Mitochondrial phylogeny and comparative mitogenomics of closely related pine moth pests (Lasiocampoidea: *Dendrolimus*) (#35083)

First submission

Editor guidance

Please submit by **13 Mar 2019** for the benefit of the authors (and your $200 publishing discount).

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Raw data check
Review the raw data. Download from the materials page.

Image check
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files
Download and review all files from the materials page.

5 Figure file(s)
5 Table file(s)
3 Other file(s)

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- Data is robust, statistically sound, & controlled.
- Speculation is welcome, but should be identified as such.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

Mitochondrial phylogeny and comparative mitogenomics of closely related pine moth pests (Lasiocampoidea: *Dendrolimus*)

Jie Qin Equal first author, [1] , **Jing Li** Equal first author, [1] , **Qiang Gao** [2] , **John J Wilson** [3] , **Ai-Bing Zhang** Corresp. [1]

[1] College of Life Sciences, Capital Normal University, Beijing, P. R. China.

[2] Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, P. R. China

[3] International College Beijing, China Agricultural University,, Beijing, P. R. China

Corresponding Author: Ai-Bing Zhang
Email address: zhangab2008@mail.cnu.edu.cn

Pine moths, *Dendrolimus* spp. (Lepidoptera; Lasiocampidea), are serious economic pests of conifer forests. Six closely related species (*Dendrolimus punctatus, D. tabulaeformis, D. spectabilis, D. superans, D. houi, and D. kikuchii*) occur in China and cause serious damage to coniferophyte. The complete mito genomes of *Dendrolimus* genus are significant to resolve the phylogenetic relationship and provide theoretical support in pest control. The complete mitogenomes of 3 species (*D. superans, D. houi and D. kikuchii)* were sequenced based on PCR-amplified with universal primers, which were used to amplify initial fragments. The phylogenetic analyses were carried out with 78 complete mitogenomes of lepidopteran species from 10 superfamilies.The complete mitochondrial genomes of these 3 species were 15,417 bp, 15,381 bp and15,377 bp in length, separately. The phylogenetic analyses produced consistent results for six *Dendrolimus* species based on complete mitogenomes, two major clades were formed, one containing *D. spectabilis* clustered with *D. punctatus* + *D. tabulaeformis*, and *D. superans* as the sister group to this three-taxon clade, the other containing *D. kikuchii and D. houi*. Comparative analyses of the congeneric mitochondrial genomes were performed, which showed that non-coding regions were more variable than the A+T rich region. The mitochondrial nucleotide diversity were more variable when compared within than among genus, and the concatenated tRNA region was the most conserved and the nd6 genes was the most variable.

Mitochondrial Phylogeny and Comparative Mitogenomics of Closely Related Pine Moth Pests (Lasiocampoidea: *Dendrolimus*)

Jie Qin[1†], Jing Li[1†], Qiang Gao[2], John James Wilson[3], Ai-bing Zhang[1*]

†Note: These authors equally contributed to the work.

[1] College of Life Sciences, Capital Normal University, Beijing, P. R. China.

[2] Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, P. R. China.

[3] International College Beijing, China Agricultural University, Beijing, P. R. China.

* Corresponding Author:

Ai-bing Zhang[1]

105 West Third Ring Road North, Haidian District, Beijing, 100048, P. R. China.

Email address: zhangab2008@mail.cnu.edu.cn

Abstract

Backgroud. Pine moths, *Dendrolimus* spp. (Lepidoptera; Lasiocampidea), are serious economic pests of conifer forests. Six closely related species (*Dendrolimus punctatus, D. tabulaeformis, D. spectabilis, D. superans, D. houi, and D. kikuchii*) occur in China and cause serious damage to coniferophyte. The complete mitogenomes of *Dendrolimus* genus are significant to resolve the phylogenetic relationship and provide theoretical support in pest control.

Methods. The complete mitogenomes of 3 species (*D. superans, D. houi and D. kikuchii)* were sequenced based on PCR-amplified with universal primers, which were used to amplify initial fragments. Phylogenetic analyses were carried out with 78 complete mitogenomes of lepidopteran species from 10 superfamilies.

Results. The complete mitochondrial genomes of these 3 species were 15,417 bp, 15,381bp and15,377bp in length, separately. The phylogenetic analyses produced consistent results for six *Dendrolimus* species based on complete mitogenomes, two major clades were formed, one containing *D. spectabilis* clustered with *D. punctatus* + *D. tabulaeformis*, and *D. superans* as the sister group to this three-taxon clade, the other containing *D. kikuchii and D. houi*. Comparative

analyses of the congeneric mitochondrial genomes were performed, which showed that non-coding regions were more variable than the A+T rich region. The mitochondrial nucleotide diversity was more variable when compared within than among genus, and the concatenated tRNA region was the most conserved and the nd6 genes was the most variable.

Keywords: Mitogenomic, Phylogeny, *Dendrolimus superans*, *D. houi, D. kikuchii*

Introduction

Pine moths in the genus *Dendrolimus* (Lepidoptera: Lasiocampidae) are major economic pests of coniferous trees, such as *Pinus*, *Larix*, *Picea* and *Abies*, and especially the Masson's pine (*Pinus massoniana*). The caterpillars feed extensively on conifer needles; the resulting damage may reduce the tree's seed yield and can lead to heavy defoliation, dieback, and death (Hou, 1987; Chen, 1990; Zhang *et al*., 2003). During an outbreak period, a pine tree can be consumed in a few days, causing withering and death of pine forests on a large-scale. Furthermore, direct contact with living or dead caterpillars, even their pupae, results in poisoning known as caterpillar arthritis, with serious consequences for human health (Hou, 1987). Twenty-seven species of *Dendrolimus* are known to occur in China, six of them (*D. houi*, *D. kikuchii, D. punctatus*, *D. spectabilis*, *D. superans*, *D. tabulaeformis*) are widely distributed (Hou, 1987; Chen, 1990).

These six major pest species are closely related and their discrimination is challenging (Zhang, Kong & Li, 2004). Morphological diagnoses have proven difficult because many of the characters commonly used to distinguish pine moth species are non-discrete and overlapping amongst the species. Furthermore, some *Dendrolimus* species are sympatric coexistence and sharing similar host plants (Tsai & Liu, 1962). Hybridization experiments and several molecular studies have been conducted, but no consensus has been achieved regarding their species status (Zhao *et al*., 1992; Dai *et al*., 2012; Zhang *et al*., 2014).

Mitochondrial genomes (mitogenomes) have been widely used in phylogenetic, population genetics and comparative genomics studies (Wilson *et al*., 2000; Simon *et al*., 2006; Salvato *et al*., 2008; Cameron, 2014; Qin *et al*., 2015). Insect mitogenomes have relatively stable structure, such as double-stranded, circular DNA molecule, 14-20 kb in size, comprising 37 genes including 13

protein-coding genes (Boore, 1999). Due to its nature of maternal inheritance, mitogenomes has a fast rate of evolution and is particularly useful in phylogenetic analysis (Hebert, Cywinska & Ball., 2003). In addition, whole mitogenome sequences can also provide sets of genome-level characters, such as the relative position of different genes, structural genomic features and compositional features, which could be quite useful in phylogenetic analysis (Thao, Baumann & Baumann.,2004; Masta & Boore., 2008).

Whole mitogenomes instead of several separated gene fragments have been used extensively to construct phylogenies (31, 32), which providing higher support levels (Boore, 2006; Yang *et al*., 2015). Within the order Lepidoptera, multiple studies have used mitogenomes to reconstruct the phylogenetic relationships among and within superfamilies (Whiting *et al*., 1997;Yang *et al*., 2009; Timmermans, Lees & Simonsen., 2014). Technological advancements have triggered rapid increases in the amount of whole mitogenomes, up to 500 of insect mitogenome have been deposited in GenBank (Timmermans, Lees & Simonsen., 2014). However, one of the most recent report shows that only 140 complete Lepidoptera mitogenomes (28 families from 12 superfamilies) have been sequenced, and only 64 are available for moth species (Ramírez-Ríos *et al*., 2016).

The ease and decreased cost of obtaining whole mitogenome sequences has provided the possibility of comparative genomic studies across short evolutionary distances (i.e., congeneric) (Curole & Kocher, 1999) providing an understanding of evolutionary dynamics and trends in a phylogenetic framework.

In this study, six complete mitogenomes from three species (*D. superans*, *D. houi* and *D. kikuchii,* 2 individuals per species) were newly sequenced. These were combined with the complete mitogenomes of three other species (*D. punctatus*, *D. tabulaeformis*, *D. spectabilis*), which have been published previously (Qin *et al*., 2015), to investigate the taxonomic status of species in the genus *Dendrolimus*. To place the relationships within the genus *Dendrolimus* within a broader context, we also conducted phylogenetic analyses of mitogenomes from other 64 lepidopteran species (mainly moth species). In order to investigate the evolutionary dynamics

among six *Dendrolimus* species, comparative analyses were conducted based on 14 mitogenomes (including 2 subspecies of *D. punctatus*), comparing nucleotide composition, codon usage, differences of overlap and non-coding regions.

Materials & Methods

Sample collection, DNA extraction, PCR amplification, sequencing, sequence assembly and annotation

Adult pine moth specimens were sampled at four locations in China (Supplemental Information 1). All specimens were preserved in 95% ethanol in the field and stored at 4℃ in the laboratory until DNA extraction. The specimens were identified by Chun-sheng Wu, Institute of Zoology, Chinese Academy of Sciences, China, using morphological characters. Six individuals of three species (*D. kikuchii*, *D. houi* and *D. superans,* 2 individuals for each species) were selected for sequencing in this study. Total genomic DNA was extracted from thoracic muscle tissue and leg muscle tissue using a DNeasy BLOOD and Tissue kit (QIAGEN) following the manufacturer's protocol.

Mitochondrial genomes were PCR-amplified and sequenced as described in our previous study (Qin *et al*., 2015). In brief, universal primers were used to amplify initial fragments. Specific fragments were then designed to amplify overlapping regions (i.e. primer walking) (Salvato *et al*., 2008; Gissi, Iannelli & Pesole., 2008). PCR recipes and conditions followed Qin *et al*. (2015). All reactions were performed using Takara LA taq (TaKaRa Co., Dalian, China). PCR fragments containing the control region were cloned into the pEASY-T3 Cloning Vector (Beijing TransGen Biotech Co., Ltd., Beijing, China) and then sequenced by using tailed primers, M13-F (CGCCAGGGTTTTCCCAGTCACGAC) and M13-R (GAGCGGATAACAATTT CACACAGG) primers.

Raw sequences were checked manually and assembled on the basis of overlapping regions with the Bioedit V7.0.5 (Caredata.com, Inc.). The tRNA genes were identified by tRNAscan-SE Search Server v.1.21 (Simon *et al*., 1994). Protein-coding and rRNA genes were determined by

comparing homologous sequences with other published Lepidoptera mitochondrial genomes (following (Qin *et al*., 2015)). The sequence data have been deposited in GenBank under accession numbers (KY000409 - KY000414).

Phylogenetic analysis

Phylogenetic analyses were carried out with 78 complete mitogenomes mined from Genbank representing lepidopteran species from 10 superfamilies (Supplemental Information 2). Four mitogenomes of Diptera species were selected as outgroups: *Anopheles darlingi* (NC_014275) (Lowe & Eddy, 1997), *Culex quinquefasciatus* (NC_014574) (Moreno *et al*., 2010), *Cydistomyia duplonotata* (NC_008756) (Behura *et al*., 2011) and *Drosophila yakuba* (NC_001322) (Cameron *et al*., 2007).

Nucleotide sequences of the 13 protein-coding genes were aligned based on the translated amino acid sequences using a customized perl script. Non-protein coding region were aligned using MUSCLE with default settings (Edgar, 2004). The separated genes and partitions were concatenated with SequenceMatrix software (Vaidya, Lohman & Meier, 2011). The concatenated sets of nucleotides were organized into two datasets: dataset 1 representing the 13 protein-coding genes (PCG) only and dataset 2 representing 37 genes (13 PCGs + 22 transfer RNA genes (tRNA) + 2 ribosomal RNA genes (rRNA)). Substitution saturations of 2 datasets were tested with software DAMBE (Xia & Xie, 2001), and both datasets were used in phylogenetic analyses, under the optimality criteria of maximum likelihood (ML) and Bayesian inference (BI) (Ronquist & Huelsenbeck, 2003).

In order to standardize the partitioning strategy as recommended for phylogenetic analyses with mitogenomes (Zardoya & Meyer, 1996), PartitionFinder v1.1.1software was used to select the optimal partitioning scheme and to find the best-fitting substitution model for each partition under the Bayesian Information Criterion (Lanfear *et al*., 2012). Not only that, optimized nucleotide substitution models could avoid being affected by the long branch attraction to some extent (Bergsten, 2005). The maximum possible partition scheme was 15 partitions: each protein-coding gene as a separate partition, the concatenated 22 tRNA genes and the concatenated rRNA

genes).

ML analysis was performed with RAxML v7.9.6 and BI analysis with a parallel version of MrBayes v 3.2.2 (Stamatakis, 2006; Ronquist *et al*., 2012). The GTR+G+I model was selected for each partition in the two datasets. Support values for the ML topologies were evaluated via bootstrap tests with 1000 iterations (in RaxML). BI analysis was conducted with two sets of four independent Markov chains run for 10 million Metropolis-coupled (MCMC) generations, with tree sampling occurring every 1000 generations, and burn-in set to 25% of the trees. After 10 million generations, all runs reached stationary as determined by the program Tracer v1.5.0 (Rambaut & Drummond, 2007).

Genetic distance analysis among closely related species of *Dendrolimus*

In order to test the intraspecific and interspecific differentiation of *Dendrolimus,* 14 mitogenomic were used to calculate the genetic distance across the two datasets described above, which including two subspecies of *D. punctatus* (*D. punctatus punctatus* and *D. punctatus wenshanensis)* and other 5 species. Genetic distances were calculated using the GTR model selected as the best model by AIC (Akaike information criterion) which performed with Modeltest 3.7 (Posada & Buckley,2004; Ronquist *et al*., 2012). Genetic distances were calculated using a custom C++ script that uses the bio++ function library (Guéguen *et al*., 2013). A correlation matrix was also estimated according to obtained genetic distance matrix. Correlation values ranged from -1 to 1, where values closer to 1 are indicative of a closer relationship. A graphical visualization of the genetic distances and correlation matrix was drawn using the corrplot.mixed function in R package (Wei, 2013).

Comparative mitogenome analyses of *Dendrolimus*

Nucleotide composition, codon usage (excluding stop codons) and Relative Synonymous Codon Usage (RSCU) were calculated across 14 mitogenomes of *Dendrolimus* with MEGA 5.0 (Tamura, 2011). Composition skew was calculated using the formulae: AT skew = (A-T)/ (A+T) and GC skew = (G-C)/ (G+C) (Perna & Koche, 1995). Sliding window analyses were used to calculate nucleotide diversity values across protein-coding genes and regions, which executed with DnaSP software (Librado & Rozas, 2009). The window size and step size were set to 100bp and 25bp,

167 separately.

168

169 **Results & Discussion**

170 **Phylogenetic analyses**

171 Phylogenetic analyses of *Dendrolimus* resulted in a fully resolved tree with robust support for
172 nearly all nodes (Figure 1). Phylogenetic analyses inferred from different datasets exhibited the
173 same topology. Six species formed two major clades: *D. punctatus* + *D. tabulaeformis* + *D.*
174 *spectabilis* + *D. superans* (Clade 1) was the sister group to *D. kikuchii* and *D. houi* (Clade 2).
175 Within 'Clade 1', *D. spectabilis* clustered with *D. punctatus* + *D. tabulaeformis*, and *D. superans*
176 was the sister group to this three-taxon clade.

177 The topology of our mitogenome *Dendrolimus* phylogeny showed some differences from the
178 topology proposed by previous studies. Zhang *et al*. (2014) constructed a phylogeny of
179 *Dendrolimus* based on one pheromone-binding proteins (PBPs) and two general odorant-binding
180 proteins (OBPs) which *D. houi* was proposed as a basal species of *Dendrolimus*. But complete
181 mitogenomes provide more information than OBPs (266 - 381bp). However, the relationships of
182 *D. tabulaeformis*, *D. punctatus*, and *D. spectabilis* were verified with mitogenomes analysis,
183 sharing a closer relationship to each other with respect to *D. superans.*

184 In the phylogenetic analyses of 78 moth mitogenomes the monophyly of each superfamily
185 was generally well-supported and there was consistency with prior studies (Yang *et al*., 2009;
186 Kawahara & Breinholt., 2014; Qin *et al*., 2015). In our study, Lasiocampoidea and Bombycoidea
187 were monophyletic and clustered together as sister groups with high support. Previous studies have
188 included Lasiocampidae within Bombycoidea (Brock, 1971; Scoble, 1992; Kawahara &
189 Breinholt., 2014), while other studies have treated Lasiocampidae as a distinct superfamily
190 Lasiocampoidea (Minet, 1991; Regier *et al*., 2009; van Nieukerken *et al*., 2011; Bazinet *et al*.,
191 2013).

192 Similar trees were obtained based on both datasets, the only difference was among the
193 superfamilies Bombycoidea, Geometroidea, Lasiocampoidea and Noctuoidea, which altogether

194 constitute approximately 73 000 described species (Minet, 1991). The 13 PCG dataset phylogeny
195 placed Geometroidea with Bombycoidea and Lasiocampoidea, and Noctuoidea as the sister group
196 to this three-taxon clade (53% BP support and 0.78 posterior probabilities) (Figure 1), which
197 revealed similar relationship with a prior study (van Nieukerken *et al*., 2011). Nonetheless, the 37
198 gene dataset phylogeny (Supplemental Information 3) placed Bombycoidea + Lasiocampoidea as
199 the sister group to Geometroidea + Noctuoidea with higher branch support (100% BP support and
200 1.0 posterior probabilities). The latter relationship was demonstrated with morphological and
201 multigenetic proofs (van Nieukerken *et al*., 1758; Regier *et al*., 2009; Bazinet *et al*., 2013;
202 Kawahara & Breinholt, 2014).
203 Within the Bombycoidea, the relationship among the families Bombycidae, Sphingidae and
204 Saturniidae has been difficult to resolve in previous study (Regier *et al*., 2013). In our study, the
205 analysis of both datasets placed the Bombycidae as the sister group to Saturniidae and Sphingidae
206 with high support (100% bootstrap), which is consistent with the phylogenetic relationship based
207 on transcriptomic data (2696 genes) (Breinholt & Kawahara, 2013).
208 **Genetic distance analyses**
209 The genetic distance analyses produced results which were consistent with the results of the
210 phylogenetic analyses. The correlation values obtained from genetic distance analysis among
211 specimens of *Dendrolimus* showed that in many cases intraspecific and interspecific values were
212 very similar. Values for intraspecific and interspecific correlations in the group comprising *D.*
213 *tabulaeformis* and two subspecies of *D. punctatus* were equal or very close to 1, which suggests
214 these sequences all have quite a few differences, which would generally be regarded within the
215 range of intraspecific variation. To illustrate the relationship of *Dendrolimus* more clearly, we re-
216 calculated genetic distance with considered *D. punctatus* and *D. tabulaeformis* as an integral taxon
217 (Group A). The genetic distance between *D. spectabilis* and Group A were 0.05, whereas *D.*
218 *superans* and Group A were 0.07 (Figure 2). Furthermore, both the correlation value between *D.*
219 *houi* - Group A and *D. kikuchii* - Group A were negative, highlighting the relatively distant genetic
220 relationship with other four species (*D. punctatus*, *D. tabulaeformis*, *D. spectabilis* and *D.*

221 *superans*) (Figure 2).

222 **Comparative mitochondrial genome characterization of *Dendrolimus***

223 **(i) Mitochondrial genome organization**

224 The complete mitochondrial genomes of *Dendromlius* ranged from 15,370 to 15,417 bp in length
225 (Table 1). The gene order was identical to other ditrysian lepidopterans with the standard *trnM*
226 gene location type (*trnM-trnI-trnQ*), and all mitochondrial genomes exhibit similar sequence
227 characteristics. The mitochondrial genes of three newly-sequenced *Dendrolimus* species (*D.*
228 *superans*, *D. houi* and *D. kikuchii*) are coded on the majority strand, except for four protein-coding
229 gene (*nd5, nd4, nd4L* and *nd1*) and eight tRNA genes (*trnQ, trnC, trnY, trnF, trnH, trnP,*
230 *trnL(CUN),*and *trnV*) (Table 1).

231 **(ii) Base composition and skewness**

232 Metazoan mitogenomes usually exhibit a clear strand bias toward adenine (A) and thymine (T) in
233 nucleotide composition. Consistent with previous observations of *Dendrolimus* mitogenomes, the
234 mitochondrial sequence of three newly-sequenced *Dendrolimus* species were biased toward A and
235 T. The A+T content of the majority strand ranged from 78.7% and 78.8% for *D. kikuchii*, 80% and
236 79.9% for *D. houi*, and 80.1% and 80.2% for *D. superans* (Supplemental Information 4). The
237 strand bias also can be measured as AT- and GC-skews. The average AT-skew across all available
238 *Dendrolimus* mitochondrial genomes was 0.028, ranging from 0.037 to 0.017, whereas the average
239 GC-skew of the *Dendrolimus* mitochondrial genomes was -0.23, ranging from -0.26 to -0.22.

240 **(iii) Start and stop codon usage**

241 Start and stop codon usage is an important characteristic in the annotation of protein-coding genes.
242 We compared the start and stop codons across the six species of *Dendrolimus* (Table 2). All
243 protein-coding genes started with the typical ATN codons except for *cox1* which used CGA. Most
244 of the start codon were consistent within the six species but a few were different (*nd2, cox2, atp8,*
245 *nd3, nd5, nd1*). This was especially the case for *atp8* and *nad3*, which were the most variable
246 among the genes. It is noteworthy that *atp8* and *nad3* are the shortest protein-coding genes when
247 compare to others in the mitochondrial genome, suggesting variability in start codon usage maybe

248 related to gene length.

249 Nine genes (*nd2, atp8, atp6, cox3, nd5, nd4l, nd6, cob, nd1*) share the same complete stop
250 codon TAA, and four genes use incomplete stop codons (*cox1, cox2 and nd4, nd3*) (Table2).
251 Incomplete stop codons are common in lepidopteran mitogenomes and are presumed to be
252 completed via post-transcriptional polyadenylation (Chen *et al*., 2016). Changes in stop codon
253 usage among *Dendrolimus* were rarer than changes in start codon usage. Only in the *cox2* and
254 *nad2* genes, did we observe changes in the stop codon used. Therefore, we can conclude that even
255 within congeneric species, start and stop codons are variable in the mitochondrial genome.

256 **(iv) Codon usage and RSCU**

257 Condon usage and relative synonymous codon usage (RSCU) results were compared across all
258 available *Dendrolimus* mitogenomes (Figure 3). The analysis showed that Leu2 (UUR), Ile, Phe,
259 Met, Asn, Gly, Ser2 (UCN), Tyr are the eight most frequent amino acids and were represented by
260 at least 50 codons per thousand codons. Two codon families, Leu2 and Ile, had at least 100 codons
261 per thousand codons. Leu2, a hydrophobic amino acid, was significantly more frequent than other
262 amino acids, which may relate to the function of chondriosomes in many transmembrane proteins.
263 The rarest used codon family was Cys.

264 The usage of both two-fold and four-fold degenerate codons was biased towards the use of
265 codons with A or T in third position (Figure 4). Codons which have relatively high G and C content
266 are likely to be abandoned, reflecting a finding across other lepidopteran insects. Examination of
267 the fourteen individual *Dendrolimus* mitogenomes showed that Leu2 (UUA), Ser2 (UCU), Arg
268 (CGA), Ala (GCU), Ser1 (AGA) are the five most frequent relative synonymous codons.

269 **(v) Non-coding regions, overlapping regions and A+T rich region**

270 All fourteen mitogenomes had six overlapping regions and the size ranged from 1 to 8 bp (Table
271 3). Nucleotide sequence of six mutual overlapping areas were almost identical, except for the
272 overlap between *nd2* and *trnW* in *D. kikuchii* which was 1bp shorter than other species of
273 *Dendrolimus.* In addition to the control region, there were 17 non-coding regions in the
274 mitogenomes of *D. punctatus*, *D. tabulaeformis* and *D. spectabilis*, 18 in *D. superans,* 16 in *D.*

kikuchii and 19 in *D. houi* (Table 3). It is noteworthy that there are 6 intergenic regions, *trnQ-nad2* (54 bp - 58 bp), *trnY-cox1* (25 bp - 41 bp), *atp6-cox3* (6 bp - 15 bp), *trnA-trnR* (9 bp - 20 bp), *trnN-trnS* (AGN) (7 bp - 25 bp), *nad4-nd4l* (19 bp - 38bp), were longer than 15 bp.

The largest intergenic spacer of whole mitogenome is the A + T rich region, which not only has the characteristics of non-coding genes, but also contains important sites for the regulation of transcription and replication (Gissi *et al*., 2008). The A+T rich region and intergenic regions might contain useful phylogenetic signals, particularly for determining congeneric relationships and relationships among recently diverged species. To investigate the utility, we constructed a phylogenetic tree of *Dendrolimus* species using only the A + T rich region and intergenic regions (Supplemental Information 5). The phylogenetic analysis using the A+T rich region produced similar but slightly different topology comparing with the whole mitogenomes. This suggests the intergenic regions might be too variable to be useful for phylogenetic analyses, nevertheless, the A+T rich region might be an effective molecular tool in solving phylogenetic relationships among recently diverged species.

Sliding-window analysis

Sliding-window analysis was conducted to compare nucleotide diversity among the mitochondrial protein-coding genes and non-coding regions of 14 individuals in *Dendrolimus* (Figure 5). The intergenic region has the highest nucleotide diversity which is likely attributable to the large indels in this region. This was followed by *nd6, cytb, cox2, atp6, cox3, nd3*, A+T rich region, *nd1, cox1, nd2, nd5, nd4, nd4l, atp8*, rRNA, tRNA. It is notable that the nucleotide diversity of the A+T rich region was moderate; lower than many protein-coding genes. The tRNA was the most conserved region and *cox1* was the most conserved protein-coding gene. In contrast, sliding-window analyses using all 78 lepidopteran mitogenomes (same dataset as the phylogenetic analyses) produced substantially similar patterns: the *nd6* gene had the highest level of divergence and tRNA was the most conserved region, while the *cox1* was the most conserved than all protein-coding genes.

Conclusion

In this study, both phylogenetic and genetic distance analyses obtained consistent results regarding the relationships among six closely related species. The whole mitogenomes failed to provide enough information to distinguish *D. tabulaeformis* from *D. punctatus*, which suggest there might not be a clear species boundary between these two species. This finding is consistent with the results of previous studies, in which *D. tabulaeformis* was regarded as ecological type of *D. punctatus* based on several DNA markers and experiments of interspecific hybridization. Meanwhile, *D. spectabilis* fell as sister to these two sibling species, and *D. superans* fell as sister to these three taxa. *D. kikuchii* and *D. houi* are sister species, having relatively close relationship comparing with other four species.

Congeneric species exhibit similar mitochondrial genome features, such as genome organization, nucleotide composition, codon usage and RSCU. Within the genus *Dendrolimus*, start and stop codons were variable in mitochondrial genome and the change of stop codons were rarer than start codons. Non-coding regions were the most variable regions in mitochondrial genomes. When comparing nucleotide diversity, the *nad6* gene had the highest level of divergence and the tRNA region was the most conserved.

Acknowledgements

The authors wish to thank Prof. Xiang-Bo Kong (Chinese Academy of Forestry) for helping with sample collection. Funding for this research was provided by the National Natural Science Foundation of China (Grant No.31425023, 31400191, 31601877 and 31772501), Beijing Municipal Natural Science Foundation (Grant No.5172005).

Competing interests

The authors have declared that no competing interests exists.

Reference:

1. Bazinet AL, Cummings MP, Mitter KT, Mitter CW. 2013. Can RNA-Seq resolve the rapid radiation of advanced moths and butterflies (Hexapoda: Lepidoptera: Apoditrysia)? An

exploratory study. *PLOS ONE* 8: e82615. DOI 10.1371/journal.pone.0082615

2. Behura SK, Lobo NF, Haas B, Lovin DD, Shumway MF, Puiu D, Romero-Severson J, Neneb V, Seversona DW. 2011. Complete sequences of mitochondria genomes of Aedes aegypti and Culex quinquefasciatus and comparative analysis of mitochondrial DNA fragments inserted in the nuclear genomes. Insect Biochemistry and Molecular Biology 41: 770-7. DOI 10.1016/j.ibmb.2011.05.006

3. Bergsten J. 2005. A review of long-branch attraction. Cladistics 21(2): 163-193. DOI10.1111/j.1096-0031.2005.00059.x

4. Boore JL. 1999. Animal mitochondrial genomes. Nucleic Acids Research 27: 1767-80. DOI 10.1093/nar/27.8.1767

5. Boore JL. 2006. The use of genome-level characters for phylogenetic reconstruction. Trends in Ecology and Evolution 21: 439-46. DOI 10.1016/j.tree.2006.05.009

6. Breinholt JW, Kawahara AY. 2013. Phylotranscriptomics: saturated third codon positions radically influence the estimation of trees based on next-gen data. Genome Biology *and Evolution* 5: 2082-92. DOI 10.1093/gbe/evt157

7. Brock J. 1971. A contribution towards an understanding of the morphology and phylogeny of the Ditrysian Lepidoptera. *Journal of Natural History* 5: 29-102. DOI 10.1080/00222937100770031

8. Cameron SL. 2014. Insect mitochondrial genomics: implications for evolution and phylogeny. *Annual Review of Entomology* 59: 95-117. DOI 10.1146/annurev-ento-011613-162007

9. Cameron SL, Lambkin CL, Barker SC, Whiting MF. 2007. A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. *Systematic Entomology* 32: 40-59. DOI 10.1111/j.1365-3113.2006.00355.x

10. Chen C, Qiang Y, Peng XY, Qian ZQ, Wang ZZ. 2016. The complete mitochondrial genome of the Sara Longwing Heliconius sara (Insecta: Lepidoptera: Nymphalidae). Mitochondrial DNA Part A 27: 3167-3168.

11. Chen CJ. 1990. In Integrated Management of Pine Caterpillar in China. Beijing, China: China Forestry Press. (in Chinese)

12. Curole JP, Kocher TD. 1999. Mitogenomics: digging deeper with complete mitochondrial genomes. Trends in Ecology and Evolution 14: 394-8. DOI 10.1016/S0169-5347(99)01660-2

13. Dai QY, Gao Q, Wu CS, Chesters D, Zhu CD, Zhang AB. 2012. Phylogenetic reconstruction and DNA barcoding for closely related pine moth species (*Dendrolimus*) in China with multiple gene markers. *PLOS ONE* 7: e32544. DOI 10.1371/journal.pone.0032544

14. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32(5): 1792–1797. DOI 10.1093/nar/gkh340.

15. Gissi C, Iannelli F, Pesole G. 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity* 101: 301-20.

16. Guéguen L, Gaillard S, Boussau B, Gouy M, Groussin M, Rochette NC, Bigot T, Fournier D, Pouyet F, Cahais V, Bernard A, Scornavacca C, Nabholz B, Haudry A, Dachary L, Galtier N, Belkhir K, Dutheil JY. 2013. Bio++: efficient extensible libraries and tools for computational molecular evolution. *Molecular Biology and Evolution* 30: 1745-50. DOI 10.1093/molbev/mst097

17. Hebert PD, Cywinska A, Ball SL. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences* 270: 313-21. DOI 10.1098/rspb.2002.2218

18. Hou T. 1987. The Pine Caterpillars in China., Beijing, China:Science Press. (in Chinese)

19. Kawahara AY, Breinholt JW. 2014. Phylogenomics provides strong evidence for relationships of butterflies and moths. *Proceeding of the Royal Society B: Biological Sciences* 281:20140970. DOI 10.1098/rspb.2014.0970

20. Lanfear R, Calcott B, Ho SY, Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* 29: 1695-701.56. DOI 10.1093/molbev/mss020

21. Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-2. DOI 10.1093/bioinformatics/btp187

22. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* 25: 955-64.

23. Masta SE, Boore JL. 2008. Parallel evolution of truncated transfer RNA genes in arachnid mitochondrial genomes. *Molecular Biology and Evolution* 25: 949-59. DOI 10.1093/molbev/msn051

24. Minet J. 1991. Tentative reconstruction of the ditrysian phylogeny (Lepidoptera: Glossata). *Insect Systematics and Evolution* 22: 69-95.

25. Moreno M, Marinotti O, Krzywinski J, Tadei WP, James AA, Achee NL,Conn JE. 2010. Complete mtDNA genomes of *Anopheles darlingi* and an approach to anopheline divergence time. *Malaria Journal* 9: 127. DOI 10.1186/1475-2875-9-127

26. van Nieukerken EJ, Kaila L, Kitching IJ, Kristensen NP, Lees D, Minet J. 2011. Order Lepidoptera Linnaeus, 1758. *Zootaxa* 3148: 212-21.

27. Perna NT, Kocher TD. 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *Journal of Molecular Evolution* 41: 353-8. DOI 10.1007/BF00186547

28. Posada D, Buckley TR. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology* 53: 793-808. DOI 10.1080/10635150490522304

29. Qin J, Zhang Y, Zhou X, Kong X, Wei S, Ward RD, Zhang AB. 2015. Mitochondrial phylogenomics and genetic relationships of closely related pine moth (Lasiocampidae: *Dendrolimus*) species in China, using whole mitochondrial genomes. *BMC Genomics* 16: 1. DOI 10.1186/s12864-015-1566-5

30. Rambaut A, Drummond A. 2007. http://beast. bio. ed. ac. uk/Tracer. Tracer v1. 4.

31. Ramírez-Ríos V, Franco-Sierra ND, Alvarez JC, Saldamando-Benjumea CI, Villanueva-Mejía DF. 2016. Mitochondrial genome characterization of Tecia solanivora (Lepidoptera:

Gelechiidae) and its phylogenetic relationship with other lepidopteran insects. *Gene* 581:107-116.

32. Regier JC, Zwick A, Cummings MP, Kawahara AY, Cho S, Weller S, Roe A, Baixeras J, Brown JW, Parr C, Davis DR, Epstein M, Hallwachs W, Hausmann. A, Janzen DH, Kitching IJ, Solis MA, Yen SH, Bazinet AL, Mitter C. 2009. Toward reconstructing the evolution of advanced moths and butterflies (Lepidoptera: Ditrysia): an initial molecular study. *BMC Evolutionary Biology* 9: 280. DOI 10.1186/1471-2148-9-280
33. Regier JC, Mitter C, Zwick A, Bazinet AL, Cummings MP, Kawahara AY, Sohn JC, Zwickl DJ, Cho S, Davis DR, Baixeras J, Brown J, Parr C, Weller S, Lees DC, Mitter KT. 2013. A large-scale, higher-level, molecular phylogenetic study of the insect order Lepidoptera (moths and butterflies). *PLOS ONE* 8: e58568. DOI 10.1371/journal.pone.0058568
34. Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572-4. DOI 10.1093/bioinformatics/btg180
35. 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. *Systematic Biology* 61: 539-42. DOI 10.1093/sysbio/sys029
36. Salvato P, Simonato M, Battisti A, Negrisolo E. 2008. The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae). *BMC Genomics* 9: 331. DOI 10.1186/1471-2164-9-331
37. Scoble MJ. 1992. The Lepidoptera. Form, function and diversity: Oxford University Press.
38. Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society Of America* 87: 651-701. DOI 10.1093/aesa/87.6.651
39. Simon C, Buckley TR, Frati F, Stewart JB, Beckenbach AT. 2006. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain

reaction primers for animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics* 37: 1 545-79. DOI 10.1146/annurev.ecolsys.37.091305.110018

40. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688-90. DOI 10.1093/bioinformatics/btl446

41. Tamura K. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood , evolutionary distance , and maximum parsimony methods. *Molecular Biology and Evolution* 28: 2731–9. DOI 10.1093/molbev/msr121

42. Timmermans MJ, Lees DC, Simonsen TJ. 2014. Towards a mitogenomic phylogeny of Lepidoptera. *Molecular Phylogenetics and Evolution* 79: 169-78. DOI 10.1016/j.ympev.2014.05.031

43. Tsai P, Liu Y. 1962. A study of the genus *Dendrolimus* (Lasiocampidae) of China with descriptions of two new species and one new subspecies. *Acta Entomol Sin* 11: 237-52.

44. Vaidya G, Lohman DJ, Meier R. 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* 27: 171-80. DOI 10.1111/j.1096-0031.2010.00329.x

45. Wei T. 2013. corrplot: Visualization of a correlation matrix. R package version 073.

46. Whiting MF, Carpenter JC, Wheeler QD, Wheeler WC. 1997. The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. *Systematic Biology* 46: 1-68. DOI 10.1093/sysbio/46.1.1

47. Wilson K, Cahill V, Ballment E, Benzie J. 2000. The complete sequence of the mitochondrial genome of the crustacean Penaeus monodon: are malacostracan crustaceans more closely related to insects than to branchiopods? *Molecular Biology and Evolution* 17: 863-874. DOI 10.1093/oxfordjournals.molbev.a026366

48. Xia X, Xie Z. 2001. DAMBE: Software Package for Data Analysis in Molecular Biology and Evolution. Journal of *Heredity* 92(4): 371-373. DOI:10.1093/jhered/92.4.371.

49. Yang L, Wei ZJ, Hong GY, Jiang ST, Wen LP. 2009. The complete nucleotide sequence of

the mitochondrial genome of Phthonandria atrilineata (Lepidoptera: Geometridae). Molecular Biology Reports 36: 1441-9. DOI 10.1007/s11033-008-9334-0

50. Yang X, Cameron SL, Lees DC, Xue D, Han H. 2015. A mitochondrial genome phylogeny of owlet moths (Lepidoptera: Noctuoidea), and examination of the utility of mitochondrial genomes for lepidopteran phylogenetics. Molecular Phylogenetics and Evolution 85: 230-7. DOI 10.1016/j.ympev.2015.02.005

51. Zardoya R, Meyer A. 1996. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. *Molecular Biology and Evolution* 13: 933-42. DOI 10.1093/oxfordjournals.molbev.a025661

52. Zhang A, Kong X, Li D, Liu Y. 2003. DNA fingerprinting evidence for the phylogenetic relationship of eight species and subspecies of *Dendrolimus* (Lepidoptera: Lasiocampidae) in China. *Acta entomologica Sinica* 47: 236-42.

53. Zhang A, Zhang Z, Wang H, Kong X. 2003. Geographical distribution of Lasiocampidae in China and its relationship with environmental factors. Journal of Beijing Forestry University. 26: 54-60.

54. Zhang SF, Zhang Z, Kong XB, Wang HB. 2014. Molecular characterization and phylogenetic analysis of three odorant binding protein gene transcripts in Dendrolimus species (Lepidoptera: Lasiocampidae). *Insect science* 21: 597-608. DOI 10.1111/1744-7917.12074

55. Zhao QS, Wu WB, Lu GP, Yuan X, Li SK, Jiang JC. 1992. Hybridization experiments with two species of *Dendrolimus*. *Acta Entomologica Sinica* 35:29-32.

Figure 1

Phylogenetic relationship of six *Dendrolimus* species

Figure 1: (A) Cladogram (ML and BI) depicting six *Dendrolimus* species constructed with Maximum Likelihood and Bayesian inference analyses of (i) 13 protein coding genes (13PCGs); (ii) 37 genes (13 protein-coding genes+22 transfer RNA genes+2 ribosomal RNA genes, 37gene). Numbers above or below branches indicate posterior probabilities and bootstrap percentages across the difference analyses and datasets (13PCGs-BI / 13PCGs-ML / 37gene-BI / 37gene-ML). (B) Cladogram constructed using Bayesian inference analysis of nucleotide sequences of 13 mitochondrial protein-coding genes of Lepidopteran (moth) species, plus outgroups. Numbers above or below branches indicate posterior probabilities.

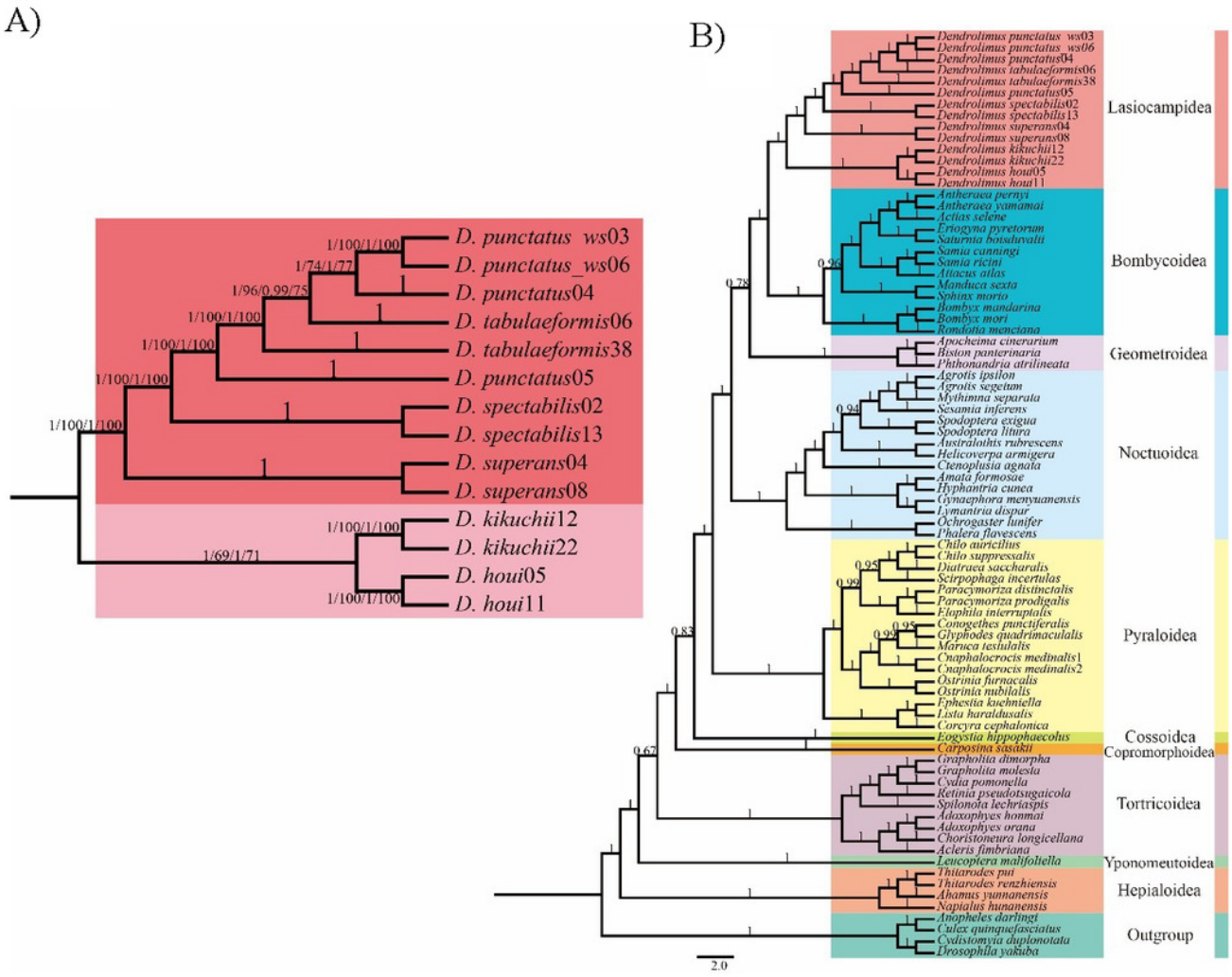

Figure 1. (A) Cladogram (ML and BI) depicting six Dendrolimus species constructed with maximum likelihood and Bayesian inference analyses of (i) 13 protein coding genes (13PCGs); (ii) 37 genes (13 PCGs + 22 tRNA genes + 2 rRNA genes, 37gene). Numbers above or below branches indicate posterior probabilities and bootstrap percentages across the difference analyses and datasets (13PCGs-BI / 13PCGs-ML / 37gene-BI / 37gene-ML). (B) Cladogram constructed using Bayesian inference analysis of nucleotide sequences of 13 mitochondrial protein-coding genes of lepidopteran (moth) species, plus outgroups. Numbers above or below branches indicate posterior probabilities.

Figure 2

Genetic distance of six *Dendrolimus* species

Figure 2 Genetic distance (below diagonal) and correlation relationship (above diagonal) of 13 concatenate protein coding genes (Left) and 37 concatenate genes (Right). The size of circle stands for the correlation values, which range from -1 to 1. Values closer to 1 indicate a closer relationship. Species names were abbreviated: *D. spectabilis* (CS02 and CS13), *D. tabulaeformis* (YS06 and YS08), *D. punctatus punctatus* (MW04 and MW05), *D. punctatus wenshanensis* (WS03 and WS06), *D. superans* (LY04 and LY08), *D. kikuchii* (SM12 and SM22), *D. houi* (YN05 and YN11).

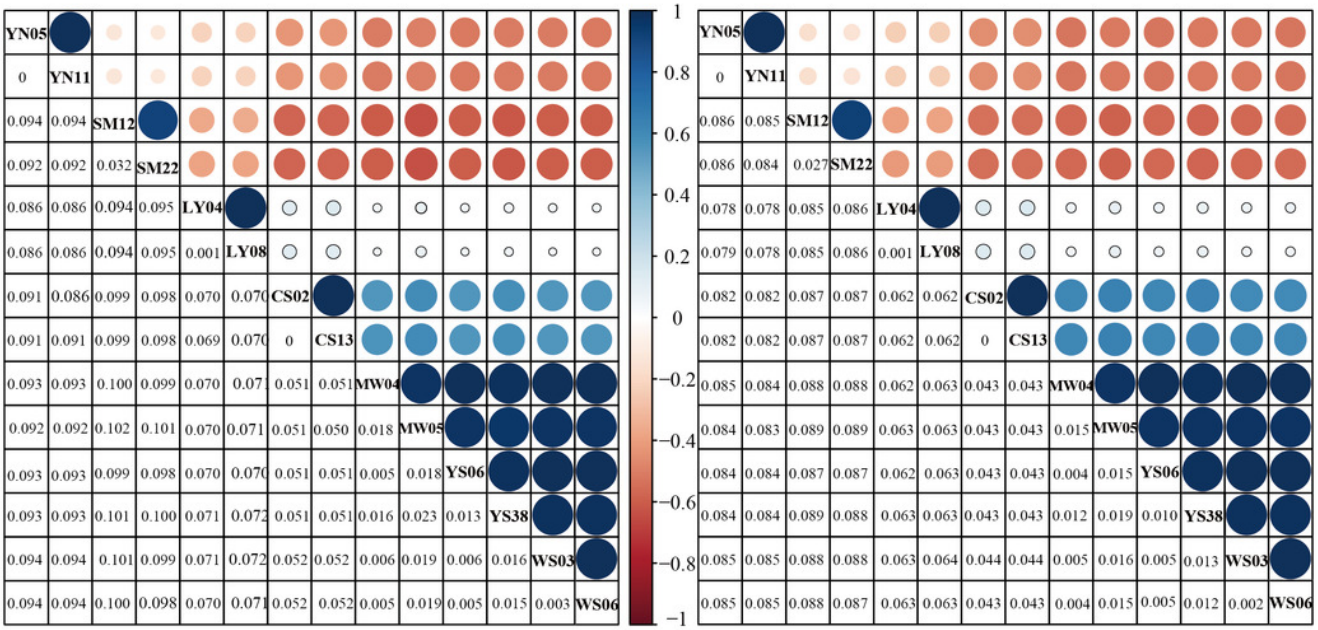

Figure 2. Genetic distance (below diagonal) and correlation relationship (above diagonal) of 13 concatenate protein coding genes (Left) and 37 concatenate genes (Right). The size of circle stands for the correlation values, which range from -1 to 1. Values closer to 1 indicate a closer relationship. Species names were abbreviated: *D. spectabilis* (CS02 and CS13), *D. tabulaeformis* (YS06 and YS08), *D. punctatus punctatus* (MW04 and MW05), *D. punctatus wenshanensis* (WS03 and WS06), *D. superans* (LY04 and LY08), D. kikuchii (SM12 and SM22), *D. houi* (YN05 and YN11).

Figure 3

Relative synonymous codon usage (RSCU) of 14 *Dendrolimus* mitochondrial genomes.

Figure 3 Relative synonymous codon usage (RSCU) of fourteen *Dendrolimus* mitochondrial genomes. Codon Families are provided on the x axis. Codons that are absent in the mitochondrial genomes are marked at the top of columns. Leu1 stands for Leu (CUN); Leu2 stands for Leu (UUR); Ser1 stands for Ser (AGN); Ser2 stands for Ser (UCN).

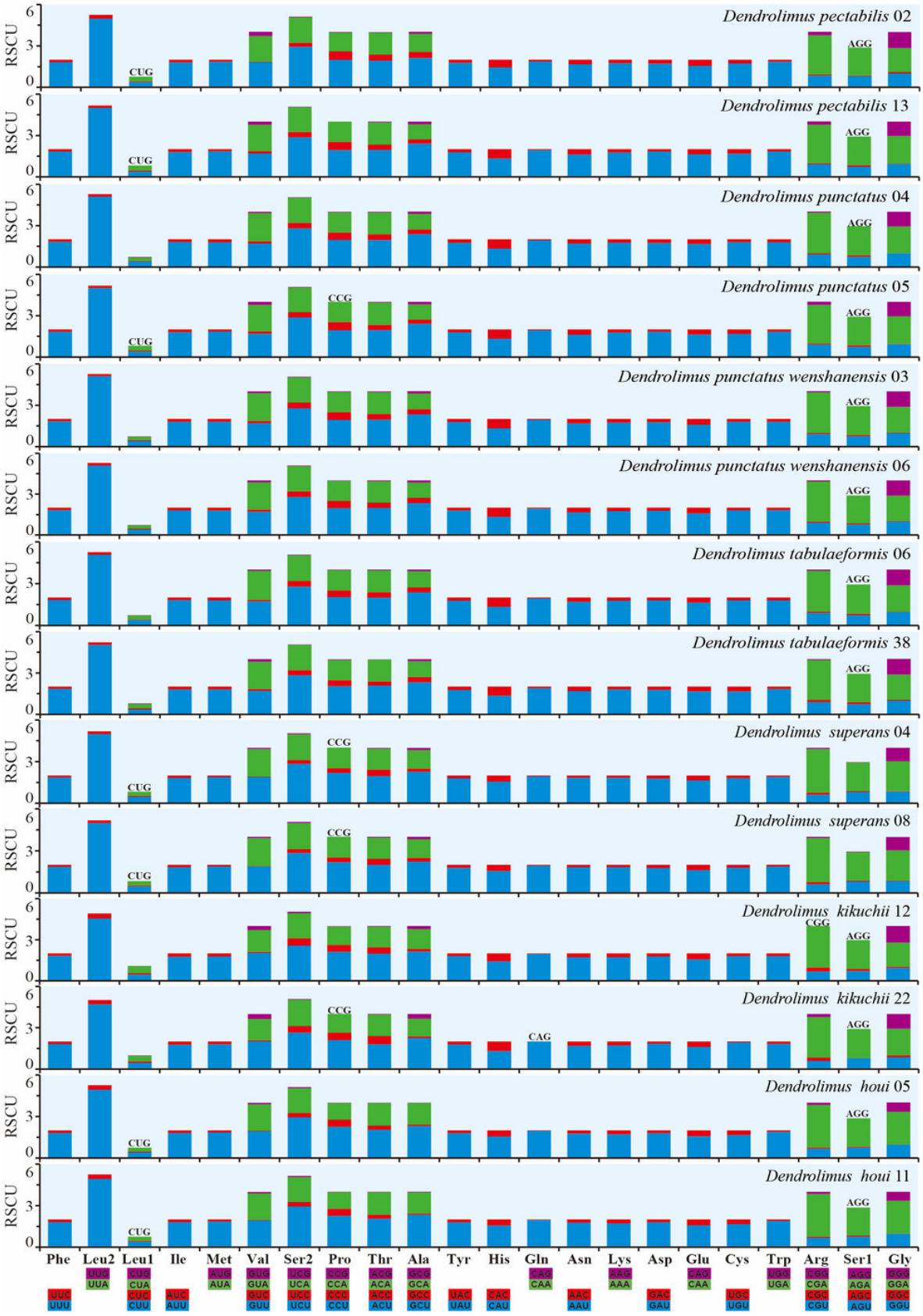

Figure 3 Relative synonymous codon usage (RSCU) of fourteen Dendrolimus mitochondrial genomes. Codon Families are provided on the x axis. Codons that are absent in the mitochondrial genomes are marked at the top of columns. Leu1 stands for Leu (CUN); Leu2 stands for Leu (UUR); Ser1 stands for Ser (AGN); Ser2 stands for Ser (UCN).

Figure 4

Codon usage of 14 *Dendrolimus* mitochondrial genomes.

Figure 4 Codon usage of fourteen *Dendrolimus* mitochondrial genomes. Numbers above the column refer to the number of codons. CDspT stands for codons per thousand codons. Codon Families are provided on the x axis. Leu1 stands for Leu (CUN); Leu2 stands for Leu (UUR); Ser1 stands for Ser (AGN); Ser2 stands for Ser (UCN).

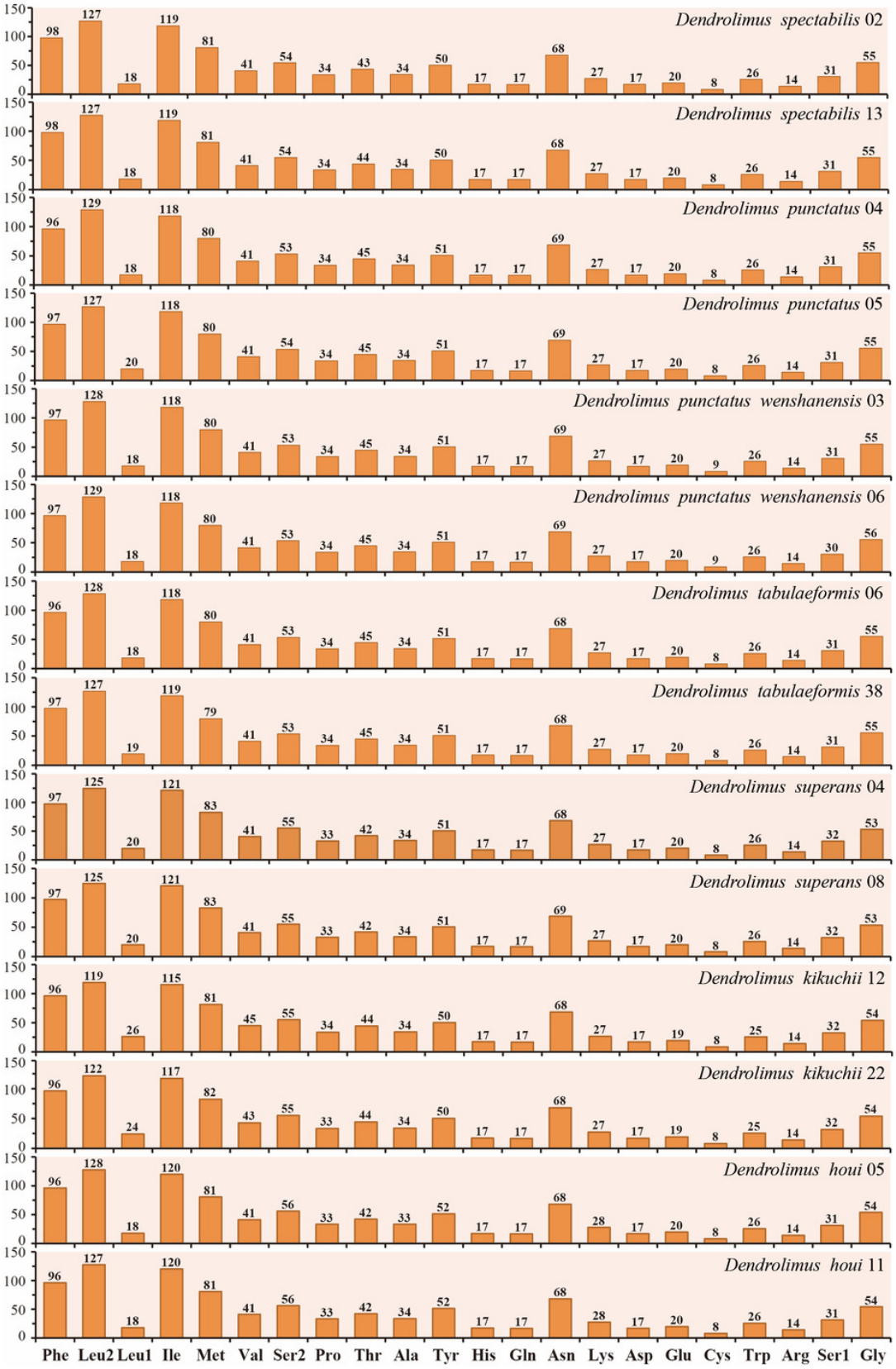

Figure 4. Codon usage of fourteen Dendrolimus mitochondrial genomes. Numbers above the column refer to the number of codons. CDspT stands for codons per thousand codons. Codon Families are provided on the x axis. Leu1 stands for Leu (CUN); Leu2 stands for Leu (UUR); Ser1 stands for Ser (AGN); Ser2 stands for Ser (UCN).

Figure 5

Sliding-window analyses of 13 protein coding genes.

Figure 5 Sliding-window analyses of 13 protein coding genes, concatenated tRNA and rRNA genes, intergenic and A+T rich region among six *Dendrolimus* species. The X-axis represents sequence length, the Y-axis nucleotide diversity. The red dotted line indicates the average nucleotide diversity.

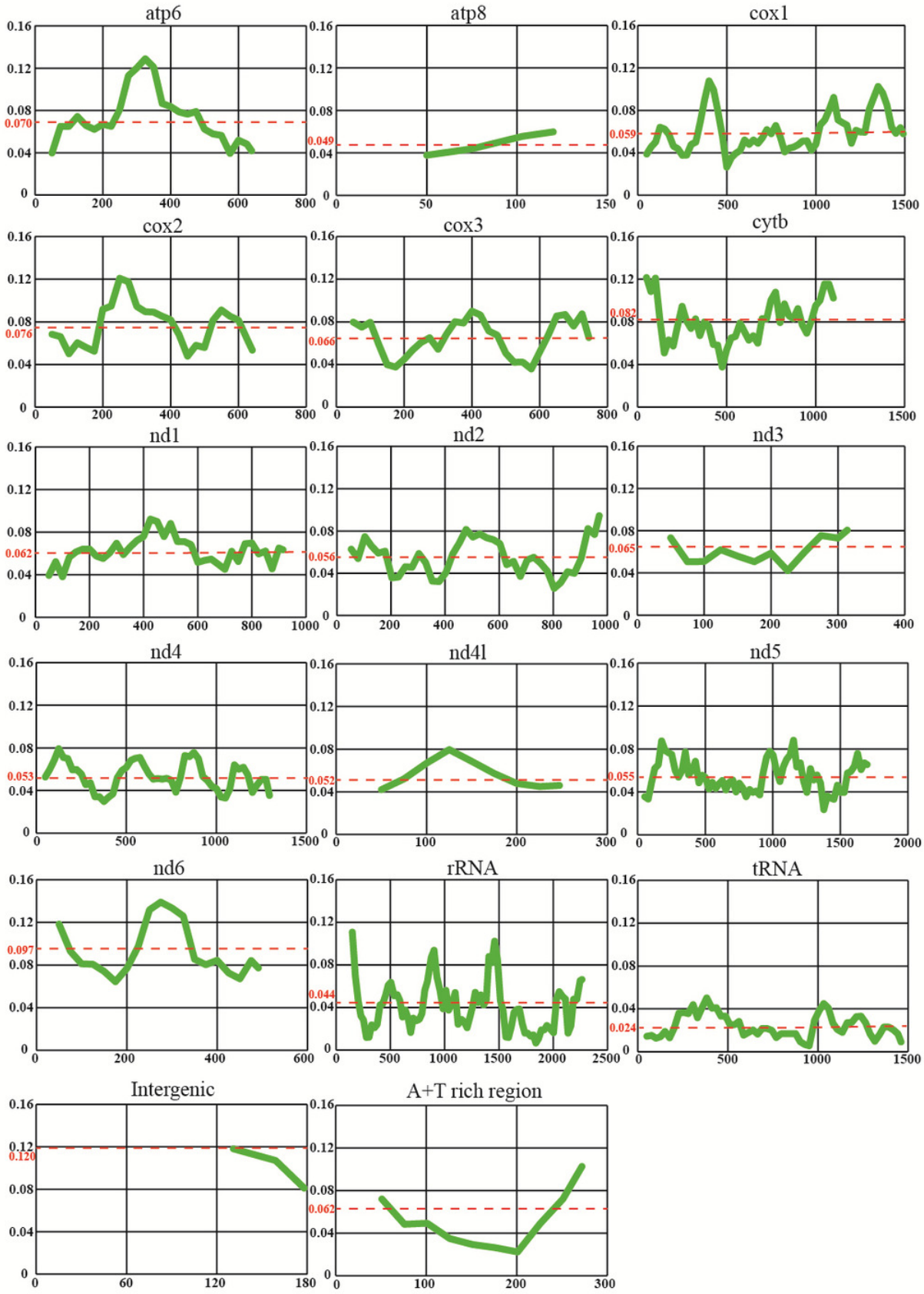

Figure 5. Sliding-window analyses of 13 protein coding genes, concatenated tRNA and rRNA genes, intergenic and A+T rich region among six Dendrolimus species. The X-axis represents sequence length, the Y-axis represents window size (bp), and the Y-axis indicates the average nucleotide diversity.

Table 1(on next page)

Genome organization of *Dendromlius kikuchii, D. houi* and *D. superans.*

Table 1. Genome organization of *Dendromlius kikuchii* (SM12 and SM22), *D. houi* (YN05 and YN11) and *D. superans* (LY04 and LY08).

Table 1. Genome organization of *Dendromlius kikuchii* (SM12 and SM22), *D. houi* (YN05 and YN11) and *D. superans* (LY04 and LY08).

Gene	Strand	Location					
		SM12	SM22	YN05	YN11	LY04	LY08
trnM	F	1-67	1-67	1-67	1-67	1-68	1-67
trnI	F	71-134	71-134	69-132	69-132	72-135	72-135
trnQ	R	132-200	132-200	130-198	130-198	133-201	133-201
nad2	F	257-1264	255-1262	254-1264	254-1264	256-1266	256-1266
trnW	F	1264-1334	1262-1332	1263-1332	1263-1332	1265-1336	1265-1335
trnC	R	1327-1392	1325-1390	1325-1391	1325-1391	1329-1394	1328-1393
trnY	R	1393-1460	1391-1458	1393-1463	1393-1463	1395-1460	1394-1459
cox1	F	1501-3031	1500-3030	1492-3022	1492-3022	1493-3023	1492-3022
trnL(UUR)	F	3032-3098	3031-3097	3023-3089	3023-3089	3024-3090	3023-3089
cox2	F	3099-3780	3098-3779	3090-3773	3090-3773	3091-3772	3090-3771
trnK	F	3781-3851	3780-3850	3775-3845	3775-3845	3773-3843	3772-3842
trnD	F	3852-3919	3851-3918	3847-3913	3847-3913	3847-3914	3846-3913
atp8	F	3920-4081	3919-4080	3914-4075	3914-4075	3915-4076	3914-4075
atp6	F	4075-4752	4074-4751	4069-4746	4069-4746	4070-4747	4069-4746
cox3	F	4759-5547	4760-5548	4762-5550	4762-5550	4760-5548	4759-5547
trnG	F	5550-5616	5551-5617	5553-5618	5553-5618	5551-5616	5550-5615
nad3	F	5617-5970	5618-5971	5619-5970	5619-5970	5617-5970	5616-5969
trnA	F	5973-6039	5974-6039	5971-6038	5971-6038	5975-6041	5974-6040
trnR	F	6058-6123	6058-6123	6048-6112	6048-6112	6055-6118	6054-6117
trnN	F	6145-6211	6145-6211	6118-6184	6118-6184	6120-6185	6119-6184
trnS(AGN)	F	6226-6293	6219-6286	6210-6277	6210-6277	6202-6269	6201-6268
trnE	F	6293-6357	6286-6350	6277-6345	6277-6345	6269-6333	6268-6333
trnF	R	6371-6437	6364-6430	6354-6420	6354-6420	6346-6412	6346-6412
nad5	R	6442-8181	6435-8174	6425-8167	6425-8167	6416-8158	6416-8158
trnH	R	8182-8246	8175-8239	8168-8231	8168-8231	8159-8226	8159-8226
nad4	R	8247-9585	8240-9578	8232-9570	8232-9570	8227-9565	8227-9565
nad4l	R	9620-9913	9611-9904	9602-9895	9602-9895	9604-9897	9604-9897
trnT	F	9918-9982	9909-9973	9900-9964	9900-9964	9905-9970	9905-9969
trnP	R	9983-10047	9974-10038	9965-10029	9965-10029	9971-10035	9970-10034
nad6	F	10056-10586	10047-10577	10038-10568	10038-10568	10044-10574	10043-10573
cob	F	10590-11738	10581-11729	10574-11722	10574-11722	10579-11727	10578-11726
trnS(UCN)	F	11737-11802	11728-11793	11726-11791	11726-11791	11731-11797	11730-11796
nad1	R	11802-12755	11793-12746	11791-12744	11791-12744	11797-12750	11796-12749
trnL(CUN)	R	12757-12827	12748-12818	12746-12813	12746-12813	12752-12820	12751-12819
rrnL	R	12828-14210	12819-14204	12814-14220	12814-14220	12821-14253	12820-14253
trnV	R	14211-14275	14205-14268	14221-14286	14221-14286	14254-14319	14254-14319
rrnS	R	14276-15058	14269-15051	14287-15062	14287-15063	14320-15101	14320-15100
AT region	F	15059-15377	15052-15370	15063-15381	15064-15382	15102-15417	15101-15417

Table 2(on next page)

Start codon and stop codon of 13 protein coding genes in six *Dendrolimus* species.

Table 2 Start codon and stop codon of 13 protein coding genes in six *Dendrolimus* species.

Table 2 Start codon and stop codon of 13 protein coding genes in six *Dendrolimus* species.

Samples	nad2	cox1	cox2	atp8	atp6	cox3	nad3	nad5	nad4	nad4l	nad6	cob	nad1
***D. spectabilis*02**	ATT/TAA	CGA/T	ATA/T	ATC/TAA	ATG/TAA	ATG/TAA	ATC/TA	ATT/TAA	ATG/T	ATG/TAA	ATA/TAA	ATG/TAA	ATG/TAA
***D. spectabilis*13**	ATT/TAA	CGA/T	ATA/T	ATC/TAA	ATG/TAA	ATG/TAA	ATC/TA	ATT/TAA	ATG/T	ATG/TAA	ATA/TAA	ATG/TAA	ATG/TAA
***D. tabulaeformis*06**	ATT/TAA	CGA/T	ATA/T	ATT/TAA	ATG/TAA	ATG/TAA	ATG/TA	ATT/TAA	ATG/T	ATG/TAA	ATA/TAA	ATG/TAA	ATG/TAA
***D. tabulaeformis*38**	ATT/TAA	CGA/T	ATA/T	ATC/TAA	ATG/TAA	ATG/TAA	ATT/TA	ATT/TAA	ATG/T	ATG/TAA	ATA/TAA	ATG/TAA	ATG/TAA
***D. punctatus*04**	ATT/TAA	CGA/T	ATA/T	ATT/TAA	ATG/TAA	ATG/TAA	ATG/TA	ATT/TAA	ATG/T	ATG/TAA	ATA/TAA	ATG/TAA	ATG/TAA
***D. punctatus*05**	ATT/TAA	CGA/T	ATA/T	ATT/TAA	ATG/TAA	ATG/TAA	ATA/TA	ATT/TAA	ATG/T	ATG/TAA	ATA/TAA	ATG/TAA	ATG/TAA
***D. punctatus_ws*03**	ATT/TAA	CGA/T	ATA/T	ATT/TAA	ATG/TAA	ATG/TAA	ATG/TA	ATT/TAA	ATG/T	ATG/TAA	ATA/TAA	ATG/TAA	ATG/TAA
***D. punctatus_ws*06**	ATT/TAA	CGA/T	ATA/T	ATT/TAA	ATG/TAA	ATG/TAA	ATG/TA	ATT/TAA	ATG/T	ATG/TAA	ATA/TAA	ATG/TAA	ATG/TAA
***D. kikuchii*12**	ATT/TAA	CGA/T	ATA/T	ATA/TAA	ATG/TAA	ATG/TAA	ATT/TAA	ATA/TAA	ATG/T	ATG/TAA	ATA/TAA	ATG/TAA	GTG/TAA
***D. kikuchii*22**	ATT/TAA	CGA/T	ATA/T	ATA/TAA	ATG/TAA	ATG/TAA	ATT/TAA	ATA/TAA	ATG/T	ATG/TAA	ATA/TAA	ATG/TAA	GTG/TAA
***D. houi*05**	ATT/TAA	CGA/T	ATA/TAG	ATT/TAA	ATG/TAA	ATG/TAA	ATT/T	ATT/TAA	ATG/T	ATG/TAA	ATA/TAA	ATG/TAA	ATG/TAA
***D. houi*11**	ATT/TAA	CGA/T	ATA/TAG	ATT/TAA	ATG/TAA	ATG/TAA	ATT/T	ATT/TAA	ATG/T	ATG/TAA	ATA/TAA	ATG/TAA	ATG/TAA
***D. superans*04**	ATC/TAA	CGA/T	ATA/T	ATT/TAA	ATG/TAA	ATG/TAA	ATC/TAA	ATT/TAA	ATG/T	ATG/TAA	ATA/TAA	ATG/TAA	ATG/TAA
***D. superans*08**	ATC/TAA	CGA/T	ATA/T	ATT/TAA	ATG/TAA	ATG/TAA	ATC/TAA	ATT/TAA	ATG/T	ATG/TAA	ATA/TAA	ATG/TAA	ATG/TAA

Table 3(on next page)

Sequence length of non-coding and overlapping regions between two genes among 14 individuals of *Dendrolimus* species.

Table 3 Sequence length of non-coding and overlapping regions between two genes among 14 individuals of *Dendrolimus* species.

Table 3 Sequence length of non-coding and overlapping regions between two genes among 14 individuals of *Dendrolimus* species.

Location*	*D. spectabilis*		*D. tabulaeformi*		*D. punctatus punctatus*		*D. punctatus wenshanensis*		*D. superans*		*D. kikuchii*		*D. houi*	
	CS02	CS13	YS06	YS38	MW04	MW05	WS03	WS06	LY04	LY08	SM12	SM22	YN05	YN11
trnM-trnI	3	3	3	3	3	3	3	3	3	4	3	3	1	1
trnI-trnQ	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
trnQ-nad2	58	58	58	58	58	58	58	58	54	54	56	54	55	55
nd2-trnW	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-1	-1	-2	-2
trnW-trnC	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
trnC-trnY	0	0	0	0	0	0	0	0	0	0	0	0	1	1
trnY-cox1	25	25	34	34	27	34	34	34	32	32	40	41	28	28
cox2-trnK	0	0	0	0	0	0	0	0	0	0	0	0	1	1
trnK-trnD	3	3	3	3	3	3	3	3	3	3	0	0	1	1
atp8-atp6	-7	-7	-7	-7	-7	-7	-7	-7	-7	-7	-7	-7	-7	-7
atp6-cox3	11	11	15	15	15	14	15	15	12	12	6	8	15	15
cox3-trnG	2	2	2	2	2	2	2	2	2	2	2	2	2	2
nad3-trnA	0	0	0	0	0	0	0	0	4	4	2	2	0	0
trnA-trnR	20	20	15	15	15	15	15	15	13	13	18	18	9	9
trnR-trnN	4	4	4	4	4	4	4	4	1	1	21	21	5	5
trnN-trnS(AGN)	18	18	11	11	11	13	11	11	16	16	14	7	25	25
trnS(AGN)-trnE	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
trnE-trnF	8	8	4	4	4	4	4	4	12	12	13	13	8	8
trnF-nad5	3	3	2	2	2	3	2	2	3	3	4	4	4	4
nad4-nd4l	23	23	24	24	24	19	24	24	38	38	34	32	31	31
nad4l-trnT	7	7	7	7	7	7	7	7	7	7	4	4	4	4
trnP-nad6	8	8	8	8	8	8	8	8	8	8	8	8	8	8
nad6-cytb	4	4	4	4	4	4	4	4	4	4	3	3	5	5
cytb-trnS(UCN)	3	3	3	3	3	3	3	3	3	3	-2	-2	3	3
trnS(UCN)-nd1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
nad1-trnL(CUN)	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Location*：Sequence length between two genes, positive value stands for non-coding regions, negative value stands for overlapping regions.