.3–12.8% and G, 7.2–7.9%. The A+T content is higher than G+C, the detailed information is listed in Table S2. Compared with the mitochondrial genomes in the 13 species, the A+T compositions of *Papilio* mtDNA are similar, and it shares with the other genomes a strong AT bias (Table S3). Guanine (G) is the rarest nucleotide; the percentage of the other three bases is roughly equal to each other, similar to other Lepidoptera animals. GC and AT skews are a measure of compositional asymmetry; in arthropods mtDNA, GC-skew values are all negative ( $G < C$ ), and the AT-skew is also negative ( $A < T$ ). In *Papilio* mtDNA, GC-skew (0.18–0.26) and AT-skew (0.01–0.03) values are in accord with this principle (Table S3).

### DNA barcoding

DNA barcoding has potential to identify species because sequence divergences are ordinarily much lower among individuals of a species than between closely related species (Hebert et al. 2003, 2004; Song et al. 2008). For establishing DNA barcoding for species, the orthology of characters is the fundamental factor. The ideal DNA barcoding should have an observable gap between intra- and interspecific levels of divergence and correctly identify species (Hebert et al. 2004; Meyer and Paulay 2005). Previously, the region of *COI* gene is recommend as the universal barcoding marker for many animal groups (Hebert et al. 2004; Ward et al. 2005; Ratnasingham and Hebert 2007). However, if the orthology assumption is violated, that is, whether paralogous sequences

are unknowingly treated as orthologs, incorrect inferences barcoding are inevitable (Funk and Omland 2003). For example, a number of molecular evolutionary processes can hinder correct amplification and identification of the orthologs (Rubinoff et al. 2006), including duplication of the gene of interest within the mitochondrial genome (Campbell and Barker 1999), nuclear integration of mtDNA (Bensasson et al. 2001), and heteroplasmy (Frey and Frey 2004). For avoiding the incorrect inferences barcoding, multiple genetic marker may be helpful. Therefore, variety of genetic markers, own the similar genetic substitution rate, may be needed. In our study, the sliding window of genetic diversity showed that there was a significant difference in the genetic diversity of each gene in the mitochondrial genome of *Papilio*. In addition, the relatively mean clock rate results suggested that the rate of evolution of each gene is not balanced and all mitochondrial genes except *ND1* and *ATP8* could act as barcoding for the identification of *Papilio* species. These results showed that the relatively mean clock rate of the *ND1* was broadly lower than those of the *COI* (about 3.81 times), *12S* (about 3.55 times), *16S* (about 3.58 times), and *COII* (about 3.98 times) genes. In conclusion, the identification of *Papilio* species may be selected two or more genetic markers, such as the combined of *COI* and *12S* genes, which was helpful to avoid the incorrect inferences barcoding.

### Disclosure statement

No potential conflict of interest was reported by the authors.

### Funding

This research was supported by the West Anhui University of Animal Husbandry Technology Co., Ltd. Jinfeng Source of Practical Education Base [No. 2016sjjd079], Anhui Provincial Colleges and Universities Provincial Quality Engineering School-Enterprise Cooperation Projects [No. 2016sjjd0], and Anhui Provincial Colleges and Universities Quality Project Excellence Talent Project [No. 2016sjjd072].

## References

- Bensasson D, Zhang D, Hartl DL, Hewitt GM. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol Evol.* 16:314–321.
- Boore JL. 1999. Animal mitochondrial genomes. *Nucleic Acids Res.* 27:1767–1780.
- Campbell NJ, Barker SC. 1999. The novel mitochondrial gene arrangement of the cattle tick, *Boophilus microplus*: fivefold tandem repetition of a coding region. *Mol Biol Evol.* 16:732–740.
- Condamine FL, Toussaint EF, Cotton AM, Genson GS, Sperling FA, Kergoat GJ. 2013. Fine-scale biogeographical and temporal diversification processes of peacock swallowtails (*Papilio* subgenus *Achillides*) in the Indo-Australian Archipelago. *Cladistics.* 29:88–111.
- Cong Q, Borek D, Otwinowski Z, Grishin NV. 2015. Tiger swallowtail genome reveals mechanisms for speciation and caterpillar chemical defense. *Cell Rep.* 10:910–919.
- Frey JE, Frey B. 2004. Origin of intra-individual variation in PCR-amplified mitochondrial cytochrome oxidase I of *Thrips tabaci* (Thysanoptera: Thripidae): mitochondrial heteroplasmy or nuclear integration? *Hereditas.* 140:92–98.
- Funk DJ, Omland KE. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu Rev Ecol Syst.* 34:397–423.
- Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astrartes fulgerator*. *Proc Natl Acad Sci USA.* 101:14812–14817.
- Hebert PDN, Ratnasingham S, Waard JRD. 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc Biol Sci.* 270:S96–S99.
- Kim I, Cha SY, Yoon MH, Hwang JS, Lee SM, Sohn HD, Jin BR. 2005. The complete nucleotide sequence and gene organization of the mitochondrial genome of the oriental mole cricket, *Gryllotalpa orientalis* (Orthoptera: Gryllotalpidae). *Gene.* 353:155–168.
- Li XS, Zhang YL, Fang JH, Schweiger O, Settele J. 2011. A butterfly hotspot in western China, its environmental threats and conservation. *J Insect Conserv.* 15:617–632.
- Maes D, Dyck HV. 2005. Habitat quality and biodiversity indicator performances of a threatened butterfly versus a multispecies group for wet heathlands in Belgium. *Biol Conserv.* 123:177–187.
- Meyer CP, Paulay G. 2005. DNA barcoding: error rates based on comprehensive sampling. *PLoS Biol.* 3:e422–2238.
- Ratnasingham S, Hebert PDN. 2007. BOLD: the barcode of life data system (<http://www.barcodinglife.org>). *Mol Ecol Notes.* 7:355–364.
- Rubinoff D, Cameron S, Will K. 2006. A genomic perspective on the shortcomings of mitochondrial DNA for “barcoding” identification. *J Hered.* 97:581–594.
- Sbordoni V, Forestiero S. 1998. Butterflies of the world. *Am Entomol.* 187–188.
- Settele J, Kudrna O, Harpke A, Kühn I, van Swaay C, Verovnik R, Warren M, Wiemers M, Hanspach J, Hickler T, et al. 2009. Climatic risk atlas of European butterflies. *BR.* 2:33.
- Shadel GS, Clayton DA. 1993. Mitochondrial transcription initiation. Variation and conservation. *J Biol Chem.* 268:16083–16086.
- Song H, Buhay JE, Whiting MF, Crandall KA. 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proc Natl Acad Sci USA.* 105:13486–13491.
- Vattikonda SR, Amanchi NR, Raja SS. 2014. Antifeedant activity of forskolin, an extract of *Coleus forskohlii*, against *Papilio demoleus* L. (Lepidoptera: Papilionidae) larvae. *Eur J Exp Biol.* 4:237–241.
- Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN. 2005. DNA barcoding Australia's fish species. *Philos Trans R Soc Lond.* 360:1847–1857.
- Wu CS. 2001. *Fauna Sinica Insect Vol. 25 Lepidoptera Papilionidae.* Beijing: Science Press.
- Zakharov EV, Caterino MS, Sperling FA. 2004. Molecular phylogeny, historical biogeography, and divergence time estimates for swallowtail butterflies of the genus *Papilio* (Lepidoptera: Papilionidae). *Syst Biol.* 53:193–215.
- Zhu LX, Wu XB, Wu CS. 2011. Phylogeographic history of the swallowtail *Papilio bianor* Cramer (Lepidoptera: Papilionidae) from China. *Orient Insects.* 45:93–102.