The complete mitochondrial genome of *Lerema accius* and its phylogenetic implications

Qian Cong[1] and Nick V. Grishin[1,2]

[1] Departments of Biophysics and Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX, United States

[2] Howard Hughes Medical Center, University of Texas Southwestern Medical Center, Dallas, TX, United States

ABSTRACT

Butterflies and moths (Lepidoptera) are becoming model organisms for genetics and evolutionary biology. Decoding the Lepidoptera genomes, both nuclear and mitochondrial, is an essential step in these studies. Here we describe a protocol to assemble mitogenomes from Next Generation Sequencing reads obtained through whole-genome sequencing and report the 15,338 bp mitogenome of *Lerema accius*. The mitogenome is AT-rich and encodes 13 proteins, 22 transfer-RNAs, and two ribosomal-RNAs, with a gene order typical for Lepidoptera mitogenomes. A phylogenetic study based on the protein sequences using both Bayesian Inference and Maximum Likelihood methods consistently place *Lerema accius* with other grass skippers (Hesperiinae).

INTRODUCTION

The order Lepidoptera contains approximately 160,000 described and half a million estimated species (*Kristensen, Scoble & Karsholt, 2007*). It represents one of the most diverse and fascinating groups of insects with many species emerging as model organisms for genetics and evolution (*Clarke & Sheppard, 1972*; *Nishikawa et al., 2013*; *Kunte et al., 2014*; *Zhan et al., 2011*; *Hines et al., 2012*; *Surridge et al., 2011*; *Engsontia et al., 2014*; *Zhang, Kunte & Kronforst, 2013*). Studies of these model organisms benefit significantly from decoding the genomes of select Lepidoptera species. Recently, we published the genome draft of Clouded Skipper *Lerema accius* using next generation sequencing techniques (*Cong et al., 2015*). Traditional genome assemblers failed to automatically assemble the Clouded Skipper mitogenome together with the nuclear genome. This failure probably resulted from a difficulty in distinguishing the mitogenome NGS reads from those of nuclear genome as well as a high erroneous k-mers frequency due to high mitochondrial DNA coverage. However, a dedicated effort should allow assembly of the mitogenome from whole-genome sequencing reads.

The insect mitogenome is circular, consisting of 14–19 kilobases (kb) that contain 13 protein-coding genes (PCGs), two ribosomal-RNA-coding genes (rRNAs), 22 transfer-RNA-coding genes (tRNAs), and an A + T rich displacement loop (D-loop) control region (*Cameron, 2014*). Because of their maternal inheritance, compact structure, lack of

genetic recombination, and relatively fast evolutionary rate, mitogenomes have been used widely in molecular phylogenetics and evolution studies (*Cameron, 2014*; *Moritz, Dowling & Brown, 1987*). Here, we assemble and annotate the complete mitogenome of *Lerema accius* from next generation sequencing reads. Phylogenetic analyses using published mitogenomes of skipper butterflies (Hesperiidae) place *Lerema accius* among other grass skippers (Hesperiinae).

METHODS

Library preparation and Illumina sequencing

We collected a male *Lerema accius* adult in the field (USA: Texas: Dallas County, Dallas, White Rock Lake, Olive Shapiro Park, 10-Nov-2013, GPS: 32.8621, −96.7305, elevation: 141 m) under permit #08-02Rev from Texas Parks and Wildlife Department (Natural Resources Program Director David H. Riskind). We removed the wings and abdomen of the deceased specimen (USA: Texas: Dallas County, Dallas, White Rock Lake, Olive Shapiro Park, 10-Nov-2013, GPS: 32.8621, −96.7305, elevation: 141 m), and used the remaining tissue to extract genomic DNA using the ChargeSwitch gDNA mini tissue kit (Life Technologies, Grand Island, NY, USA). About 500 ng of genomic DNA was used to prepare 250 bp and 500 bp paired-end libraries, respectively, following the Illumina TruSeq DNA sample preparation guide using enzymes from NEBNext Modules (New England Biolabs, Ipswich, MA, USA). These two libraries were pooled (and they occupied about 60% of one illumina lane) together with other libraries (not used for the mitogenome assembly) to sequence 150 bp from both ends with a rapid run on the Illumina HiSeq 2500 platform at the UT Southwestern Medical Center genomics core facility. The sequencing reads have been deposited in NCBI SRA database under accession numbers: SRR2089773– SRR2089775.

Mitogenome assembly

Sequencing reads were processed by MIRABAIT (*Chevreux, Wetter & Suhai, 1999*) to remove contamination from sequence adapters and trimmed low-quality regions (quality score <20) at both ends. Using the mitogenomes of four skippers (*Carterocephalus silvicola*, *Potanthus flavus*, *Polytremis nascens* and *Polytremis jigongi*) as references, we applied mitochondrial baiting and iterative mapping (MITObim) v1.6 (*Hahn, Bachmann & Chevreux, 2013*) to extract the sequencing reads of the mitogenome in the 250 bp and 500 bp libraries.

About 1,161,000 reads (1.04% of all reads) were extracted using MITObim. Because the average size of the reference mitogenomes is 15,400 bp, we expected an average coverage of about 22,600 fold (1,161,000×150×2/15,400). We used JELLYFISH software (*Marcais & Kingsford, 2011*) to obtain the frequencies of 15-mers in these reads. The frequencies of some 15-mers were much lower than expected. They might come from regions in the mitogenome that were poorly covered in the sequencing reads; alternately, they might arise from sequencing errors, heterogeneity in different copies of mitochondrial DNA and reads from the nuclear genome. The second scenario could cause problems in *de novo* assembly, and thus we applied QUAKE (*Kelley, Schatz & Salzberg, 2010*) to correct

errors in 15-mers with frequencies lower than 1,000 and excluded reads containing low-frequency 15-mers after error correction. We assembled the error-corrected reads into contigs *de novo* with Platanus (*Kajitani et al., 2014*). The contigs were further assembled into scaffolds using all the reads (including the ones containing 15-mers with frequencies lower than 1,000).

This automatic procedure assembled a draft mitogenome of 15,332 bp without any gaps. However, since the genome assembler Platanus is not deigned to assemble circular mitogenomes, the linear representation of the circular DNA may either (1) miss a fragment after its 3′-terminus and before its 5′-terminus or (2) have redundant fragments that appear both at the 3′-terminus and 5′-terminus. We manually inspected the sequences at the 5′- and 3′-termini and revealed that there was no redundant fragment but instead a fragment of six base pairs was missing. We determined the sequence of the missing fragment by searching for the two 32 bp fragments at the 5′- and 3′-termini of the draft mitogenome in the sequencing reads and selected the sequence between them. A majority (99.8%) of the reads revealed the same missing fragment (others likely contained sequencing errors) and we manually added it into the mitogenome. We also adjusted the linear representation of the circular DNA by circular permutation so that the sequence started with the trnM(cau) gene, which was the convention for most Lepidoptera sequences deposited in the database.

Annotation and analysis of the mitochondrial genome

The mitogenome sequence was annotated using the MITOS web server (*Bernt et al., 2013*). We translated the sequences of PCGs to protein sequences using the genetic code for invertebrate mitogenomes. The predictions from MITOS were manually curated using other published skipper mitogenomes as references, and the starts and ends of genes were modified, if necessary, to be consistent with other species. The open reading frames (after modification) of the protein coding genes were validated. Secondary structures of tRNA genes were predicted using the same server.

Assembly quality assessment

We mapped the 250 bp and 500 bp paired-end reads to the mitogenome using bowtie2 v2.2.3 (*Langmead & Salzberg, 2012*) and processed the results with SAMtools (*Li et al., 2009*). Coverage depth at each position was calculated based on this mapping result. As the sequencing reads that could map partly to the 5′-terminus and partly to the 3′-terminus would map only to one terminus or fail to map, the coverage at the termini could be under-estimated. Therefore, we recalculated the coverage for the 1,000 bp segments in the 5′- and 3′-termini based on the mapping result to another linear representation of the circular mitogenome that was obtained by connecting the 5′-terminal half to the end of 3′-terminal half.

Only two regions of the mitogenome showed coverage below 1,000 fold. One of them was a low complexity region that contained mostly (85%) T and another was a 46 bp fragment of AT repeats. Such AT-rich regions tend to be underrepresented in the sequencing libraries as they break easier during the library preparation (*Benjamini & Speed, 2012*).

Table 1 List of taxa analyzed in present paper.

Species	Length	Identity[a]	Accession	References
Ampittia dioscorides	15,313	91.2%	KM102732.1	XW Yang et al., 2014, unpublished data
Apocheima cinerarium	15,722	n.a.	NC_024824.1	*Liu et al. (2014)*
Biston suppressaria	15,628	99.7%	NC_027111.1	*Chen et al. (2015)*
Carterocephalus silvicola	15,765	99.0%	NC_024646.1	*Kim et al. (2014)*
Celaenorrhinus maculosa	15,282	n.a.	NC_022853.1	*Wang, Hao & Zhao (2013)*
Choaspes benjaminii	15,300	85.9%	NC_024647.1	*Kim et al. (2014)*
Ctenoptilum vasava	15,468	n.a.	NC_016704.1	*Hao et al. (2012)*
Daimio tethys	15,350	96.8%	NC_024648.1	*Kim et al. (2014)*
Erynnis montanus	15,530	99.8%	NC_021427.1	*Wang et al. (2014)*
Graphium timur	15,226	97.9%	NC_024098.1	*Chen et al., (2014)*
Hasora anura	15,290	n.a.	NC_027263.1	*Wang et al. (2015)*
Lobocla bifasciata	15,366	95.9%	NC_024649.1	*Kim et al. (2014)*
Ochlodes venata	15,622	78.5%	NC_018048.1	C Xu et al., 2012, unpublished data
Papilio glaucus	15,306	99.5%	NC_027252.1	*Shen, Cong & Grishin (2015)* and *Cong et al., (2015)*
Parnassius apollo	15,404	98.6%	NC_024727.1	*Kim et al. (2009)*
Phthonandria atrilineata	15,499	99.9%	NC_010522.1	*Yang et al. (2009)*
Polytremis jigongi	15,353	99.6%	NC_026990.1	*Jiang et al. (2015)*
Polytremis nascens	15,392	83.6%	NC_026228.1	*Jiang et al. (2015)*
Potanthus flavus	15,267	99.3%	NC_024650.1	*Kim et al., (2014)*

Notes.

[a]Identity: the lowest sequence identity to independently sequenced mitochondrial DNA of the same species in the Non-redundant database identified by BLAST.
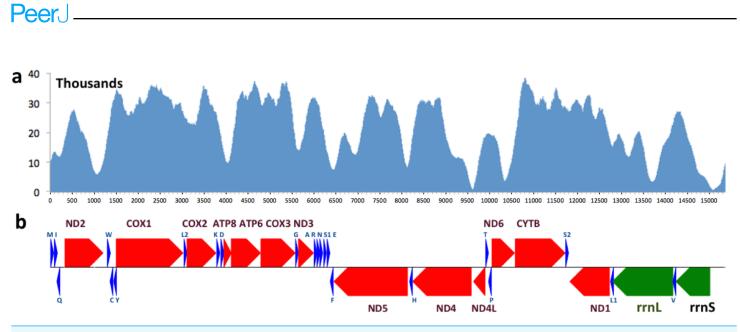
n.a.: there is no other mitochondrial sequences of the same species in the Genbank for cross-validation.
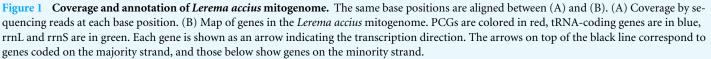
Manual inspection by searching the flanking regions of these poorly covered fragments in the sequencing reads revealed variation in the length of the poly-T sequence in the first fragment and the number AT-repeats in the second fragment, respectively. The variation might correspond to the heterogeneity in different copies of mitochondrial DNA in the specimen. We confirmed that the mitogenome produced by the de novo assembler did represent the dominant form of the possible variations.

We further assessed the quality of our assembly by its consistency with other published skipper mitogenomes in the protein-, rRNA- and tRNA-coding regions. We aligned the rRNA- and tRNA-coding sequences directly and aligned translated sequences for PCGs. Alignments confirmed that our sequences were consistent with the majority of available mitogenomes, and gaps were only in regions that are poorly conserved among other skipper species. In addition, the COI barcode (5′-terminal region of cytochrome oxidase subunit 1 coding gene) of *Lerema accius* was reported previously (Genbank accession: GU088418.1) and this sequence agreed 100% with the corresponding region in our mitogenome.

Phylogenetic analysis

The mitogenomes of 13 other skipper species that were available (up to June, 2015) were downloaded from NCBI (Table 1). Three moths from the Geometridae family and three species of the Papilionidae family were used as outgroups. A blast search against all the available sequences of the same species in the non-redundant database was used to validate

Figure 1 **Coverage and annotation of *Lerema accius* mitogenome.** The same base positions are aligned between (A) and (B). (A) Coverage by sequencing reads at each base position. (B) Map of genes in the *Lerema accius* mitogenome. PCGs are colored in red, tRNA-coding genes are in blue, rrnL and rrnS are in green. Each gene is shown as an arrow indicating the transcription direction. The arrows on top of the black line correspond to genes coded on the majority strand, and those below show genes on the minority strand.
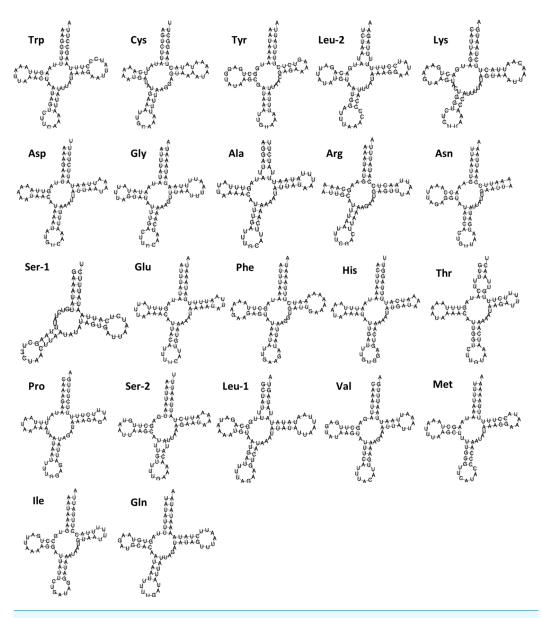
each mitogenome sequence. For most species, some individual genes in the mitogenome were sequenced independently and the mitogenome sequence was consistent with these gene sequences (sequence identity >95%). However, four skippers were excluded from downstream analyses (*Ampittia dioscorides, Choaspes benjaminii, Ochlodes venata* and *Polytremis nascens*, three of which are unpublished but available from GenBank) due to poor agreement for at least one gene sequence found in GenBank.
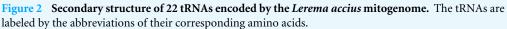
Protein sequences of the 13 protein-coding genes were aligned by MAFFT. We manually checked the alignments, corrected annotation errors based on consensus and removed positions with long gaps and their surrounding regions with uncertain alignment. The processed alignments were concatenated and analyzed with Bayesian Inference and Maximum likelihood methods using Phylobayes-MPI v1.5a (*Lartillot, Lepage & Blanquart, 2009*) (model: CATGTR (*Lartillot & Philippe, 2004*)) and RaxML v8.1.17 (*Stamatakis, 2014*) (model: PROTGAMMAAUTO), respectively. The resulting phylogenetic trees were visualized in FigTree v1.4.2.

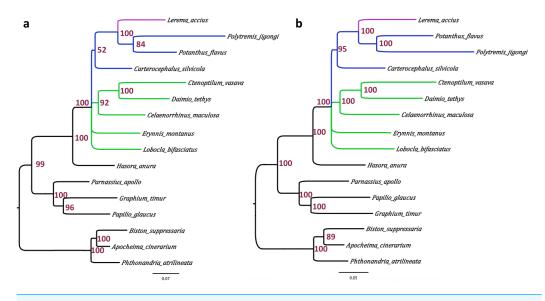## RESULTS AND DISCUSSION

Annotation of the mitogenome

The complete mitogenome of *Lerema accius* is deposited in GenBank of NCBI under accession number KT598278. The length of this mitogenome is 15,338 bp and it retains the typical insect mitogenome gene set and gene order, including 13 PCGs (nd1-6, nd4l, cox1-3, atp8, atp6, and cytb), 22 tRNA genes (two for Serine and Leucine and one for each of the rest of the amino acids), 2 ribosomal RNAs (rrnL and rrnS), and an A + T rich D-loop control region. The annotation of the mitogenome is illustrated in Fig. 1. The cox1 gene uses start codon CGA, which is consistent with many other insect mitogenomes (*Kim et al., 2009*). All the rest of the genes start with the typical ATN. cox1, cox2 and nd4

Figure 2 Secondary structure of 22 tRNAs encoded by the *Lerema accius* mitogenome. The tRNAs are labeled by the abbreviations of their corresponding amino acids.
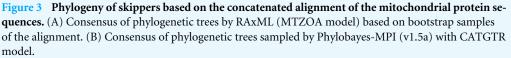
use an incomplete stop codon T (*Ojala, Montoya and Attardi, 1981*), and a complete TAA codon will likely be formed during mRNA maturation (*Ojala, Montoya and Attardi, 1981*; *Boore, 1999*).

The lengths of tRNA-coding genes range from 60 bp to 70 bp. Secondary structures predicted by MITOS suggest that all tRNAs adopt a typical cloverleaf structure except for trnS1(gcu) (Fig. 2). The dihydrouridine (DHU) arm of trnS1(gcu) does not form a stable stem-loop structure, which is very common in butterfly mitogenomes (*Lu et al., 2013*; *Kim et al., 2014*). A 488 bp A + T rich region (A + T content: 94.7%) connects rrnS and trnM(cau). This region contains an "ATAGA" motif located 22 bp downstream from rrnS

Figure 3 **Phylogeny of skippers based on the concatenated alignment of the mitochondrial protein sequences.** (A) Consensus of phylogenetic trees by RAxML (MTZOA model) based on bootstrap samples of the alignment. (B) Consensus of phylogenetic trees sampled by Phylobayes-MPI (v1.5a) with CATGTR model.

and is followed by 15 bp of poly-T stretch that is a gene regulation element commonly found in Lepidoptera (*Lu et al., 2013*; *Salvato et al., 2008*).

Phylogenetic analysis

We built a phylogenetic tree of the 10 skipper species with published mitogenomes, based on the concatenated alignment of the mitochondrial protein sequences. Three Papilionidae and three Geometridae mitogenomes were used as outgroups. Maximum likelihood method RAxML automatically selected MTZOA, a general mitochondrial amino acid substitution model, as the most appropriate, and placed *Lerema accius* among other grass-skippers (Subfamily Hesperiinae). A Bayesian analysis with the CATGTR model supported a tree with exactly the same topology. This topology is largely consistent with previously reported phylogenetic studies on the basis of standard gene markers and morphology (*Warren, Ogawa and Brower, 2008*; *Warren, Ogawa and Brower, 2009*; *Yuan et al., 2015*). Notably, the subfamily Coeliadinae (represented by *Hasora anura*) is a sister to all other Hesperiidae. Topology between the subfamilies Eudaminae (*Lobocla bifasiatus*), Pyrginae (other branches shown in green in Fig. 3) and remaining Skippers is unresolved. The tribes Celaenorrhini (*Celaenorrhinus maculosa*) and Tagiadini (*Daimio* and *Ctenoptilum*) group together (in the absence of Pyrrhopygini). The subfamily Heteropterinae (represented by *Carterocephalus silvicola*) is a sister to grass skippers (Hesperiinae).

Interestingly, based on the mitochondrial genome, the two Asian grass skippers (*Potanthus* from the tribe Taractrocerini and *Polytremis* from the tribe Baorini) are grouped together, and *Lerema* (from the tribe Moncini) is their sister. The sequences of two nuclear markers, EF1a and wingless, are available from these species in the database; however, they support different topologies at low confidence. While the maximal likelihood

tree based on EF1a favors (bootstrap: 52%) the same topology as the mitogenome, the tree based on wingless groups *Potanthus* with Lerema with a 63% bootstrap support and places *Polytremis* as their sister. The phylogeny between these tribes could become clear when more sequences from more taxa become available.

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Qian Cong conceived and designed the experiments, performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Nick V. Grishin conceived and designed the experiments, contributed reagents/materials/analysis tools, wrote the paper, reviewed drafts of the paper.

Field Study Permissions

The following information was supplied relating to field study approvals (i.e., approving body and any reference numbers):

Texas Parks and Wildlife Department (Natural Resources Program Director David H. Riskind) issued permit #08–02Rev for Texas State Parks.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The complete mitogenome of Lerema accius is deposited in GenBank of NCBI under accession number KT598278.

Data Availability

The following information was supplied regarding data availability:

The sequencing reads have been deposited in NCBI SRA database under accession numbers: SRR2089773– SRR2089775.

REFERENCES

Benjamini Y, Speed TP. 2012. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Research* **40**:e72 DOI 10.1093/nar/gks001.

Bernt M, Donath A, Juhling F, Externbrink F, Florentz C, Fritzsch G, Putz J, Middendorf M, Stadler PF. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* **69**:313–319 DOI 10.1016/j.ympev.2012.08.023.

Boore JL. 1999. Animal mitochondrial genomes. *Nucleic Acids Research* **27**:1767–1780 DOI 10.1093/nar/27.8.1767.

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.

Chen Y, Gan S, Shao L, Cheng C, Hao J. 2014. The complete mitochondrial genome of the Pazala timur (Lepidoptera: Papilionidae: Papilioninae). *Mitochondrial DNA* **27(1)**:533–534 DOI 10.3109/19401736.2014.905843.

Chen SC, Wang XQ, Wang JJ, Hu X, Peng P. 2015. The complete mitochondrial genome of a tea pest looper, Buzura suppressaria (Lepidoptera: Geometridae). *Mitochondrial DNA* Epub ahead of print Feb 11 2015.

Chevreux B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information. In: *Computer science and biology: proceedings of the German conference on bioinformatics (GCB)*, 45–56. *Available at* http://www.bioinfo.de/isb/gcb99/talks/chevreux/.

Clarke CA, Sheppard PM. 1972. The genetics of the mimetic butterfly Papilio polytes L. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **263**:431–458 DOI 10.1098/rstb.1972.0006.

Cong Q, Borek D, Otwinowski Z, Grishin Nick V. 2015. Tiger swallowtail genome reveals mechanisms for speciation and caterpillar chemical defense. *Cell Reports* **10**:910–919 DOI 10.1016/j.celrep.2015.01.026.

Engsontia P, Sangket U, Chotigeat W, Satasook C. 2014. Molecular evolution of the odorant and gustatory receptor genes in lepidopteran insects: implications for their adaptation and speciation. *Journal of Molecular Evolution* **79**:21–39 DOI 10.1007/s00239-014-9633-0.

Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research* **41**:e129 DOI 10.1093/nar/gkt371.

Hao J, Sun Q, Zhao H, Sun X, Gai Y, Yang Q. 2012. The complete mitochondrial genome of *Ctenoptilum vasava* (Lepidoptera: Hesperiidae: Pyrginae) and its phylogenetic implication. *Comparative and Functional Genomics* **2012**:328049 DOI 10.1155/2012/328049.

Hines HM, Papa R, Ruiz M, Papanicolaou A, Wang C, Nijhout HF, McMillan WO, Reed RD. 2012. Transcriptome analysis reveals novel patterning and pigmentation

genes underlying Heliconius butterfly wing pattern variation. *BMC Genomics* **13**:288 DOI 10.1186/1471-2164-13-288.

Jiang W, Zhu J, Yang Q, Zhao H, Chen M, He H, Yu W. 2015. Complete mitochondrial DNA genome of Polytremis nascens (Lepidoptera: Hesperiidae). *Mitochondrial DNA* Epub ahead of print Feb 18 2015.

Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M, Nagayasu E, Maruyama H, Kohara Y, Fujiyama A, Hayashi T, Itoh T. 2014. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Research* **24**:1384–1395 DOI 10.1101/gr.170720.113.

Kelley DR, Schatz MC, Salzberg SL. 2010. Quake: quality-aware detection and correction of sequencing errors. *Genome Biology* **11**:R116 DOI 10.1186/gb-2010-11-11-r116.

Kim MI, Baek JY, Kim MJ, Jeong HC, Kim KG, Bae CH, Han YS, Jin BR, Kim I. 2009. Complete nucleotide sequence and organization of the mitogenome of the red-spotted apollo butterfly, Parnassius bremeri (Lepidoptera: Papilionidae) and comparison with other lepidopteran insects. *Molecular Cell* **28**:347–363 DOI 10.1007/s10059-009-0129-5.

Kim MJ, Wang AR, Park JS, Kim I. 2014. Complete mitochondrial genomes of five skippers (Lepidoptera: Hesperiidae) and phylogenetic reconstruction of Lepidoptera. *Gene* **549**:97–112 DOI 10.1016/j.gene.2014.07.052.

Kristensen NP, Scoble MJ, Karsholt O. 2007. Lepidoptera phylogeny and systematics: the state of inventorying moth and butterfly diversity. *Zootaxa* **1668**:699–747.

Kunte K, Zhang W, Tenger-Trolander A, Palmer DH, Martin A, Reed RD, Mullen SP, Kronforst MR. 2014. doublesex is a mimicry supergene. *Nature* **507**:229–232 DOI 10.1038/nature13112.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**:357–359 DOI 10.1038/nmeth.1923.

Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**:2286–2288 DOI 10.1093/bioinformatics/btp368.

Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution* **21**:1095–1109 DOI 10.1093/molbev/msh112.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R Genome project data processings. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**:2078–2079 DOI 10.1093/bioinformatics/btp352.

Liu S, Xue D, Cheng R, Han H. 2014. The complete mitogenome of Apocheima cinerarius (Lepidoptera: Geometridae: Ennominae) and comparison with that of other lepidopteran insects. *Gene* **547**:136–144 DOI 10.1016/j.gene.2014.06.044.

Lu HF, Su TJ, Luo AR, Zhu CD, Wu CS. 2013. Characterization of the complete mitochondrion genome of diurnal Moth Amata emma (Butler) (Lepidoptera: Erebidae)

and its phylogenetic implications. *PLoS ONE* **8**:e72410
DOI 10.1371/journal.pone.0072410.

Marcais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**:764–770
DOI 10.1093/bioinformatics/btr011.

Moritz C, Dowling TE, Brown WM. 1987. Evolution of animal mitochondrial-DNA—relevance for population biology and systematics. *Annual Review of Ecology and Systematics* **18**:269–292 DOI 10.1146/annurev.es.18.110187.001413.

Nishikawa H, Iga M, Yamaguchi J, Saito K, Kataoka H, Suzuki Y, Sugano S, Fujiwara H. 2013. Molecular basis of wing coloration in a Batesian mimic butterfly, Papilio polytes. *Scientific Reports* **3**:3184 DOI 10.1038/srep03184.

Ojala D, Montoya J, Attardi G. 1981. tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290**:470–474 DOI 10.1038/290470a0.

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.

Shen J, Cong Q, Grishin NV. 2015. The complete mitochondrial genome of *Papilio glaucus* and its phylogenetic implications. *Meta Gene* **5**:68–83
DOI 10.1016/j.mgene.2015.05.002.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312–1313
DOI 10.1093/bioinformatics/btu033.

Surridge AK, Lopez-Gomollon S, Moxon S, Maroja LS, Rathjen T, Nadeau NJ, Dalmay T, Jiggins CD. 2011. Characterisation and expression of microRNAs in developing wings of the neotropical butterfly Heliconius melpomene. *BMC Genomics* **12**:62
DOI 10.1186/1471-2164-12-62.

Wang K, Hao J, Zhao H. 2013. Characterization of complete mitochondrial genome of the skipper butterfly, Celaenorrhinus maculosus (Lepidoptera: Hesperiidae). *Mitochondrial DNA* **26(5)**:690–691 DOI 10.3109/19401736.2013.840610.

Wang J, James JY, Xuan S, Cao T, Yuan X. 2015. The complete mitochondrial genome of the butterfly Hasora anura (Lepidoptera: Hesperiidae). *Mitochondrial DNA* Epub ahead of print Oct 14 2015.

Wang AR, Jeong HC, Han YS, Kim I. 2014. The complete mitochondrial genome of the mountainous duskywing, *Erynnis montanus* (Lepidoptera: Hesperiidae): a new gene arrangement in Lepidoptera. *Mitochondrial DNA* **25**:93–94
DOI 10.3109/19401736.2013.784752.

Warren AD, Ogawa JR, Brower AVZ. 2008. Phylogenetic relationships of subfamilies and circumscription of tribes in the family Hesperiidae (Lepidoptera: Hesperioidea). *Cladistics* **24**:642–676 DOI 10.1111/j.1096-0031.2008.00218.x.

Warren AD, Ogawa JR, Brower AVZ. 2009. Revised classification of the family Hesperiidae (Lepidoptera: Hesperioidea) based on combined molecular and morphological data. *Systematic Entomology* **34**:476–523.

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**:1441–1449 DOI 10.1007/s11033-008-9334-0.

Yuan X, Gao K, Yuan F, Wang P, Zhang Y. 2015. Phylogenetic relationships of subfamilies in the family Hesperiidae (Lepidoptera: Hesperioidea) from China. *Scientific Reports* **5**:11140 DOI 10.1038/srep11140.

Zhan S, Merlin C, Boore JL, Reppert SM. 2011. The monarch butterfly genome yields insights into long-distance migration. *Cell* **147**:1171–1185 DOI 10.1016/j.cell.2011.09.052.

Zhang W, Kunte K, Kronforst MR. 2013. Genome-wide characterization of adaptation and speciation in tiger swallowtail butterflies using de novo transcriptome assemblies. *Genome Biology and Evolution* **5**:1233–1245 DOI 10.1093/gbe/evt090.