Journal of Medical Entomology, 54(5), 2017, 1107–1108
doi: 10.1093/jme/tjx136
Advance Access Publication Date: 14 July 2017
Letter to the Editor

Comment on Seri Masran and Ab Majid 2017

Shu-Ping Tseng and Chin-Cheng (Scotty) Yang[1]

Research Institute for Sustainable Humanosphere, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan (magic760812@gmail.com; ccyang@rish.kyoto-u.ac.jp), and [1]Corresponding author, e-mail: ccyang@rish.kyoto-u.ac.jp

Subject Editor: William Reisen

Received 22 March 2017; Editorial decision 22 March 2017

TO THE EDITOR:

We refer to a paper recently published in the *Journal of Medical Entomology* on the phylogenetic relationships and mtDNA diversity of tropical bed bug, *Cimex hemipterus* (Hemiptera: Cimicidae), populations in Malaysia (Seri Masran and Ab Majid 2017). We feel that the methodology, analyses, discussion, and conclusion are seriously flawed based on the following reasons:

First, the sequence data generated from the study were unreliable. The authors pooled three individuals of the tropical bed bug from the same site into a single DNA mixture and used this for mitochondrial DNA (mtDNA) PCR and sequencing. In addition, we were unable to understand the rationale behind sequencing pools. By doing so, the determination of the mtDNA haplotype will either be unlikely or be inaccurate, as multiple PCR products in a single sequencing reaction will generate overlapping peaks in all or part of the electropherogram. This will make base calling ambiguous or impossible. Even if the authors managed to obtain a single "clean" sequence, we questioned how the authors excluded the possibility of preferential PCR amplification of a certain haplotype over another, if several happened to exist. Moreover, it was extremely inappropriate to consider the pooled-DNA-based sequencing data as a "population", a term that was frequently misused in the discussion section because the authors failed to recover "real" sequence variations from the three individuals from which the DNAs were extracted. A dataset composed of individual sequences generated from multiple bed bugs (a sample size ranging 5 to 25 individuals is often considered sufficient to guarantee accurate estimates of population-level genetic variation, Goodall-Copestake et al. 2012) collected from a site would be appropriate for a population-level study. Pooled DNA was not necessary to ensure greater DNA quality and better sequencing output as claimed by the authors, because many studies have demonstrated that sufficient DNA may be obtained from a single individual or tissue from a single specimen to perform PCR, followed by direct sequencing of the PCR products. For instance, genomic DNA extracted from half of the thorax and legs of a single bed bug (*Cimex* spp.) was sufficient for mtDNA sequencing or nuclear DNA genotyping (Balvín et al. 2015, Booth et al. 2015).

Second, the amplified sequences differed in length by as many as 120 base pairs (see Fig. 1 and the results section: "The average size fragment of the *COI* gene from the amplified *C. hemipterus* was 473 bp. The amplified *COI* gene sequences obtained varied from 411 to 531 bp"). These samples purported were from a single species, and therefore the length of all amplified sequences should be similar. It is possible that variation in amplified sequences may exist due to noisy trace peaks or sequencing hard stops. In addition, variation could arise if either the authors failed to trim all the generated sequences in a proper manner, or the template DNAs were from different species of bed bug. If the latter, then possibly the bugs were not identified correctly.

Third, the quality of sequences submitted to the Genbank database was poor and contained a number of flaws that were most likely associated with improper handling of the sequences. The presence of primer sequences and numerous "N"s at the two ends of submitted sequence contigs suggested that sequences with background noise were not "trimmed out" (Fig. 1). It is the contributor's responsibility to ensure that the submitted sequence data contain correct information before the submission to Genbank. The authors clearly failed to do so here, and this may lead to serious confusion, such as misinterpretation of haplotype diversity. For instance, many of sequences differed in only a single nucleotide (e.g., KT851503 vs. KT851521), and the respective nucleotide that separated the two sequences was labeled with "N", which denotes that the nucleotide could be any base. Most of the "N"s resided on the primer sequence region that was supposed to be trimmed off when the sequences were edited. If the authors had performed a proper sequence editing and cleanup, they would realize that the sequences they generated resembled only two mtDNA haplotypes, and not as many as stated by the authors. In addition, the sequence alignment used in this study should be deposited into a public database such as DRYAD. Otherwise, the analyses in this study are not reproducible by others.

Correcting for the abovementioned issues, the *C. hemipterus* specimens were represented by only two mtDNA haplotypes with a single nucleotide difference. The authors, however, treated the two haplotypes as "two different groups (Ch01 and Ch02)," which is a serious flaw. The phylogenetic tree is therefore redundant and not required. Aside from the necessity of a phylogenetic tree, it was also inappropriate to use sequences from the body louse as the outgroup species, especially as there are so many mtDNA sequences from *Cimex* species in the Genbank database. Using a distantly related species as an outgroup may lead to incorrect evolutionary and phylogenetic inferences (Graham et al. 2002). Also, the primer information given was unclear, making it impossible for readers to interpret whether the sequences of the body louse (AY239287) and *C. hemipterus* were from the same COI regions. Moreover, the alignment of all sequences was ambiguous (Fig. 1), especially for the outgroup sequence where the GUIDANCE sequence score was as low as 0.739 (cut-off for GUIDANCE alignment confidence score is usually set at 0.9, Penn et al. 2010), suggesting that inclusion of body louse as the outgroup was inappropriate.

Finally, the existence of two mtDNA haplotypes in the sampled populations renders any further interpretations on population demography

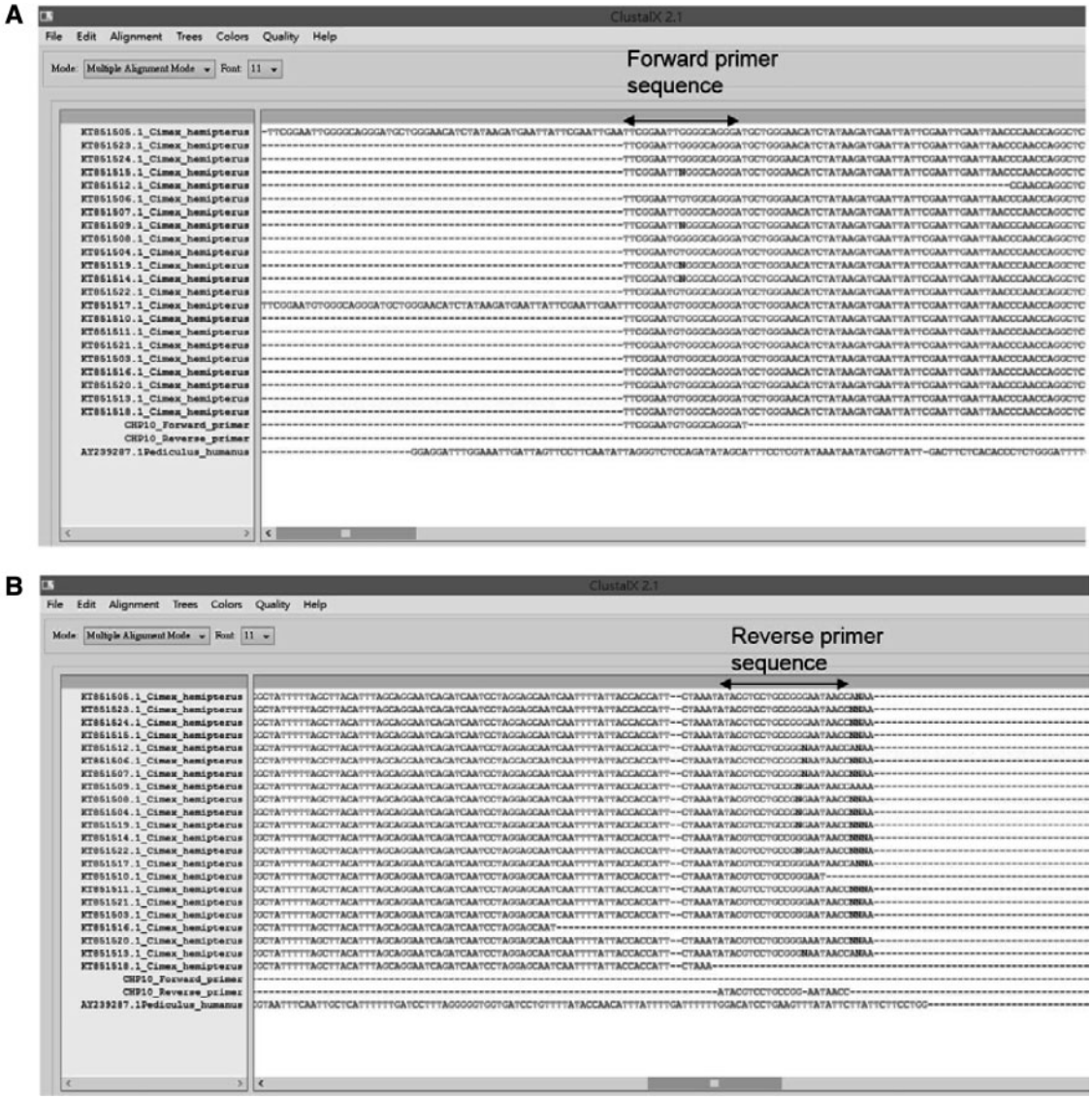

Fig. 1. The Clustal X alignment of the sequences used in Seri Masran and Ab Majid (2017). Only the left end (**a**) and the right end (**b**) of the alignment are shown. Note the presence of multiple "N"s and untrimmed primer sequences on both ends of the alignment.

of *C. hemipterus* speculative and unfounded. For example, the authors stated that "The monophyletic clade suggests that *C. hemipterus* in Malaysia might have undergone repeated cycles of population fluctuation, thus they appear as the populations with a low genetic diversity." "Monophyletic clade" is not a proper phylogenetic term for describing the genetic grouping of the tropical bed bug in Malaysia, as all sequences generated in this study were simply classified as two haplotypes based on our analyses. In addition, the authors did not perform any demographic inference (e.g., Skyline-plot methods, reviewed in Ho and Shapiro 2011) before drawing such a conclusion. We felt that low genetic diversity of *C. hemipterus* found in Malaysia could be best explained by other factors such as founder effect or genetic drift, as the authors have neither direct evidence to show any genetic signature of repeated cycles of population fluctuation nor history of bed bug control in every sampled site.

References Cited

Balvín, O., S. Roth, and J. Vilímová. 2015. Molecular evidence places the swallow bug genus *Oeciacus* Stål within the bat and bed bug genus *Cimex* Linnaeus (Heteroptera: Cimicidae). Syst. Entomol. 40: 652–665.

Booth, W., O. Balvín, E. L. Vargo, J. Vilímová, and C. Schal. 2015. Host association drives genetic divergence in the bed bug, *Cimex lectularius*. Mol. Ecol. 24: 980–992.

Goodall-Copestake, W. P., G. A. Tarling, and E. J. Murphy. 2012. On the comparison of population-level estimates of haplotype and nucleotide diversity: a case study using the gene cox1 in animals. Heredity 109: 50–56.

Graham, S. W., R. G. Olmstead, and S.C.H. Barrett. 2002. Rooting phylogenetic trees with distant outgroups: A case study from the commelinoid monocots. Mol. Biol. Evol. 19: 1769–1781.

Ho, S.Y.W., and B. Shapiro. 2011. Skyline-plot methods for estimating demographic history from nucleotide sequences. Mol. Ecol. Resour. 11: 423–434.

Penn, O., E. Privman, H. Ashkenazy, G. Landan, D. Graur, and T. Pupko. 2010. GUIDANCE: A web server for assessing alignment confidence scores. Nucleic Acids Res. 38: W23–W28.

Seri Masran, S.N.A., and A. H. Ab Majid. 2017. Genetic diversity and phylogenetic relationships of *cytochrome c oxidase* subunit I in *Cimex hemipterus* (Hemiptera: Cimicidae populations in Malaysia. J. Med. Entomol. 54: 974–979.