Common but not connected: high genetic structure and cryptic genetic diversity in the ubiquitous biting midge *Culicoides peregrinus* Kieffer

Mintara, R.¹, Pramual, P.¹*

¹Department of Biology, Faculty of Science, Mahasarakham University, Maha Sarakham 44150 Thailand

*Corresponding author: pairot.p@msu.ac.th

**ARTICLE HISTORY**
Received: 14 June 2023  
Revised: 30 June 2023  
Accepted: 3 July 2023  
Published: 30 September 2023

**ABSTRACT**

The biting midge *Culicoides peregrinus* Kieffer is a significant pest and vector species, and knowledge of its genetic diversity and genetic structure is critically important for designing an effective control program. However, such information is limited to only small sample-size DNA barcoding studies. Therefore, in this study, we used mitochondrial cytochrome c oxidase I (COI) to examine genetic structure and diversity of *C. peregrinus* from northeastern Thailand. In addition, we also inferred genetic relationships between *C. peregrinus* from Thailand and those reported from other countries across the geographic range of the species. Maximum intraspecific genetic divergence (3.83%) within Thai specimens was relatively high compared to other *Culicoides* species. Genetic structure analysis revealed that 71% (32 from 45) of population comparisons were highly significantly different. A high level of genetic structure among populations, even between those in close geographic proximity (22 km geographic distance) suggested that there has been little or no movement between local populations. This is possibly due to the ability to exploit diverse types of breeding site and a generalist feeding habit which enables *C. peregrinus* to complete its life cycle within cattle pens. Genetic relationships between Thai *C. peregrinus* and those reported from other countries revealed three genetically divergent lineages (A, B and C) associated with geographic origins. Specimens from Thailand + China formed lineage A, those from Australia formed lineage B and India + Bangladesh belonged to lineage C. These genetically divergent lineages also agree with morphological variation of the wing pale marking spots. Further investigation using independent genetic loci from nuclear genes will be very useful to resolve taxonomic status of these divergent lineages.

**Keywords:** Biting midge; DNA barcode; genetic diversity; genetic structure; population genetics.

**INTRODUCTION**

Biting midges of the genus *Culicoides* Latreille (Diptera: Ceratopogonidae) are small (<2.5 mm in body length) blood-sucking insects. There are >1,300 extant species recorded globally (Borkent & Dominik, 2020). Many of these are vectors of pathogens that can be transmitted to humans and other animals, causing diseases such as Oropouche fever in humans, bluetongue disease in ruminants, African horse sickness in equines, and leucocytozoonosis in chickens (Wirth & Hubert, 1989; Yanase et al., 2013; Harsha & Mazuda, 2015). Based on the biting habit, *C. peregrinus* was considered as a general feeder because it feeds on diverse hosts including humans, horse, goat, sheep, cow, buffalo, pig, dog and birds (Wirth & Hubert, 1989; Jomkumsing et al., 2021; Kar et al., 2022b; Sunantaraporn et al., 2022; Kamyngkird et al., 2023).

Despite its significance as a pest and potent vector, little is known about the genetic diversity and genetic structure of *C. peregrinus* except for DNA barcoding studies based on small sample sizes (Harrup et al., 2016; Liu et al., 2018; Jomkumsing et al., 2021; Gopurenko et al., 2022). Understanding the genetic diversity and genetic structure of the vector species is important for disease epidemiology and might be used to design an effective control program (Tabachnick & Black, 1995; McCoy, 2008). For example, the level of genetic differentiation could indicate rate of individual movement between local populations. This information is critically important for vector control strategy (Tabachnick & Black, 1995). For species that are widely distributed, such as those found across different biogeographic regions, it is also important to test species status between geographically distant populations as it could be...
related to the vector-pathogen competency (McCoy, 2008; Powell, 2018). Morphological identification of Culicoides is difficult due to their small size and low within-species morphological variation which provide only limited diagnostic characters distinguishing closely related species (Harrup et al., 2015). Therefore, genetic markers such as DNA barcodes can be very helpful for testing species status (Ander et al., 2013; Harrup et al., 2016). An example of this type of study is the discovery of morphologically similar but genetically different species within the so-called C. arakawae (Arakawa) in Thailand (Pramual et al., 2021b).

Culicoides arakawae is geographically widely distributed from southern, eastern and southeastern Asia and Japan (Wirth & Hubert, 1989). DNA barcode analyses revealed that some specimens formerly identified as C. arakawae are actually different species which were then formally described as C. mahasarakhamense (Pramual et al., 2021b). Recent studies have revealed that C. mahasarakhamense is a potent vector of several disease-causing agents including Leucocytozoon, Plasmudium juxtanucleare, P. gallinaceum and Leishmania martiniquensis (Pramual et al., 2021a; Sunantaraporn et al., 2021; Songumpai et al., 2022).

In this study, we examined the genetic diversity and genetic structure based on the mitochondrial cytochrome c oxidase I (COI) gene sequences of C. peregrinus collected from cattle pens in northeastern Thailand. By using a population genetic approach, the level of genetic differentiation between local populations can be inferred. This information can then be used to indicate inter-population movement which can potentially be used for designing effective control programs (Tabachick & Black, 1995). We also examined genetic relationships between specimens of C. peregrinus from Thailand with those reported from other countries. This enabled determination of species status of populations from geographically widespread regions.

**MATERIALS AND METHODS**

**Specimen collection and identification**

Adult fly specimens of C. peregrinus were collected around cattle pens from nine locations (Table 1 and Figure 1) in northeastern Thailand between 19 November 2020 and 12 November 2022. Specimens were collected by a sweep net swept around the animal host and in the air around the cattle pens. Specimens were preserved in 80% ethanol and stored at -20°C in a freezer until use. The species was identified using keys and descriptions of Southeast Asian biting flies.

**Table 1.** Sampling locations, number of specimen (N), number of haplotypes, nucleotide diversity (\(\pi\)) and haplotype diversity (h) of Culicoides peregrinus from Thailand

<table>
<thead>
<tr>
<th>Location</th>
<th>Latitude/Longitude</th>
<th>N (no. of haplotypes)</th>
<th>(\pi)±S.E.</th>
<th>h±S.E</th>
<th>Collection date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Prangku, Si Sa Ket (SR)</td>
<td>14.83035 N/104.06043 E</td>
<td>13 (8)</td>
<td>0.0085±0.0050</td>
<td>0.8077±0.1131</td>
<td>2 Jan 2021/4 Jun 2022</td>
</tr>
<tr>
<td>2. Huai Mek, Kalasin (KS1)</td>
<td>16.60528 N/103.22320 E</td>
<td>18 (6)</td>
<td>0.0019±0.0014</td>
<td>0.5621±0.1342</td>
<td>23 Jan 2021</td>
</tr>
<tr>
<td>3. Nadun, Maha Sarakham (MK1)</td>
<td>15.68681 N/103.23162 E</td>
<td>9 (9)</td>
<td>0.0066±0.0041</td>
<td>1.0000±0.0524</td>
<td>20 Feb 2021</td>
</tr>
<tr>
<td>4. Ban Don Suan, Kantharawichai, Maha Sarakham (MK2)</td>
<td>16.25278 N/103.27525 E</td>
<td>12 (8)</td>
<td>0.0098±0.0057</td>
<td>0.9091±0.0649</td>
<td>26 Feb 2021</td>
</tr>
<tr>
<td>5. Non Sang, Nong Bua Lamphu (NL)</td>
<td>16.82201 N/102.57699 E</td>
<td>6 (6)</td>
<td>0.0047±0.0034</td>
<td>1.0000±0.0962</td>
<td>27 Feb 2021</td>
</tr>
<tr>
<td>6. Ban Hua Na Kam, Kalasin (KS2)</td>
<td>16.40762 N/103.20411 E</td>
<td>18 (10)</td>
<td>0.0063±0.0037</td>
<td>0.8497±0.0775</td>
<td>19 Nov 2020</td>
</tr>
<tr>
<td>7. Ban Pho Si, Nakhon Phanom (NP)</td>
<td>17.44680 N/104.46520 E</td>
<td>10 (9)</td>
<td>0.0200±0.011</td>
<td>0.9778±0.0540</td>
<td>20 Nov 2020</td>
</tr>
<tr>
<td>8. Khongchiam, Ubon Ratchathani (UB)</td>
<td>15.31459 N/105.49841 E</td>
<td>16 (13)</td>
<td>0.0084±0.0048</td>
<td>0.9667±0.0357</td>
<td>12 Nov 2022</td>
</tr>
<tr>
<td>9. Ban Nong Hua Mu, Udon Thani (UD)</td>
<td>17.36152 N/102.75725 E</td>
<td>10 (7)</td>
<td>0.0068±0.0042</td>
<td>0.9111±0.0773</td>
<td>13 Oct 2022</td>
</tr>
<tr>
<td>10. Trang (TR)*</td>
<td>7.83065 N/99.33763 E a</td>
<td>3 (1)*</td>
<td>0</td>
<td>0</td>
<td>29 Sep 2012 a</td>
</tr>
</tbody>
</table>

Total 115 (64) 0.0082±0.0044 0.9466±0.0138

*Data from Gopurenko et al. (2022).
midge (Wirth & Hubert, 1989) and wing pictorial illustration of Dyce et al. (2007).

**DNA extraction, polymerase chain reaction (PCR), and sequencing.** DNA was extracted from individual flies using the GF-1 Nucleic Acid Extraction Kit (Vivantis Technologies Sdn. Bhd, Malaysia). A fragment of approximately 650 bp of mitochondrial cytochrome c oxidase I (COI) was amplified using the primers LCO1490 (5'-GGTCAACAAATCATAAAGATATTGG-3') and HCO2198 (5'-TAAACTTCAGGGTGACCAAAAAATCA-3') (Folmer et al., 1994). PCR reaction conditions followed those described in Tangkawanit et al. (2018). 1% agarose gel electrophoresis was used to check the PCR products. Successful amplifications were purified using PureDireX PCR CleanUp & Gel Extraction Kit (BioHelix, Taiwan). Purified PCR products were sent for sequencing at the ATCG Company Limited (Thailand Science Park (TSP), Pathumthani, Thailand) using the same primers as for PCR.

**Data analysis**

A total of 112 COI sequences (GenBank accession nos. OR123081 - OR123192) were obtained in the present study. The COI sequences of *C. peregrinus* publicly available in NCBI GenBank and BOLDs (79 sequences) were retrieved and included in data analysis. Intraspecific genetic divergence was calculated using the uncorrected p-distance in TaxonDNA (Meier et al., 2006). Haplotype diversity and nucleotide diversity were estimated in ARLEQUIN ver. 3.5 (Excoffier & Lischer, 2010). Genetic relationships between haplotypes were inferred in the software Network ver. 10.2.0 (http://www.fluxus-engineering.com) using the median-joining (MJ) network method (Bandelt et al., 1999). To test the genetic differentiation between populations, pairwise FST values were calculated in ARLEQUIN. Tests of statistical significance of the FST values were based on 1023 permutations. To avoid type I error due to multiple comparisons, the Bonferroni correction was used to adjust the significance level. To examine relationships between level of genetic (based on FST values) and geographic distances (km) between sampling locations, the isolation-by-distance (IBD) model (Mantel, 1967) was tested. The IBD analysis was performed in IBD ver. 1.52 (Bohonak, 2002). Statistical significance of the IBD was based on 1,000 randomizations.

**RESULTS AND DISCUSSION**

Nucleotide diversity in populations ranged from 0 to 0.0098 with an overall value of 0.0082. The haplotype diversity within populations ranged between 0 and 1 with an overall value of 0.9466 (Table 1). Intraspecific genetic divergence across individual specimens (n = 112) obtained in the present study ranged between 0 – 3.14%. When COI sequences of *C. peregrinus* (n = 34) from Thailand previously reported (Jomkumsing et al., 2021; Gopurenko et al., 2022) were included, there was a slight increase of maximum intraspecific genetic divergence to 3.83%. It has been suggested that, for *Culicoides* species, the p-distance is usually <3% (Gopurenko et al., 2015). Species with >3% based on COI sequences are likely to be comprised of cryptic species, if there is also support from other evidence such as additional genetic markers or morphological characteristics or with a highly genetically different lineage at an early stage of speciation. Based on the MJ network (Figure 2) of the Thai specimens, there were no indications of deep genetic divergence of lineages except for three genetically divergent haplotypes. If these haplotypes were omitted, maximum genetic divergence within species of *C. peregrinus* in Thailand was only 2.44%. This level of intraspecific genetic divergence is similar to that reported for *C. peregrinus* from India (2.0%) (Harrup et al., 2016) although our study had much larger sample size (146 vs 10).

![Figure 2. Median joining network of 112 mitochondrial cytochrome c oxidase I (COI) sequences of Culicoides peregrinus from northeastern Thailand obtained in present study and three COI sequences from southern Thailand (TR) retrieved from Gopurenko et al. (2022). Each circle represents a haplotype and sizes are relative to number of individuals sharing such haplotype. Haplotypes are labelled according to the sampling location as in Figure 1.](image-url)
In contrast to the low level of genetic divergence, population pairwise FST analysis revealed a surprisingly high level of genetic structuring. Among 45 pairwise comparisons, 32 (71%) were significantly different (Table 2). The many population genetic studies of the Culicoides species have usually found low or no genetic structuring, particularly between geographically proximal populations (Calvo et al., 2009; Onyango et al., 2015a, 2015b; Jacquett et al., 2016a, 2016b; Mignotte et al., 2021). In Thailand, a population genetic study of the geographically co-distributed and common species, C. mahasarakhamense, also found a low level of genetic differentiation between populations (Pramal et al., 2022).

The very high level of genetic structuring between populations of C. peregrinus is unexpected given that it is among the most abundant and common species, particularly near animal farms. The high level of genetic differentiation among populations of C. peregrinus indicated that gene flow is limited. The IBD analysis revealed no significant relationship between level of genetic (i.e. FST value) and geographic distances (R² = 0.252, P = 0.084). Therefore, geographic distance cannot explain the genetic structuring found in this species. A possible explanation for the high level of genetic differentiation between populations of C. peregrinus is that there is limited movement between local populations i.e. between cattle pens. The MJ network of the specimens obtained in the present study (Figure 2) indicated more closely related haplotypes within the same population than those from others. Among 64 haplotypes identified in this study, only 6 were shared by >2 populations. This was also supported by the generally low nucleotide diversity in each population. The ability to use diverse types of breeding habitats suggest that immature stages of C. peregrinus could be using moist soil around the cattle pens. Furthermore, the generalist feeding habit enables this species to search for host blood sources within or around the cattle pens where different animals including buffalos, cows, chickens and humans are available. Taken together, these factors facilitate C. peregrinus to complete its life cycle within a cattle pen without any need for searching for host blood sources in other places. This will promote genetic differentiation between local populations.

Although genetic divergence within Thai specimens was relatively low, it was very high (max. p-distance of 5.40%) when sequences from other countries (Australia, China, India, Bangladesh) recorded in GenBank and BOLDs were included. The maximum genetic divergence was found between specimens from Thailand and Australia. This level of genetic divergence is lower than those reported in BOLDs (max. p-distance of 6.25%) for BIN BOLD:AAJ7131 (https://www.boldsystems.org/index.php/PublicBarcodeCluster?clusteruri=BOLD:AAJ7131), a single BIN recorded in BOLDs for C. peregrinus thus far (accessed on June 11, 2023). The lower value of genetic divergence in this study compared to BOLD is due to smaller sample size because many sequences from some countries (Indonesia, Timor-Leste, Malaysia, Singapore, Pakistan) used for p-distance calculation in BOLD are not publicly available and thus were not included in our analysis.

Despite the limitations of accessing COI sequences from BOLDs, we have detected cryptic genetic diversity in C. peregrinus specimens included in the present study. The MJ network analysis revealed three divergent lineages (A, B, C) (Figure 3) within C. peregrinus from Thailand, China, Bangladesh, India and Australia. Specimens from Thailand and China formed lineage A. In fact, five specimens from China shared a haplotype with those from Thailand. Sequences of C. peregrinus from Australia formed lineage B and those from India and Bangladesh belonged to lineage C. Genetic differentiation between lineage A (Thailand + China) and B (Australia) ranged between 2.44% and 5.40% and between A and C (India + Bangladesh) ranged between 1.22% – 4.18%. Lower genetic divergence was found between lineage B and C with the p-distance varying between 1.57% and 2.26%.

The type locality of C. peregrinus is in India (Wirth & Hubert, 1989), therefore, we assume that the COI sequences reported from that country are more likely to be the true C. peregrinus. Genetically divergent lineages of specimens from Thailand + China and Australia from those of India + Bangladesh could either be indicative of different, but morphologically similar species or of genetic structuring within species (Gopurenko et al., 2015). Wirth and Hubert (1989) has noted that some females collected from Chonburi province, central Thailand, are different from typical C. peregrinus by having more extensive, yellowish pale wing marking. In addition, the pale spot at the base of the mediocubital fork is continuous along veins M3+4 and Cu1 to the wing margin. Our specimens used in this study collected from northeastern Thailand, have the pale wing marking (Figure 4) which agrees with those from central Thailand described in Wirth and Hubert (1989). In addition, based on wing photographs of C. peregrinus from China provided in Liu et al. (2018) from which the COI sequences (accession nos. KY433456 – 59) were included in present study, these variations of pale wing marking were also found in their specimens. This is in agreement with the COI sequences where Thailand and China belonged to the same lineage. For those specimens of C. peregrinus from Australia, according to the wing photograph (Dyce et al., 2007), no such variations were found. Based on the COI barcoding sequences and morphological differentiations, we hypothesize that specimens so-called C. peregrinus in Thailand and China are different species from those in India and Australia. To test this hypothesis, further study using additional independent molecular marker from nuclear genes is needed to resolve species status of these COI genetically divergent lineages (Gopurenko et al., 2015).

Table 2. Population pairwise FST values between 10 populations of Culicoides peregrinus from Thailand

<table>
<thead>
<tr>
<th>SR</th>
<th>KS1</th>
<th>MK1</th>
<th>MK2</th>
<th>NL</th>
<th>KS2</th>
<th>NP</th>
<th>UB</th>
<th>UD</th>
<th>TR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KS1</td>
<td>0.08472</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MK1</td>
<td>0.16539</td>
<td>0.36470</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MK2</td>
<td>0.13525</td>
<td>0.20072</td>
<td>0.23096</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>0.07576</td>
<td>0.13991</td>
<td>0.23445</td>
<td>0.09923</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KS2</td>
<td>0.35899</td>
<td>0.44416</td>
<td>0.39091</td>
<td>0.09777</td>
<td>0.35503</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP</td>
<td>0.11227</td>
<td>0.28477</td>
<td>0.14857</td>
<td>0.18794</td>
<td>0.17016</td>
<td>0.37281</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UB</td>
<td>0.36576</td>
<td>0.47122</td>
<td>0.41449</td>
<td>0.1779</td>
<td>0.36474</td>
<td>0.08434</td>
<td>0.30978</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UD</td>
<td>0.23869</td>
<td>0.34084</td>
<td>0.33406</td>
<td>0.00078</td>
<td>0.22682</td>
<td>-0.00259</td>
<td>0.25978</td>
<td>0.06262</td>
<td></td>
</tr>
<tr>
<td>TR</td>
<td>0.70778</td>
<td>0.85877</td>
<td>0.68557</td>
<td>0.31396</td>
<td>0.75177</td>
<td>0.30466</td>
<td>0.58066</td>
<td>0.11713</td>
<td>0.31494</td>
</tr>
</tbody>
</table>

Note: Bold number indicates statistical significance after Bonferroni correction.
Figure 3. Median joining network of 191 mitochondrial cytochrome c oxidase I (COI) sequences of *Culicoides peregrinus* from northeastern Thailand and those reported from other countries reported in public database. Each circle represents a haplotype and sizes are relative to number of individuals sharing such haplotype. Haplotypes are labelled according to the country of origin.

Figure 4. Photographs of female right wings of *Culicoides peregrinus* in northeastern Thailand collected in this study. Arrow indicates variations of the pale spot at the base of mediocubital fold and along the veins M3+4 and Cu1.
In conclusion, we found a high level of genetic structuring within *C. peregrinus* from northeastern Thailand. The ability to utilize diverse breeding habitats as well as a generalist feeding habit could enable *C. peregrinus* to complete its life cycle within cattle pens. This would limit dispersal for host searching or mating and thereby facilitate genetic differentiation even between geographically proximal populations. At the wider geographic scale (i.e. comparisons of specimens from different countries), we found three genetically divergent lineages within *C. peregrinus*. Specimens from Thailand and China belonged to the same lineage and were different from those of India + Bangladesh and Australian lineages. The genetically divergent lineages based on COI sequences were also supported by morphological characteristics. Therefore, these lineages are potentially different, but morphologically similar species. Additional study using the independent genetic markers from nuclear DNA genes will be useful to resolve the species status of *C. peregrinus* recorded in Thailand.

ACKNOWLEDGEMENTS

This study was financially supported by Mahasarakham University (Postgraduateotoral Fellowship). We would like to thank Dr. Adrian Plant for valuable comments and English language editing of the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

REFERENCES


