Identification and analysis of the complete mitochondrial genome of *Thaumetopoea pityocampa* (Lepidoptera: Notodontidae)

Kai Wu, Jinge Yang, Yuyang Ni, and Qiuning Liu

Abstract

The mitochondrial genome (mitogenome) provides important information for phylogenetic analysis and understanding evolutionary origins. *Thaumetopoea pityocampa* is a forest pest that harms nearly all cedar and pine species. In this study, the *T. pityocampa* mitochondrial genome was sequenced, assembled, and annotated. The sequence length of the genome was found to be 15,737 bp, containing 13 protein-coding genes (PCGs), two rRNA genes, 22 tRNA genes, and an A+T-rich region compared with the genomes of other lepidopterans. The overall nucleotide composition is: 37.3% T, 40.5% A, 14.6% C, and 7.6% G, demonstrating an AT bias (A+T: 77.8%). Our phylogenetic tree analysis results showed that *T. pityocampa* and *Ochrogaster lunifer* were the most similar species, with the closest evolutionary distance. The mitogenome sequence determined in this study will contribute to improved understanding of Notodontidae evolution.

Lepidoptera, which contains butterflies and moths, is the second largest order of insects, surpassed only by Coleoptera. Noctuoidea is among the largest superfamilies within Lepidoptera, with almost 42,400 described species. *Thaumetopoea pityocampa* (Lepidoptera: Notodontidae) is a pest that consumes the needles of conifers such as *Cedrus*, *Pinus*, and *Pseudotsuga* (Colacci et al. 2018). Besides damaging forest production, the setae of *T. pityocampa* are also a threat to human and animal health because they contain allergens such as Tha p 1 and Tha p 2 (Rebollo et al. 2002; Vega et al. 2011; Rodriguez-Mahillo et al. 2012; Kaszak et al. 2015; Berardi et al. 2017). Insect genomic information improves our understanding of various aspects of pests such as their physiology, biochemistry, reproduction, migration, and tolerance to extreme environments. The *T. pityocampa* genome is 537 Mb in length; de novo transcriptomic analysis of two phenologically divergent populations has identified 9625 unigenes while 29,701 functional genes and a long-chain non-coding region (AT enrichment region), at a length of 15,737 bp. The overall nucleotide composition is: 37.3% T, 40.5% A, 14.6% C, and 7.6% G, demonstrating an AT bias (A+T: 77.8%). Our phylogenetic tree analysis results showed that *T. pityocampa* and *Ochrogaster lunifer* were the most similar species, with the closest evolutionary distance. The mitogenome sequence determined in this study will contribute to improved understanding of Notodontidae evolution.

T. *pityocampa* eggs used for studying were collected in October 2017 in Venosta, Italy (46°37’N, 10°46’E) and extracted DNA (YTU-20171001008) was stored at Jiangsu Provincial Key Laboratory of Coastal Wetland Bioresources in Yancheng Teachers University. The National Centre for Biotechnology Information (NCBI) BLAST (http://blast.ncbi.nlm.nih.gov/Blast) and DNAStar packages (DNAStar Inc. Madison, WI) were used to annotate the genome sequence. The MAFFT sequence alignment programme was used to compare genome sequences from *T. pityocampa* and other species (Katoh et al. 2002). Bayesian inference (BI) and maximum likelihood (ML) analyses were performed using the MrBayes version 3.2.1 and IQ-TREE software, respectively. We selected mtMet + F + I + G4 as the best-fit model for amino acid sequences, as determined by the Modelfinder tool using the Bayesian information criterion (BIC). The mitochondrial genome (mitogenome) is considered a powerful marker for resolving phylogenetic relationships (Galtier et al. 2009). To analyse phylogenetic relationships, we obtained the complete mitogenomes of other species from the GenBank database, and aligned the amino acid sequences of the 13 PCGs using ML and BI methods to reconstruct the phylogenetic tree.

Like those of other lepidopterans, the *T. pityocampa* mitochondrial genome (GenBank accession no. MH286070) has 37 functional genes and a long-chain non-coding region (AT enrichment region), at a length of 15,737 bp. The overall
nucleotide composition is: 37.3% T, 40.5% A, 14.6% C, and 7.6% G, demonstrating an AT bias (A + T: 77.8%).

From the BI tree of concatenated amino acid sequences from the 13 PCGs, phylogenetic analysis showed that T. pityocampa is very closely related to *Ochrogaster lunifer* (Figure 1). *Thaumetopoea pityocampa* was phylogenetically distant from two the outgroup species, *Phthonandria atrilineata* and *Biston panterinaria*.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding**

This work was supported by “Science and Technology Research Project of Education Department of Jiangxi Province [grant No. GJJ170925 and GJJ181171]” and “Project of Science and Technology of Jiangxi [grant No. 20192BAB214009, 20171BAB216038]”.

**References**

Berardi L, Pivotto M, Arrigoni G, Mitali E, Trentin AR, Olivieri M, Kerdelhue C, Dorkeld F, Nidelet S, Dubois E, et al. 2017. Proteome analysis of urticating setae from *Thaumetopoea pityocampa* (Lepidoptera: *Notodontidae*). J Med Entomol. 54:1560–1566.

Colacci M, Kavallieratos NG, Athanassiou CG, Boukouvvala MC, Rumbos CI, Kontodimas DC, Pardo D, Sancho J, Benavent-Fernandez E, Galvez-Settier S, et al. 2018. Management of the pine processory moth, *Thaumetopoea pityocampa* (Lepidoptera: *Thaumetopoidea*), in urban and suburban areas: trials with trunk barrier and adhesive barrier trap devices. J Econ Entomol. 111:227–238.

Galtier N, Nabholz B, Glemin S, Hurst GD. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Mol Ecol. 18:4541–4550.

Gschloessl B, Dorkeld F, Berges H, Beydon G, Bouchez O, Branco M, Bretaudeau A, Burban C, Dubois E, Gauthier P, et al. 2018. Draft genome and reference transcriptomic resources for the urticating pine defoliator *Thaumetopoea pityocampa* (Lepidoptera: *Notodontidae*). Mol Ecol Resour. 18:602–619.

Gschloessl B, Vogel H, Burban C, Heckel D, Streiff R, Kerdelhue C. 2014. Comparative analysis of two phenologically divergent populations of the pine processory moth (*Thaumetopoea pityocampa*) by de novo transcriptome sequencing. Insect Biochem Mol Biol. 46:31–42.

Kaszkak I, Planellas M, Dworecka-Kaszkak B. 2015. Pine processory caterpillar, *Thaumetopoea pityocampa* Denis and Schiffmüller, 1775 contact as a health risk for dogs. Ann Parasitol. 61:159–163.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.

Kerdelhue C, Zane L, Simonato M, Salvato P, Rousselet J, Roques A, Battisti A. 2009. Quaternary history and contemporary patterns in a currently expanding species. BMC Evol Biol. 9:220.

Rebollo S, Moneo I, Vega JM, Herrera I, Caballero ML. 2002. Pine processionary caterpillar allergenicity increases during larval development. Int Arch Allergy Immunol. 128:310–314.
Rodriguez-Mahillo AI, Gonzalez-Munoz M, Vega JM, Lopez JA, Yart A, Kerdelhue C, Camafeita E, Garcia Ortiz JC, Vogel H, Petrucco Toffolo E, et al. 2012. Setae from the pine processionary moth (Thaumetopoea pityocampa) contain several relevant allergens. Contact Dermatitis. 67: 367–374.

Rousselet J, Zhao RX, Argal D, Simonato M, Battisti A, Roques A, Kerdelhue C. 2010. The role of topography in structuring the demographic history of the pine processionary moth, Thaumetopoea pityocampa (Lepidoptera: Notodontidae). J Biogeogr. 37:1478–1490.

Vega JM, Moneo I, Ortiz JC, Palla PS, Sanchis ME, Vega J, Gonzalez-Munoz M, Battisti A, Roques A. 2011. Prevalence of cutaneous reactions to the pine processionary moth (Thaumetopoea pityocampa) in an adult population. Contact Dermatitis. 64:220–228.