

RESEARCH ARTICLE

# Population Genetic Structure and Potential Incursion Pathways of the Bluetongue Virus Vector *Culicoides brevitarsis* (Diptera: Ceratopogonidae) in Australia

W. T. Tay<sup>1\*</sup>, P. J. Kerr<sup>1,2</sup>, L. S. Jermin<sup>1</sup>

**1** CSIRO, Black Mountain Laboratories, Canberra, ACT, 2601, Australia, **2** School of Biological Sciences, The University of Sydney, Sydney, 2006, Australia

\* [weetek.tay@csiro.au](mailto:weetek.tay@csiro.au)



**OPEN ACCESS**

**Citation:** Tay WT, Kerr PJ, Jermin LS (2016) Population Genetic Structure and Potential Incursion Pathways of the Bluetongue Virus Vector *Culicoides brevitarsis* (Diptera: Ceratopogonidae) in Australia. PLoS ONE 11(1): e0146699. doi:10.1371/journal.pone.0146699

**Editor:** Tzen-Yuh Chiang, National Cheng-Kung University, TAIWAN

**Received:** April 18, 2015

**Accepted:** December 20, 2015

**Published:** January 15, 2016

**Copyright:** © 2016 Tay et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All sequence data are available from GenBank (GenBank accession numbers: KP201844-KP201872). Aligned DNA and amino acid sequences used in phylogenetic study are provided as S1 and S2 Files, respectively.

**Funding:** Funding was by the Australian Biosecurity CRC. The funder has no role in the project design, data analyses, and interpretation of results.

**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

*Culicoides brevitarsis* is a vector of the bluetongue virus (BTV), which infects sheep and cattle. It is an invasive species in Australia with an assumed Asian/South East Asian origin. Using one mitochondrial marker (i.e., part of the cytochrome oxidase subunit I gene) and six nuclear markers, we inferred population genetic structure and possible incursion pathways for Australian *C. brevitarsis*. Nine mitochondrial haplotypes, with low nucleotide sequence diversity (0.0–0.7%) among these, were identified in a sample of 70 individuals from seven sites. Both sets of markers revealed a homogeneous population structure, albeit with evidence of isolation by distance and two genetically distinct clusters distributed along a north-to-south cline. No evidence of a cryptic species complex was found. The geographical distribution of the mitochondrial haplotypes is consistent with at least two incursion pathways into Australia since the arrival of suitable livestock hosts. By contrast, 15 mitochondrial haplotypes, with up to four times greater nucleotide sequence diversity (0.0–2.9%) among these, were identified in a sample of 16 individuals of the endemic *C. marksii* (sampled from a site in South Australia and another in New South Wales). A phylogenetic tree inferred using the mitochondrial marker revealed that the Australian and Japanese samples of *C. brevitarsis* are as evolutionarily different from one another as some of the other Australian species (e.g., *C. marksii*, *C. henryi*, *C. pallidothorax*) are. The phylogenetic tree placed four of the species endemic to Australia (*C. pallidothorax*, *C. bundyensis*, *C. marksii*, *C. henryi*) in a clade, with a fifth such species (*C. bunrooensis*) sharing a common ancestor with that clade and a clade comprising two Japanese species (*C. verbosus*, *C. kibunensis*).

## Introduction

Accurate delineation of a vector species' status and its population and evolutionary genetics are fundamental to biosecurity and mitigation of the emergence of exotic diseases. Cryptic species (i.e., morphologically indistinguishable but genetically different species) are widespread among eukaryotes and prokaryotes, including all major orders of insects (e.g., Lepidoptera [1], Diptera

[2], and Hemiptera [3]). The detection of cryptic species in the dipteran *Culicoides variipennis* species complex and the subgenus *Culicoides* [4] had major implications for our understanding of vector competence for Bluetongue virus (BTV) [5].

Delineation of cryptic species and rapid identification of morphologically similar species are now feasible using genetic markers such as the nuclear ribosomal RNA (rRNA) genes (e.g., [6–9], but see [10] for caution relating to concerted evolution and the use of rRNA as a genetic marker) and the mitochondrial cytochrome oxidase subunit I (COI) [4, 11–14] and cytochrome *b* (Cyt *b*) [15] genes. Important population genetics features, such as migration, gene flow, and mating behaviour (e.g., [16–18]), all benefit from the use of nuclear DNA (nuDNA) and mitochondrial DNA (mtDNA) (e.g., [19–22]), and the use of mtDNA to delineate, and infer the phylogeny of, closely related species has been augmented by using nuDNA markers and morphological data (e.g., [14, 20, 23]).

Globally, there are currently 1401 recognized species (1355 extant species, 46 fossil species) and 31 subgenera of *Culicoides* [24–26], and many of them are thought to be cryptic species groups. Traditionally, species identification of *Culicoides* has relied on external (e.g., head or wing patterns [12, 14]) and internal (e.g., genitalia [27, 28]) traits. Using external traits to identify species can lead to confusion and hamper the identification and confirmation of virus vector status and competency (e.g., *C. brevitarsis*, *C. actoni*, and *C. suzukii* in the *Avaritia* group [29], and the *C. variipennis* species complex [5]). By contrast, different molecular markers have been used successfully to identify and delineate species of *Culicoides* from Africa, Europe, and the USA (e.g., [9, 12, 30–38]), but so far not from Australia.

In Australia, *Culicoides* consists of introduced and endemic species [39]. Dyce *et al.* [27] divided the Australasian species of *Culicoides* (145 described, 120 un-described) into 23 subgenera, with the subgenus *Avaritia* divided into five complexes that contain four introduced BTV vector species (i.e., *C. actoni* (*Actoni* complex), *C. wadai* (*Boophagus* complex), *C. brevitarsis* (*Imicola* complex), and *C. fulvus* (*Orientalis* complex)) of Asian or South-east Asian origins [39, 40]. Oviposition sites are highly specialised in *Culicoides* species [41, 42]. Oviposition by *C. brevitarsis* appears to be limited to cattle dung [42–44], with adults feeding on both cattle and sheep [45, 46]. Hence, it seems likely that *C. brevitarsis* only was capable of founding breeding populations in Australia after the introduction of domesticated cattle by European settlers.

*Culicoides wadai*, *C. actoni* and *C. fulvus*, while competent vectors of BTV, have only been found in northern Australia [47]. BTV is asymptomatic in cattle across the ranges of *C. brevitarsis*, *C. wadai*, *C. fulvus*, and *C. actoni* in Australia. By contrast, sheep are highly susceptible to BTV but are farmed predominantly in the southern parts of Australia, where they overlap geographically with *C. brevitarsis* in its southern-most range of New South Wales (NSW). Therefore, a transfer of BTV into sheep could occur in these parts of Australia, posing a risk to this region's livestock industry.

Given that *C. brevitarsis* is a competent vector of BTV and regarded as the most likely vector of BTV between Australian cattle and sheep, it is important that a better understanding of this species be developed. In particular, information on its population genetic structure is needed, but we also need to determine whether there is evidence of it being a cryptic species complex, whose members could have different levels of vector competency for BTV. Finally, because *C. brevitarsis* may have been recently introduced to Australia, comparing Southeast Asian and Australian samples of *Culicoides* might reveal the evolutionary relationship among these samples, and point to the likely incursion pathways. Such information could become critical for a successful Australian response to a BTV epizootic event like that in Europe (reviewed in [48–50]). The *C. variipennis* species complex illustrates the importance of such information: different subspecies in this complex have different levels of susceptibility to BTV infection and different virus competency rates [5].

Using molecular data obtained from Australian samples of *C. brevitarsis* and *C. marksi* (included for comparative purposes) as well as published sources (GenBank), this study aims to address some of the issues mentioned above. In particular, it aims to determine whether Australian samples of *C. brevitarsis*: (i) are genetically diverse; (ii) form a hitherto unknown cryptic species complex; and (iii) differ genetically from their Southeast Asian counterparts.

## Materials and Methods

### Sampling method, sampling sites and species identification

*Culicoides brevitarsis* and *C. marksi* were collected using light traps at the National Arbovirus Monitoring Program's (NAMP) monitoring sites in the New South Wales (NSW), Northern Territory (NT), Western Australia (WA), and South Australia (SA) (Table 1). Australian state, territory, and federal government officers are responsible for monitoring the NAMP. Samples obtained for this study were provided in-kind by the relevant government departments via the Officer in Charge, so no specific written permissions were required. The Officer in Charge at the various government departments (e.g., NSW Department of Primary Industry; Department of Agriculture, Forestry and Fishery; Department of Agriculture; and Northern Australia Quarantine Strategy) identified the name of each specimen using the keys by Dyce *et al.* [27]. *Culicoides brevitarsis* were sampled from: Kalumburu in WA ('A' 29-Dec-08; 'B' 25-Jun-09), Douglas-Daly (28-Nov-08) and Katherine (28-Feb-08) in NT, and Lismore (14-Jan-08), Grafton (14-Dec-07), Bellingen (14-Jan-08), and Paterson (6-Feb-08) in NSW. Two populations of *C. marksi* (n = 16), a species endemic to Australia, were also sampled from Murray Bridge in SA (20-Jan-09) and Moree in NSW (10-Feb-08) to enable comparison of nucleotide sequence diversity at the population level (Fig 1). Four other Australian endemic species (i.e., *C. henryi* (Lismore; 13-Nov-07), *C. bundyensis* (Kalumburu; 19-Dec-08), *C. pallidothorax* (Kalumburu; 29-Dec-08), and *C. bunrooensis* (Denman, NSW; 13-Mar-08)) were included in the study to allow direct comparison of estimates of inter-species nucleotide sequence diversity and contribute to our understanding of phylogenetic relationships between Australian and Asian/South East Asian *Culicoides* species. Individuals of *C. pallidothorax*, *C. henryi*, *C. bundyensis*, and *C. bunrooensis* from this study have been deposited in the Australian National Insect Collection's database (ANIC Database: No. 29035987, 2903588, 29035989, 29035990, respectively) and thus are available to future research efforts, such as characterisation of the species based on morphological characters (e.g., wing patterns). Wing patterns and partial mtDNA COI sequence [GenBank: KJ162974] of a *C. brevitarsis* individual from Grafton (also collected by H. McKenzie on the same date) have been presented [51].

### DNA extraction, mtDNA PCR markers design, PCR conditions, and DNA sequencing

Genomic DNA (gDNA) was extracted from individual specimens of *Culicoides* using Qiagen's DNeasy Blood and Tissue Kit (Cat. # 69506) following the recommended protocol. gDNA from individual samples were eluted in 5µl of the elution buffer and stored at -20°C. Forward and reverse primers for PCR amplification of partial mtDNA COI regions (Table 2) were derived from a multiple sequence alignment of mtDNA COI sequences from: *C. brevitarsis* [GenBank: AB360994], *C. wadai* [GenBank: AB361005], *C. dubius* [GenBank: AB361000], and *C. oxystoma* [GenBank: AB361002]. PCR amplification of each sample was carried out in a total of 25µL PCR reaction volume that included 2.5µL of gDNA, 0.25mM of both forward and reverse primers, 0.1mM of dNTP's, 1x PCR buffer (NEB), 1.5mM of MgCl<sub>2</sub>, and 0.025U of Taq DNA Polymerase (NEB). PCR involved an initial 5 minutes template-denaturing step at 95°C,

**Table 1. *Culicoides* samples used in this study including known virus vectored and taxonomic status.**

Location	<i>Culicoides</i> spp.	N	Subgenus <sup>1</sup> /Sp. Gp. <sup>2</sup> /Complex <sup>3</sup>	Date collected	Viruses vectored
Kalumburu, WA	<i>pallidothorax</i>	1	unplaced <sup>1a</sup> / <i>Victoriae</i> <sup>1a</sup>	29/12/08	?
	<i>bundyensis</i>	1	unplaced <sup>1a</sup> / <i>Victoriae</i> <sup>1a</sup>	19/12/08	?
	<i>brevitarsis</i> (A)	9	<i>Avaritia</i> <sup>1a</sup> / <i>Imicola</i> <sup>3a</sup>	29/12/08	BTV/BEF/EHDV/AV
	<i>brevitarsis</i> (B)	8	<i>Avaritia</i> <sup>1a</sup> / <i>Imicola</i> <sup>3a</sup>	25/06/09	BTV/BEF/EHDV/AV
Douglas-Daly, NT	<i>brevitarsis</i>	7	<i>Avaritia</i> <sup>1a</sup> / <i>Imicola</i> <sup>3a</sup>	28/11/08	BTV/BEF/EHDV/AV
Katherine, NT	<i>brevitarsis</i>	4	<i>Avaritia</i> <sup>1a</sup> / <i>Imicola</i> <sup>3a</sup>	28/2/08	BTV/BEF/EHDV/AV
Lismore, NSW	<i>henryi</i>	2	unplaced <sup>1a</sup> / <i>Victoriae</i> <sup>2a</sup>	13/11/07	?
	<i>brevitarsis</i>	7	<i>Avaritia</i> <sup>1a</sup> / <i>Imicola</i> <sup>3a</sup>	13/11/07	BTV/BEF/EHDV/AV
Grafton, NSW	<i>brevitarsis</i>	12	<i>Avaritia</i> <sup>1a</sup> / <i>Imicola</i> <sup>3a</sup>	14/12/07	BTV/BEF/EHDV/AV
Bellingen, NSW	<i>brevitarsis</i>	4	<i>Avaritia</i> <sup>1a</sup> / <i>Imicola</i> <sup>3a</sup>	14/01/08	BTV/BEF/EHDV/AV
Paterson, NSW	<i>brevitarsis</i>	19	<i>Avaritia</i> <sup>1a</sup> / <i>Imicola</i> <sup>3a</sup>	06/02/08	BTV/BEF/EHDV/AV
Denman, NSW	<i>bunrooensis</i>	1	unplaced <sup>1a</sup> / <i>Clavipalpis</i> <sup>2a</sup>	13/03/08	?
Moree, NSW	<i>marksi</i>	9	<i>Marksomyia</i> <sup>1a</sup> / <i>Marks</i> <sup>2a</sup>	10/2/08	BEF <sup>†</sup>
Murray Bridge, NSW	<i>marksi</i>	7	<i>Marksomyia</i> <sup>1a</sup> / <i>Marks</i> <sup>2a</sup>	20/1/09	BEF <sup>†</sup>
GenBank Accession Number	<i>Culicoides</i> spp.	N		Nucleotide positions (start.end)	
AB360994.1, AB360995.1	<i>brevitarsis</i>	2	<i>Avaritia</i> <sup>1a</sup> / <i>Imicola</i> <sup>3a</sup>	326.872 (123. .957) <sup>e</sup>	BTV/BEF/EHDV/AV
KJ162974.1	<i>brevitarsis</i>	1	<i>Avaritia</i> <sup>1a</sup> / <i>Imicola</i> <sup>3a</sup>	365.704	BTV/BEF/EHDV/AV
KJ162966.1, KJ162972.1, KJ162973.1	<i>brevitarsis</i>	3	<i>Avaritia</i> <sup>1a</sup> / <i>Imicola</i> <sup>3a</sup>	324.646	BTV/BEF/EHDV/AV
KJ162971.1	<i>brevitarsis</i>	1	<i>Avaritia</i> <sup>1a</sup> / <i>Imicola</i> <sup>3a</sup>	324.646	BTV/BEF/EHDV/AV
KJ162970.1, KJ162969.1, KJ162967.1	<i>brevitarsis</i>	3	<i>Avaritia</i> <sup>1a</sup> / <i>Imicola</i> <sup>3a</sup>	324.658	BTV/BEF/EHDV/AV
KJ162975.1	<i>brevitarsis</i>	1	<i>Avaritia</i> <sup>1a</sup> / <i>Imicola</i> <sup>3a</sup>	324.646	BTV/BEF/EHDV/AV
KJ162968.1	<i>brevitarsis</i>	1	<i>Avaritia</i> <sup>1a</sup> / <i>Imicola</i> <sup>3a</sup>	324.646	BTV/BEF/EHDV/AV
AB361000.1	<i>dubius</i>	1	<i>Culicoides</i> <sup>1b</sup>	578.1,124	?
AB360979.1	<i>oxystoma</i>	8	<i>Remmia</i> <sup>1a</sup> / ( <i>Oecacta</i> <sup>1c</sup> )?	4,679.5,225	BTV
AB360978.1, AB360980.1, AB360981.1, AB360982.1, AB360983.1, AB360984.1, AB360985.1				326.872	BTV
AB361001.1, AB361002.1	<i>japonicus</i>	2	<i>Beltranmyia</i> <sup>1b,e</sup>	326.872	?
AB646612.1, AB646613.1	<i>pictimargo</i>	2	<i>Oecacta</i> <sup>1d</sup>	326.872	?
AB646615.1	<i>verbosus</i>	1	<i>Oecacta</i> <sup>1c</sup>	326.872	?
AB360973.1, AB360974.1, AB360975.1, AB360972.1	<i>arakawae</i>	5	<i>Meijerehelea</i> <sup>1a,d</sup> / ( <i>Beltranmyia</i> <sup>1b</sup> )?	541.1,087	BTV
AB361004.1				1,994.2,540	
AB364649.1, AB364650.1	<i>matsuzawai</i>	2	<i>Trithecoides</i> <sup>1b</sup>	326.872	?
AB646611.1	<i>kibunensis</i>	2	<i>Oecacta</i> <sup>1c</sup>	326.872	?
AB646610.1				1,376.1,922	

(Continued)

Table 1. (Continued)

AB360988.1, AB360989.1	<i>punctatus</i>	2	<i>Culicoides</i> <sup>b</sup>	326.872	?
AB361005.1	<i>wadai</i>	3	<i>Avaritia</i> <sup>1a</sup> / <i>Boophagus</i> <sup>3a</sup>	1,885.2,431	BTV
AB360996.1, AB360997.1				578.1,124	
AB360999.1	<i>nipponensis</i>	1	<i>Culicoides</i> <sup>b</sup>	326.875	?
AB361003.1	<i>peregrinus</i>	1	<i>Hoffmania</i> <sup>1a</sup> / ( <i>Culicoides</i> <sup>1b</sup> )?	326.875	BTV
AB360990.1, AB360991.1	<i>jacobsoni</i>	2	<i>Avaritia</i> <sup>1a</sup> / <i>Jacobsoni</i> <sup>3a</sup>	326.872	?
AB361006.1	<i>cylindratus</i>	2	<i>Culicoides</i> <sup>1b</sup> / ( <i>Hoffmania</i> <sup>1d</sup> )?	323.869	?
AB361007.1				5,692.6,238	
AB360998.1	<i>brevipalpis</i>	1	<i>Avaritia</i> <sup>1a</sup> / <i>Imicola</i> <sup>3a</sup>	326.872	BTV
AB364651.1	<i>humeralis</i>	3	<i>Trithecoides</i> <sup>1b</sup>	1,809.2,355	?
AB360992.1, AB360993.1				326.872	
AB360986.1, AB360987.1	<i>maculatus</i>	2	<i>Avaritia</i> <sup>1a</sup> / <i>Orientalis</i> <sup>3</sup>	326.872	?
AB360971.1	<i>actoni</i>	1	<i>Avaritia</i> <sup>1a</sup> / <i>Actoni</i> <sup>3</sup>	6,015.6,561	BTV
AB360976.1, AB360977.1	<i>ohmorii</i>	2	<i>Hoffmania</i> <sup>1a</sup> / ( <i>Culicoides</i> <sup>1b</sup> )?	326.872	?

*Culicoides* samples used in this study including known virus vectored and taxonomic status. *Culicoides* species collected from Australian sites (Western Australia (WA); Northern Territory (NT); New South Wales (NSW); South Australia (SA)), as well as from GenBank used in this study. Number of individuals used (N), collection dates, taxonomic classification for subgenus, species group (Sp. Gp.) and complex also indicated. The nucleotide start and end positions for GenBank samples included in the calculation of nucleotide diversities are listed. Viruses vectored are: bluetongue virus (BTV), bovine ephemeral fever (BEF), epizootic haemorrhagic disease virus (EHDV), Akabane virus (AV), unknown (?).

† indicates virus replication detected but transmission status not confirmed.

**Note:** Dyce *et al.* [27]

(a); Encyclopedia Of Life <<http://eol.org>>

(b); Arthropod Vectors of interest for Animal health database <<http://avabase.cirad.fr/>>

(c); FLYTREE Assembling the Diptera Tree of Life <http://www.inhs.illinois.edu/files/9613/9136/7590/CulicoidesSubgenera.pdf>

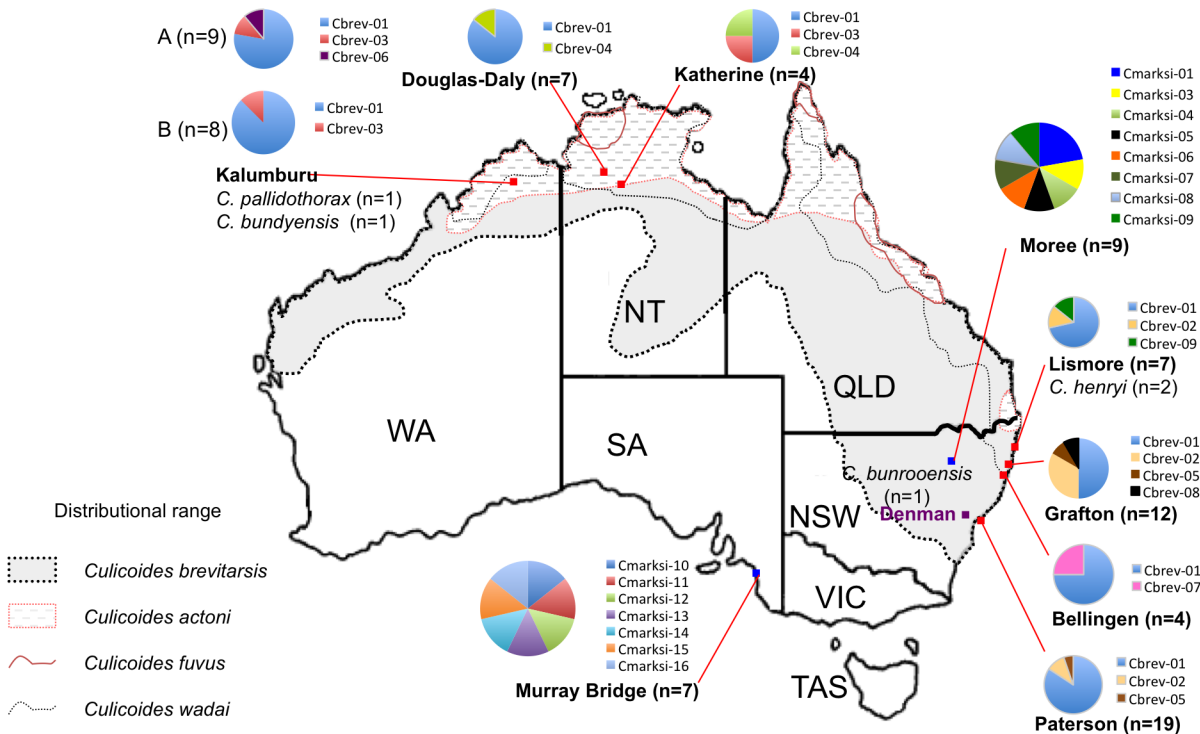
(d). The subgenus status for the following species is uncertain: (i) *C. oxystoma* is unclear and has been grouped within the *Remmia* subgenus by Dyce *et al.* [27] or within the subgenus *Oecacta* <<http://www.inhs.uiuc.edu/research/FLYTREE/index.html>>; (ii) *C. peregrinus* and *C. ohmorii* are either placed in the subgenera of *Hoffmania* [27] or *Culicoides* <<http://avabase.cirad.fr/>>; (iii) *C. wadai* placed either in the subgenera of *Avaritia* [27] or *Boophagus* (FLYTREE); (iv) *C. arakawae* placed either in the subgenera of *Meijerehelea* [27]; FLYTREE) or *Beltranmyia* <<http://avabase.cirad.fr/>>.

(e) 835bp also used in pairwise sequence divergence estimates between *C. brevitarsis* haplotypes from Australia and Japan. *C. maculatus* (= *C. tainanus*)<sup>a</sup>

doi:10.1371/journal.pone.0146699.t001

followed by 35 cycles of denaturation at 95°C, primer annealing at 50°C, and template extension at 72°C of 1 minute each. A final template-extension step at 72°C for 5 minutes was included after the 35 PCR cycles, before incubation at 10°C. PCR products (amplicons) were confirmed by gel electrophoresis (1% 1xTAE agarose gel) stained with GelRed (Biotium; cat # 41003) at 85V for 45 minutes.

Amplicons from individual samples were purified and eluted in 25µL ddH<sub>2</sub>O using Qiagen's Qiaquick PCR purification kit (Cat. no. 28104). For sequencing of individual samples, 2µL of purified amplicon for each sample was used as template in individual DNA sequencing reactions, which also contained 1.6mM primer, 1.5µL of 5x BigDye sequencing buffer and 1µL of BigDye V3 sequencing reaction mix, made up to a final volume of 10µL. The sequencing profile followed that of Behere *et al.* [16] prior to electrophoresis at John Curtin Medical School of Research, Australian National University, Canberra, Australia.



**Fig 1. *Culicoides* sampling sites.** Geographical range of the four BTV competent *Culicoides* vector species in Australia (modified from [40]) and locations from which samples of *Culicoides* were collected for this study: *C. brevitarsis* (from Kalumburu, Western Australia (WA); Douglas-Daly and Katherine, Northern Territory (NT); Lismore, Grafton, Bellingen, and Paterson, New South Wales (NSW)); *C. marksii* (from Moree, NSW; Murray Bridge, South Australia (SA)); *C. pallidothorax* and *C. bundyensis* (also from Kalumburu, WA); *C. henryi* (from Lismore, NSW); and *C. bunrooensis* (from Denman, NSW). *Culicoides brevitarsis* and *C. marksii* mtDNA haplotypes (respectively labelled 'Cbrev' and 'Cmarks') identified from all sampling sites are also indicated. *C. brevitarsis* and *C. marksii* individuals analysed per site are indicated by 'n'. Kalumburu sampling dates are 29-Dec-08 'A'; 25-Jun-09 'B'.

doi:10.1371/journal.pone.0146699.g001

## Design of exon-primed intron-crossing markers and microsatellite DNA markers

We also designed ribosomal protein (Rp) gene exon-primed intron-crossing (EPIC) markers for *C. brevitarsis* based on the methods reported in [52] and [19]. Briefly, six *C. sonorensis* Rp mRNA sequence entries in GenBank (RpL27A [GenBank: AAV84242.1], RpS2A [GenBank: AAV84247.1], RpS2B [GenBank: AAV84248.1], RpS4 [GenBank: AAV84250.1], RpS6 [GenBank: AAV84251.1], RpS8 [GenBank: AAV84252.1]) were searched for the presence of introns of between 120bp and 400bp by comparison with the homologs in *Bombyx mori*, and primers anchoring exon boundaries were designed to maximise conserved sequences between *Aedes aegypti*, *Ae. albopictus*, *Anopheles gambiae*, and *Culex quinquefasciatus*. EPIC primers were designed using Oligo version 7.17 (Molecular Biology Insights, Inc. Cascade, CO 80809, USA) with criteria that included  $T_m \geq 60^\circ\text{C}$ , minimal/no primer duplex and hairpin formation, and minimal false priming sites. The resulting primers were examined for PCR amplification efficacies, expected amplicon lengths, and the presence of polymorphism in selected samples of *C. brevitarsis* from Lismore, Katherine, Douglas, Kalumburu, and Paterson prior to the population-wide screening and scoring.

Microsatellite DNA markers were designed by searching reported microsatellite loci sequences available in the GenBank. We followed the methods of Tay *et al.* [53] to screen against the potential of incorporating flanking primer-annealing sequences that might be due

**Table 2. Primers for mtDNA COI regions and exon-primed intron-crossing (EPIC) primers for ribosomal protein nuclear genes used to infer population structure of *Culicoides* species.**

Primer name	Primer sequence	Temp (°C)	Size (bp)	Fluorescent dye
CO1R1	CCAACWGTAAAYATRTGATGWGCTC	54		
CO1F2A	GCWGTWAATTTYWTWACHACWATTAT	48		
CO1R2	CTATGTTCDGHWGGNGGWARATTTTG	56		
CO1209F	CWATYATAATTGGDGGRTTYGG	51		
CO1265F	GATATRGCVWTTYCCKCGWATAAATAA	52		
CO1R1295	AATATCGTCGAGGTATTCTCTGA	51		
CO1F272	ATTAGGAGCTCCTGATATAGC	50		
CO1hmbF	GCYATTTTATTAYTWTATCTTTAC	46		
CO1hmbR	ACATAATGRAARTGWGCTACTAC	50		
RpS2B-f	GGYGTSARTGCAGCAAGGAAGT		509	FAM
RpS2B-r	TGTGGCCTTKGCRAAGTTACC			
RpS6-f	TAGGYGATGAATGGAARGGCTATGT		234	FAM
RpS6-r	CAATRCACCCACGAACKGAYTTACG			
RpL27A-f	GAAARAAGACCAGRAAGCTACG		233	TET
RpL27A-r	CGAARTAWCCRRGGRTGGTAYTTGTC			
RpS8-f	TGCMACBGGTGGYAAGAAGG		163	TET
RpS8-r	GGYATCHAGACGVAGVGCAC			
Cimi69-f	TCACAAGGAGTCGTTACTTTTGC		232	TET
Cimi69-r	CAAGGATGAATAACAGTAAACAG			
Cimi85-f	CAATAACACTCGCTGACTTGCTG		241	FAM
Cimi85-r	TGCCTATACATGTTAGTAGATAG			

Primer sequences for partial mtDNA COI region, *Culicoides* ribosomal protein (Rp) gene EPIC markers, and microsatellite DNA markers (Cimi; based on *C. imicola* sequences [GenBank: DQ008972; GenBank: DQ008973] that contained simple sequence repeat units) used to infer population genetic structure of *Culicoides* species in Australia. Primer annealing temperature (°C) and fluorescence dye modifications (FAM, TET) added to 5' end of EPIC-PCR and microsatellite DNA markers are provided. IUPAC nucleotide ambiguity codes used. Approximate expected amplicon size (in bp) for individual nuclear DNA and mtDNA markers are also provided.

**Note:** expected amplicon sizes for: COIF2A/COIR1: 407bp; COIF2A/COIR2: 1,031bp; COIF2A/COIR1295: 839bp; COI209F/COIR1: 673bp; COI209F/COIR2: 1,297; COI209F/COIR1295: 1,105bp; COIF272/COIR1: 630bp; COIF272/COIR2: 1,254bp; COIF272/COIR1295: 1,062bp; COI209F/COIR1295: 980bp; CO1hmbF/CO1hmbR: 547bp.

doi:10.1371/journal.pone.0146699.t002

to Transposable Elements (TEs). Briefly, four microsatellite DNA loci (i.e., Cimi-12 [GenBank: DQ008970], Cimi-66 [GenBank: DQ008971], Cimi-69 [GenBank: DQ008972] and Cimi-85 [GenBank: DQ008973]) were searched for possible target-site duplication sequences, as well as reverse-translated using BLASTX [54] to identify the likely presence of a reverse transcriptase gene, and to search for the presence of target-site duplication sequences that would indicate Class I retrotransposable TE sequence insertions. We note that using BLASTX to identify the presence of reverse transcriptase is only effective against Class I (RNA-associated/'copy and paste') TEs but ineffective against Class II (DNA-base/'cut and paste') TEs (see [55]).

Conditions used during PCR amplification and expected amplicon sizes for the EPIC and microsatellite DNA markers are provided in Table 2. We used the Fragment Profiler software within the MegaBACE Genetic Profiler suite v1.2 (GE Healthcare Life Sciences, UK) for allele scoring of EPIC and microsatellite DNA loci of biting midges samples in this study. EPIC and microsatellite DNA markers were labelled either with TET or FAM [19], and electrophoresis of PCR amplicons was carried out at the Genetic Analysis Facility, James Cook University (Townsville, Queensland, Australia), as described in Behere *et al.* [19].

## Inter- and intra-species nucleotide diversity and genetic distances in *Culicoides*

Primary sequencing reads were assembled into contigs using Pregap4 and Gap4 from the Staden package [56] and the resulting contigs were then aligned using Se-AL version 2.0a11 (<http://tree.bio.ed.ac.uk/software/seal/>). To identify unique haplotypes, the contigs were clustered using a K-align rapid multiple sequence alignment (MSA) algorithm for DNA sequences (<http://www.ebi.ac.uk/Tools/msa/kalign>). Nucleotide sequence diversity among the *C. brevitarsis* haplotypes (i.e., Cbrev-01, . . . , Cbrev-09; 835 base pair (bp)), and among the *C. marksii* haplotypes (Cmarksii-01, Cmarksii-03, . . . , Cmarksii-16 (547bp) were calculated as: (i) *p* distances (i.e., the proportion of sites with different nucleotides within pairs of sequences); or (ii) evolutionary distances using Tamura and Nei's [57] nucleotide substitution model. In practice, we used MEGA version 5.05 [58] to obtain these estimates. Rate-heterogeneity across sites was modelled using a discrete  $\Gamma$  distribution with four rate categories and a shape parameter ( $\alpha = 0.1759$ ; model inferred using MODELTEST version 3.7 [59]). Standard errors of genetic distance estimates were obtained using nonparametric bootstrapping with 500 replicates. Mean genetic distances between *Culicoides* species were also obtained from an alignment of 455bp under the conditions described above for the within-species comparison. Unique haplotypes have been deposited in GenBank [GenBank accessions: KP201844-KP201872].

MtDNA COI sequences from *Culicoides* species previously deposited into GenBank, covering the same 547bp region as the *C. marksii* haplotypes reported in this study, were also included (Table 1) in our study in order to estimate nucleotide sequence divergence between and within species. GenBank entries were found using BLASTN [54] search with *Culicoides* as the Entrez [60] query against the Cmarksii-01 haplotype query sequence and specifying 'Other (nr etc.)' as the database.

## Comparison of *C. brevitarsis* from Australia, Solomon Islands, Timor-Leste, China and Japan

Twelve *C. brevitarsis* mtDNA COI sequences, ranging in length from 646bp to 703bp, were available through GenBank [51, 61]. Of these samples, two were from Japan [GenBank: AB360994, AB360995], one from Australia (Grafton [GenBank: KJ162974]), three from China (Hainan Province [GenBank: KJ162966, KJ162972, KJ162973]), four from Solomon Islands (Malaita [GenBank: KJ162971], Western Province [GenBank: KJ162970, KJ162969, KJ162967]), and two from Timor-Leste (Oecusse [GenBank: KJ162975], Cova Lima [GenBank: KJ162968]). These published mtDNA COI haplotypes allowed us to formulate different hypotheses regarding possible incursions of *C. brevitarsis* into Australia.

The 10 *C. brevitarsis* mtDNA COI sequences of Bellis *et al.* [51] represented four haplotypes, of which the samples from Grafton, Cova Lima, and the Solomon Islands represented one haplotype, the Chinese samples represented two haplotypes, and the sample from Oecusse represented the final haplotype. For the purpose of aligning the mtDNA COI haplotypes by Bellis *et al.* [51] with those generated by the present study, all *C. brevitarsis* sequences were trimmed either at the 5' and/or 3' regions, resulting in overlapping of either 323bp (China,  $n = 3$ ; Timor-Leste,  $n = 2$ ; Solomon Islands,  $n = 1$ ), 335bp (Solomon Islands,  $n = 3$ ), or 339bp (Australian,  $n = 1$ ; Japanese,  $n = 2$ ) with our *C. brevitarsis* haplotypes. Reducing the length of the alignment of our *C. brevitarsis* mtDNA COI region to nucleotide positions 1 to 339 (i.e., the shared region with GenBank samples) also reduced the total number of *C. brevitarsis* mtDNA haplotypes identified from nine to seven (i.e., Cbrev-01/05, Cbrev-02/08, Cbrev-03, Cbrev-04, Cbrev-06, Cbrev-07, Cbrev-09) (S4 Table).



The mtDNA COI haplotype of the individual from Grafton [GenBank: KJ162974] is likely to be Cbrev-01 or Cbrev-05 (due to the presence of a C in position 646, which it shares with Cbrev-01 and Cbrev-05, but not the other Australian sequences; [S4 Table](#)). To differentiate between Cbrev-01 and Cbrev-05 would require a C or a T at position 1,144, but this position was unfortunately not sequenced. Both Cbrev-01 and Cbrev-05 have been found at Grafton, so KJ162974 is most likely Cbrev-01 or Cbrev-05. The sequence from Grafton [KJ162974] was also identical over a shared 646–655bp region to those from the Solomon Islands [GenBank: KJ162971, KJ162970, KJ162969, KJ162967] and Timor-Leste [GenBank: KJ162968].

### MtDNA COI haplotype network

Haplotype networks were drawn using TCS version 1.21 [\[62\]](#) for the 835bp mtDNA COI haplotypes of *C. brevitarsis*, and for the 547bp mtDNA COI haplotypes of *C. marksii*. The resulting networks were subsequently refined by hand.

### Analysis of Australian *C. brevitarsis* population structure based on partial mtDNA COI gene sequence and EPIC-microsatellite DNA markers

**(I) Inference of population structure based on mtDNA COI partial gene.** Using 835bp of the partial mtDNA COI gene, we estimated genetic differentiation among and between *C. brevitarsis* populations across northern and eastern Australia using Weir and Cockerham's [\[63\]](#) single locus  $F$ -statistics, as implemented in GENEPOP version 4.2 [\[64, 65\]](#). Estimates of diversity indices (haplotype diversity ( $h$ ) and nucleotide sequence diversity ( $\pi$ )) in *C. brevitarsis* (based on 835bp) and *C. marksii* (based on 547bp), and the Analysis of Molecular Variance (AMOVA) [\[66\]](#) for *C. brevitarsis* was performed using Arlequin version 3.11 [\[67\]](#). To identify whether the Australian *C. brevitarsis* population could be subdivided into smaller groups, we performed an AMOVA whilst considering the following entities: (i) a single population (i.e., where all the Australian specimens of *C. brevitarsis* belong to a single population); (ii) a northern population (comprising specimens from Kalumburu, Douglas-Daly, and Katherine) and an eastern population (comprising specimens from Lismore, Grafton, Bellingen, and Paterson), and (iii) a large number of site-specific populations (i.e., where each collection site has its own population). The reason for assuming the northern and eastern populations was based on the general close proximity of collection sites to each other and did not reflect the midges' flight capability, which is usually dependent on environmental factors (e.g., weather conditions/wind patterns [\[29, 68\]](#)). We also calculated the inter-population  $F_{ST}$  between the seven *C. brevitarsis* populations using GENEPOP (Option 6) to determine whether geographic distance between locations may have contributed to an underlying population structure; to investigate this, we plotted geographic distances (km, estimated from GPS data) against the  $F_{ST}/(1 - F_{ST})$  values.

**(II) Inference of *C. brevitarsis* population structure based on nuclear markers.** Using Arlequin, we also obtained results from an AMOVA based on the EPIC (RpS4, RpS2A, RpS2B, RpL27A, RpS6, and RpS8) and microsatellite (Cimi-69, and Cimi-85) markers. We excluded RpS4 and RpS2A and the two microsatellite loci from our analyses because: RpS4 and RpS2A were monomorphic; Cimi-66 was found to potentially contain a 5-bp (TTGCA) target-site duplication sequences surrounding the (CA)<sub>n</sub> SSR units; and Cimi-12 amplified inconsistently with >3 alleles detected in various *C. brevitarsis* individuals (data not shown). Estimates of nucleotide sequence diversity ( $\pi$ ), haplotype (gene) diversity ( $h$ ),  $F_{IS}$  and their associated  $P$  values, pair-wise  $F_{ST}$  values between populations, average number of alleles ( $a$ ), allele richness ( $r$ ), and observed and expected heterozygosity estimates ( $H_O$ ,  $H_E$ ) across all nuclear markers for each population were also obtained using Arlequin, and departures from Hardy-Weinberg equilibrium were calculated using GENEPOP. Pair-wise  $F_{ST}$  estimates between populations

with significant  $P$ -values were adjusted using the Holm-Bonferroni correction [69] at the family-wise error rate of  $\alpha = 0.05$ . We determined the presence of isolation by distance for our *C. brevitarsis* populations using the transformed (i.e.,  $F_{ST}/(1-F_{ST})$ ) pair-wise population  $F_{ST}$  estimates (i.e., Slatkin's [70] linearized  $F_{ST}$ ), plotted against natural logarithm of geographic distances between locations of individual populations, with the significance of observed associations determined using Spearman Rank Correlation coefficient with 5,000 permutations, as implemented in GENEPOP.

## Structure analysis

We used the Bayesian MCMC simulation program STRUCTURE [71] to determine the most likely number of ancestral gene clusters based on the current population nuclear marker (four EPIC and two microsatellite) genetic data for *C. brevitarsis*. The rationale and parameters used to calculate the *ad hoc* statistic  $\Delta K$  values (this provides an indication of the second-order rate-of-change to the estimated log probability of data  $\Pr(X|K)$  between increasingly large values of  $K$  (i.e., number of genetic clusters)) were as described in Behere *et al.* [19]. Once the best  $\Delta K$  values were found, a final simulation (comprising 500,000 MCMC samples and a burn-in of 50,000) was conducted to obtain the most likely ancestral population structure.

## Phylogenetic analysis

A phylogenetic tree was inferred using a multiple sequence alignment (MSA) of 182 codons from the mtDNA-encoded COI genes of the 136 *Culicoides* sequences (Table 1). Orthologous sequences from *Anopheles quadriannulatus* [GenBank: DQ792581], *A. gambiae* [GenBank: Nc\_002084], *A. dirus* D [GenBank: AJ877572], and *A. janconnae* [GenBank: HQ335348] were aligned to these 136 sequences, allowing us to root of the inferred tree of *Culicoides*. The MSA (see S1 and S2 Files) was inferred using SeaView version 4.4.2 [72].

Initially, the MSA of codons was translated into amino acids and then partitioned into MSAs of codon sites. Whenever sequences in these four MSAs were found to be identical to one another, all but one of these sequences were removed to reduce bias in the phylogenetic estimates.

Most molecular phylogenetic methods assume that the sequences have evolved under globally stationary, reversible, and homogeneous (SRH) conditions (defined in [73–77]). To determine whether these conditions were met by the data, we assessed each alignment using a matched-pairs test of symmetry [73, 78]. In practice, we used Homo <[www.csiro.au/homo](http://www.csiro.au/homo)>, a program that returns a  $\chi^2$ -distributed test statistic for all pairs of sequences. The corresponding  $P$  values were then charted in a PP-plot against the expected  $P$  values. A J-shaped distribution of dots in this plot, with a larger-than-expected proportion of low  $P$  values, indicates that the assumption of evolution under globally SRH conditions is violated by the data.

Given the results of our assessment of model misspecification (described above) and other considerations (described below), we conducted our phylogenetic analysis on the MSA of first and second codon sites. In practice, we used IQTREE version 0.9.3 [79] to find the most likely (ML) tree. In so doing, we first identified the best model of evolution for the data using the-m TESTONLY option. Next, we inferred the most likely tree and completed the phylogenetic analysis with a nonparametric bootstrap analysis (using the-bb 1000 option). The result was displayed using FigTree version 1.40 <<http://tree.bio.ed.ac.uk/software/figtree>>.

## Ethics

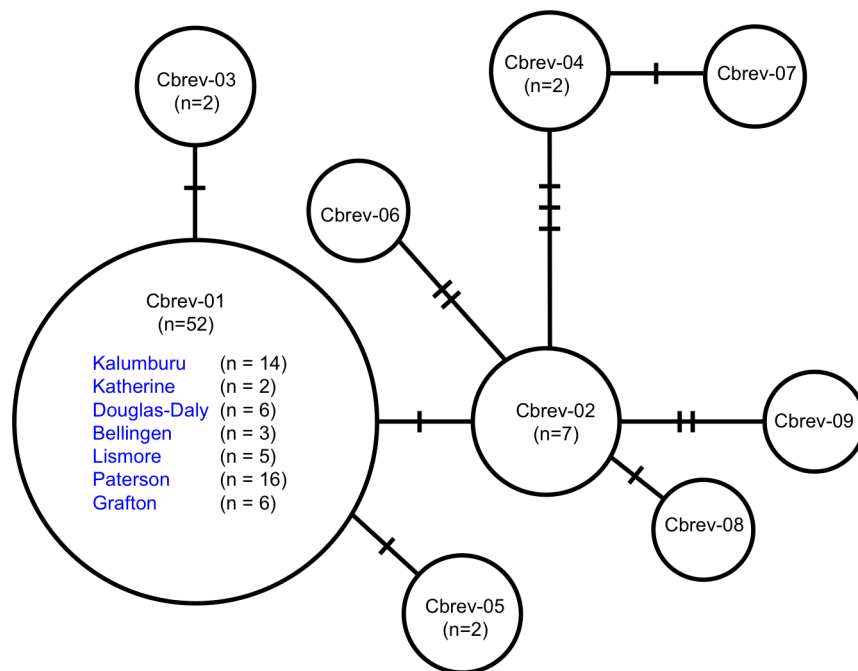
The authors declare that the research does not involve human subjects, human material or human data. The study also does not involve vertebrates or any regulated invertebrates, and therefore does not require any ethics committees' approval.

## Results

### Analyses of the mtDNA marker from Australian *C. brevitarsis* and *C. marksii*

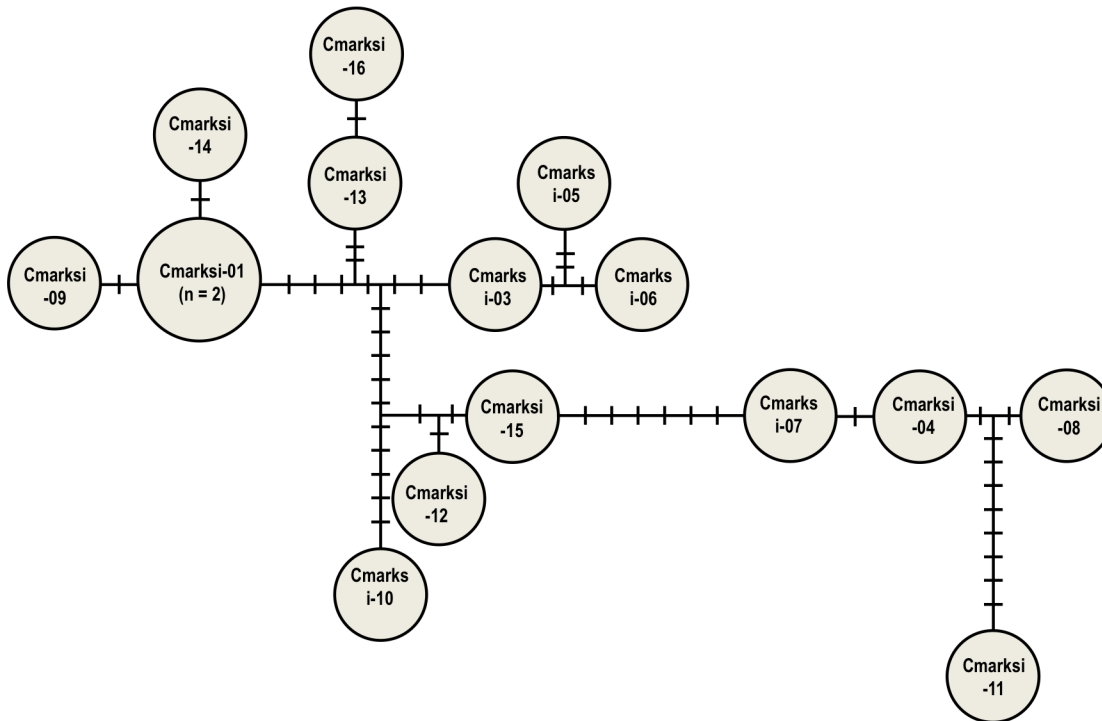
A total of nine haplotypes was found in our alignment of 70 partial sequences (835bp) of the mtDNA COI gene from *C. brevitarsis*. The most common haplotype was Cbrev-01 (n = 52), followed by Cbrev-02 (n = 7), Cbrev-03 (n = 3), Cbrev-04 (n = 2), and Cbrev-05 (n = 2), while Cbrev-06, Cbrev-07, Cbrev-08, and Cbrev-09 were found once each (Table 1, Fig 1). Cbrev-01 was found at all sampling sites (Table 1, Figs 1 and 2) while Cbrev-03, Cbrev-04 and Cbrev-06 were found only at our sites in WA (Kalumburu) and NT (Douglas-Daly, and Katherine). The other haplotypes (Cbrev-02 (three sites), Cbrev-05 (two sites), Cbrev-07, Cbrev-08, and Cbrev-09 (one site each)) were found at our sites in NSW (Lismore, Grafton, Bellingen, and Paterson). Differences at six sites in the DNA separated the most diverged haplotypes (e.g., Cbrev-07 vs. Cbrev-03; Cbrev-07 vs. Cbrev-09; Cbrev-07 vs. Cbrev-06; Cbrev-07 vs. Cbrev-05; Fig 2). Estimates of the evolutionary distance between our Australian samples of *C. brevitarsis* were low, ranging from 0 to 0.7% (S1 Table). The corresponding estimates between the Australian and Japanese samples of *C. brevitarsis* [GenBank: AB360994.1, AB360995.1] [61] were, surprisingly, one order of magnitude higher (8.2% to 8.9%, S1 Table).

A total of 15 haplotypes was found in our alignment of 16 partial sequences (547bp) of the mtDNA marker from *C. marksii* (all from Murray Bridge in SA and Moree in NSW. Two of the



**Fig 2. *Culicoides brevitarsis* mtDNA haplotype network based on 835 bp of mtDNA COI gene sequence.** 2 *Culicoides brevitarsis* mtDNA COI haplotype network (835bp) from seven sites across Australia (Kalumburu, WA; Douglas-Daly and Katherine, NT; Lismore, Grafton, Bellingen, and Paterson, NSW). The total number of *C. brevitarsis* surveyed was n = 70. The most common haplotype is Cbrev-01 (it includes n = 52 individuals from all seven sites), followed by Cbrev-02 (n = 7) from Lismore, Grafton, and Paterson. Haplotypes Cbrev-03, Cbrev-04, and Cbrev-05 are represented by two individuals each; all remaining haplotypes are represented by one individual each. The haplotype network was generated using TCS version 1.21 [62], and then refined manually. The numbers of substitutions separating haplotypes are indicated by short bars.

doi:10.1371/journal.pone.0146699.g002



**Fig 3. *Culicoides marksii* mtDNA haplotype network based on 547 bp of mtDNA COI partial gene sequence.** *Culicoides marksii* mtDNA COI haplotype network (547bp) from Moree (NSW) and Murray Bridge (SA). The total number of *C. marksii* samples surveyed was 16. Haplotypes Cmarks-i-01, and Cmarks-i-03, . . . , Cmarks-i-09 are from Moree, while Cmarks-i-10, . . . , Cmarks-i-16 are from Murray Bridge. The haplotype network was obtained using TCS version 1.21, and then refined manually. Apart from Cmarks-i-01, which was represented by two individuals, the other haplotypes were represented by one individual each. The numbers of substitutions separating haplotypes are indicated by short bars.

doi:10.1371/journal.pone.0146699.g003

samples from Moree were identical; Figs 1 and 3), indicating a high level of genetic diversity among these samples. The evolutionary divergence between pairs of these sequences ranged from 0.0 to 2.8% (S2 Table), which is up to four times greater than those between the Australian *C. brevitarsis*.

The average genetic distances between all *Culicoides* species sampled in Australia and those available from GenBank (S5 Table) indicate high levels of genetic diversity between the species, with an average genetic distance of  $0.153 \pm 0.009$  SE (range from  $0.110 \pm 0.014$  SE (*C. pallidothorax* vs. *C. henryi*) to  $0.264 \pm 0.020$  SE (*C. humeralis* vs. *C. actoni*)). When third codon sites were excluded (reason given below), the genetic distances between the species were much lower (average =  $0.059 \pm 0.007$  SE; range from  $0.007 \pm 0.004$  SE, (*C. kibunensis* vs. *C. verbosus*) to  $0.172 \pm 0.021$  SE (*C. actoni* vs. *C. jacobsoni*)). The average genetic distances within species with at least two haplotypes ranged from 0.0 (e.g., in *C. wadai*) to 0.065 (in *C. matsuzawai*) across the trimmed 547bp region. Differentiation of *Culicoides* species based purely on mtDNA COI sequence divergence may be impossible (for a discussion, see Meier *et al.* [80]) due to over-lapping intra- and inter-specific genetic distances (i.e., the lower range between species is 0.096 (*C. pallidothorax* vs. *C. henryi*) while the maximum intra-specific genetic distance is 0.082 (e.g., *C. brevitarsis*) (S5 Table), a difference of only 1.4%).

Estimates of haplotype diversity ( $h$ ) and nucleotide sequence diversity ( $\pi$ ) were both higher for the Australian endemic *C. marksii* ( $h = 0.976 \pm 0.012$ ;  $\pi = 0.017 \pm 0.001$ ) than for the Australian *C. brevitarsis* ( $h = 0.586 \pm 0.172$ ;  $\pi = 0.001 \pm 0.001$ ). We did not detect any isolation-by-distance effect between the Australian populations of *C. brevitarsis* based on the mitochondrial DNA COI partial gene marker, with these populations lacking significant substructure as

**Table 3. AMOVA results of Australian *Culicoides brevitarsis* populations based on mtDNA and nuDNA markers.**

Hierarchical levels	Variance	Percentage of variation	Fixation indices	P-value
<b>(a)</b>				
Among groups ( $F_{CT}$ )	0.0001	0.03	0.0003	0.431±0.014
	(0.004)	(1.78)	(0.018)	(0.305±0.015)
Among populations within groups ( $F_{SC}$ )	0.023	5.18	0.052	0.064 ± 0.009
	(0.004)	(1.64)	(0.017)	(0.297±0.014)
Within populations ( $F_{IT}$ )	0.415	(94.79)	0.0518	0.107 ± 0.009
	(0.227)	(96.58)	(0.034)	(0.225±0.014)
<b>(b)</b>				
Among groups ( $F_{CT}$ )	0.003	0.46	0.005	0.235±0.014
Among populations within groups ( $F_{SC}$ )	0.014	2.00	0.020	<b>0.047±0.008</b>
Among individuals within populations ( $F_{IS}$ )	0.030	4.21	0.043	0.172±0.010
Within individuals ( $F_{IT}$ )	0.665	93.32	0.067	0.088±0.009

Results from Analysis of Molecular Variance (AMOVA) of *Culicoides brevitarsis* populations in Australia based on (a) mtDNA COI partial gene sequence 829bp (502bp in parentheses for comparisons with *C. marksii*), and (b) four EPIC and two microsatellite DNA loci.

**Note:** Fixation indices based on mtDNA COI partial sequence for: among groups ( $F_{CT}$ ) are northern Australian (Kalumburu, Douglas-Daly, and Katherine) and eastern Australian sites (Lismore, Grafton, Bellingen, and Paterson); among populations within groups ( $F_{SC}$ ), and within populations ( $F_{ST}$ ). All *P* values were non-significant. Fixation indices estimated from EPIC-PCR and microsatellite DNA markers were also non-significant for among groups ( $F_{CT}$ ), for among individuals within populations ( $F_{IS}$ , i.e., non-random mating among individuals within subpopulations), for  $F_{IT}$  (i.e., average reduction in heterozygosity of an individual relative to the total population, and represents contributions due to non-random mating within demes (subpopulations) and effects of random genetic drift among demes ( $F_{ST}$ )), but significant for among populations within groups ( $F_{SC}$ ; *P*-value = 0.047 ± 0.008), indicating substantial variance existed at this population structure hierarchical level.

doi:10.1371/journal.pone.0146699.t003

indicated by a low  $F_{ST}$  value ( $F_{ST} = 0.026$ ), and non-significant correlation between geographic distances and  $F_{ST}/(1 - F_{ST})$  values ( $R^2 = 0.0016$ ), possibly due to the wide geographical distribution of the Cbrev-01 haplotype (Fig 1). Hierarchical population structures were not found using AMOVA-estimated fixation indices:  $F_{CT}$  (among groups),  $F_{SC}$  (among populations within groups) and  $F_{ST}$  (within populations) (none of the associated *P*-values were significant, Table 3). The AMOVA showed that 96.6% of the variance occurs at the within-populations level, while 1.8% and 1.6% of the variance were at the among-groups level (i.e., within Australia) and among-populations within-groups level (i.e., between northern and eastern populations), respectively.

### Analyses of the nuDNA markers from Australian *C. brevitarsis*

Two exon-primed intron-crossing (EPIC) DNA markers (i.e., RpS4 and RpS2A) were found to be monomorphic across all the populations of *C. brevitarsis* that we screened. We sequenced the amplicons of RpS4 and RpS2A from individuals of *C. brevitarsis*, confirming the primer annealing sites and the partial coding region sequences for RpS4 and RpS2A. Despite the expected presence of an intron within the sequences examined when compared against the *Bombyx mori* homologs, the expected introns in these two Rp gene regions were missing (sequences available on request), providing the reason for the lack of polymorphisms for these two EPIC DNA markers. This, therefore, leaves the availability of four EPIC markers (i.e., RpS2B, RpS6, RpS8, RpL27A) as well as two microsatellite DNA markers (i.e., Cimi69, Cimi85) to infer the population genetic structure of *C. brevitarsis*.

The results obtained from the AMOVA of the six nuDNA markers from *C. brevitarsis* were largely in agreement with those obtained from the mtDNA marker, with the exception that the

fixation index for the among populations within groups comparisons ( $F_{SC} = 0.020$ ) was significant at both the mean and lower estimates, but not at the higher estimate ( $P$  value =  $0.047 \pm 0.008$ ). The availability of results from analyses of nuDNA markers enabled a fine-scale investigation of within population variation, which showed that 93.3% of the variation was at the within individual level relative to the total population ( $F_{IT}$ ) level (Table 3). Pair-wise  $F_{ST}$  estimates between all populations obtained using the mtDNA marker did not reveal any population substructure (Table 4) due to the presence of the Cbrev-01 haplotype in all populations. By contrast, the nuDNA markers disclosed significant signatures of limited gene flow between the northern region populations (e.g., Kalumburu B, Douglas-Daly) and four of the southern region populations (Table 5; six out of 28 comparisons). However, following a Holm-Bonferroni correction, only two pair-wise  $F_{ST}$  estimates were significant (i.e., Douglas-Daly vs. Lismore; Douglas-Daly vs. Paterson; Table 5). Based on the  $F_{IS}$  estimates (Table 6), no significant inbreeding at the within subpopulation level could be detected.

Significant isolation-by-distance was detected using the nuDNA markers ( $R^2 = 0.192$ ,  $P$  value = 0.0152), indicating that neighbouring populations tended to be genetically more similar to one another than they are to distant populations. Overall, the six nuDNA markers appeared to assort independently, and using Fisher’s exact test did not disclose any departure from Hardy-Weinberg equilibrium, the only exceptions being Cimi-85 for Kalumburu and Douglas-Daly, and RpS6 and RpL27A for Paterson (due to excess of homozygotes; results not shown).

### STRUCTURE analysis

Using the method of Evanno *et al.* [81] to calculate the number of genetically homogenous groups of individuals in a population, we found that the nuDNA markers data (Fig 4) was best explained as two genetically homogeneous sets of individuals ( $\Delta K = 8.81$ ). We did not apply the mtDNA data to the STRUCTURE analysis as this method is best applied to sequence data with sufficient recombination rates (i.e., its application based on non-recombination genomes such as the sex chromosomes (i.e., Y-chromosome) and genomes with predominantly uni-parental inheritance mode (i.e., chloroplast or mitochondrial DNA) should be interpreted with caution [71]). The STRUCTURE result suggests the Australian *C. brevitarsis* population structure was best explained by unequal contributions from two genetically distinct ancestral clusters, with the ‘Red’ and ‘Green’ genetic clusters following latitudinal clines from the north-western/northern populations to the south-eastern/southern populations, although increasing sampling sites and marker numbers will be needed to better explore this genetic signature in

**Table 4. Australian *Culicoides brevitarsis* population mtDNA COI pairwise  $F_{ST}$  estimates and associated  $P$  values  $\pm$  SE.**

Populations	1	2	3	4	5	6	7
1. Kalumburu	-	0.999±0.001	0.538±0.015	0.999±0.001	0.121±0.010	0.999±0.001	0.429±0.012
2. Douglas-Daly	-0.059	-	0.484±0.014	0.999±0.001	0.101±0.010	0.999±0.001	0.758±0.013
3. Katherine	-0.019	0.018	-	0.570±0.017	0.534±0.015	0.999±0.001	0.090±0.009
4. Lismore	-0.055	-0.044	-0.028	-	0.428±0.017	0.999±0.001	0.760±0.012
5. Grafton	0.092	0.122	0.008	-0.021	-	0.450±0.014	0.077±0.009
6. Bellingen	-0.088	-0.067	-0.067	-0.105	0.027	-	0.363±0.016
7. Paterson	-0.018	-0.040	0.163	-0.041	0.117	-0.020	-

Note: Australian *Culicoides brevitarsis* population mtDNA COI pairwise  $F_{ST}$  estimates (lower triangle) and associated  $P$  values  $\pm$  SE (upper triangle) based on 829bp of mtDNA COI partial gene sequence. Intermediate level of population substructure is detected between Grafton and Douglas-Daly ( $F_{ST} = 0.122$ ) and between Grafton and Paterson ( $F_{ST} = 0.117$ ) but with non-significant  $P$  values and SD from 500 bootstrap replications.

doi:10.1371/journal.pone.0146699.t004

**Table 5. Australian *Culicoides brevitarsis* population pairwise  $F_{ST}$  estimates and associated  $P$ -values inferred from nuDNA markers.**

	1	2	3	4	5	6	7	8
1 Kalumburu A	-	0.261 ± 0.044	0.054 ± 0.020	0.072 ± 0.018	0.541 ± 0.054	0.676 ± 0.046	0.496 ± 0.039	0.937 ± 0.028
2 Kalumburu B	0.027	-	0.450 ± 0.073	0.532 ± 0.032	0.721 ± 0.045	0.036 ± 0.015	0.198 ± 0.033	0.252 ± 0.049
3 Katherine	0.057	0.015	-	0.009 ± 0.009	0.369 ± 0.042	0.306 ± 0.015	0.505 ± 0.039	0.009 ± 0.009
4 Douglas-Daly	0.092	0.007	0.157	-	0.045 ± 0.020	<b>0.000 ± 0.000</b>	<b>0.000 ± 0.000</b>	0.054 ± 0.020
5 Bellingen	-0.008	-0.011	0.002	0.075	-	0.369 ± 0.055	0.496 ± 0.051	0.387 ± 0.078
6 Lismore	-0.015	0.087	0.018	<b>0.216</b>	0.005	-	0.820 ± 0.037	0.180 ± 0.043
7 Paterson	-0.003	0.027	0.001	<b>0.119</b>	-0.005	-0.015	-	0.216 ± 0.053
8 Grafton	-0.029	0.016	0.076	0.056	0.000	0.025	0.010	-

Pairwise  $F_{ST}$  estimates (lower triangle) and corresponding  $P$  values (upper triangle) between *Culicoides brevitarsis* populations as inferred based on genotype data from four EPIC and two microsatellite DNA markers. Significant  $F_{ST}$   $P$  values ( $< 0.05$ ) are in red based on 110 permutations as calculated using Arlequine V3.11. Kalumburu A and Kalumburu B are two Western Australian populations collected from the same sites but from different years ('A' 29/12/08; 'B' 25/6/09). Katherine and Douglas-Daly are from the Northern Territory, and Bellingen, Lismore, Paterson and Grafton are New South Wales populations.

**Note:** Only  $F_{ST}$  estimates for Douglas-Daly/Lismore and Douglas-Daly/Paterson remained significant following a Holm-Bonferroni correction at the  $\alpha = 0.05$  level (bolded). Treating Kalumburu A & B as a single population (due to the insignificant  $F_{ST}$  value between these two samples) indicated the  $F_{ST}$  estimate between Kalumburu and Lismore was not significant ( $F_{ST} = 0.029$ ,  $P = 0.135 \pm 0.023$ ).

doi:10.1371/journal.pone.0146699.t005

future studies. This analysis therefore shows that although the northern and southern *C. brevitarsis* populations are the same species and that gene flow has occurred between them, there is also evidence that they represent two overall genetic groups of individuals (exemplified by differences between populations in the north (e.g., Kalumburu) and populations in the south (e.g., Paterson)). This may have implications for the capacity of the BTV to spread (see [Discussion](#)).

### Phylogenetic analysis

[Fig 5](#) shows four PP-plots with results from the matched-pairs test of symmetry for the four MSAs of amino acids, first codon sites, second codon sites, and third codon sites. In terms of being suitable for phylogenetic analysis using time-reversible Markov models, the data sets are ranked (from best to worst): amino acids, second codon sites, first codon sites, and third codon

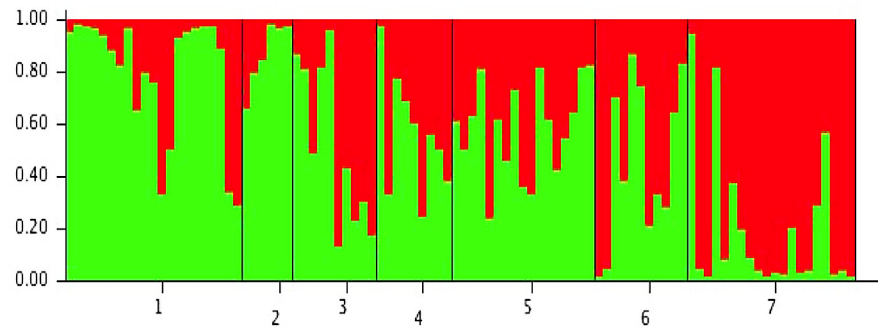
**Table 6. Nuclear DNA marker statistics for Australian *C. brevitarsis* populations.**

Locations	n	a	r	$H_o$	$H_e$	$F_{IS}$	$P$
Kalumburu A	8	3.167 ± 1.344	0.443 ± 0.289	0.521 ± 0.357	0.488 ± 0.305	0.400	0.098
Kalumburu B	8	3.000 ± 1.291	0.638 ± 0.473	0.578 ± 0.320	0.526 ± 0.277	0.228	0.094
Douglas-Daly	6	2.833 ± 1.344	0.568 ± 0.377	0.461 ± 0.360	0.494 ± 0.351	-0.059	0.779
Katherine	4	3.333 ± 2.055	0.801 ± 0.557	0.602 ± 0.308	0.572 ± 0.278	0.103	0.257
Bellingen	3	3.667 ± 2.494	0.732 ± 0.514	0.535 ± 0.342	0.611 ± 0.173	-0.001	0.556
Lismore	6	2.500 ± 1.893	0.000 ± 0.000	0.563 ± 0.419	0.599 ± 0.431	0.034	0.433
Paterson	20	4.500 ± 2.063	0.659 ± 0.422	0.582 ± 0.335	0.591 ± 0.136	-0.063	0.525
Grafton	12	2.000 ± 2.828	0.669 ± 0.473	0.217 ± 0.313	0.223 ± 0.319	0.022	0.562

Australian *Culicoides brevitarsis* population statistics based on four EPIC-PCR and two microsatellite DNA loci.

**Note:** n = number of individuals analysed; a = average number of alleles ( $\pm$  s.d.), r = allele richness (average gene diversity over loci),  $H_o$  = observed heterozygosity,  $H_e$  = expected heterozygosity,  $P$  (Random  $F_{IS} \geq$  observed  $F_{IS}$ ). Markers were overall in Hardy Weinberg equilibrium for all populations with non-significant  $P$  values. All  $F_{IS}$  estimates for all populations were also not significant, although Kalumburu (A & B)  $F_{IS}$  estimates were high, but low ( $F_{IS}$ : 0.087,  $P = 0.252$ ) when both populations combined as one.

doi:10.1371/journal.pone.0146699.t006



**Fig 4. *Culicoides brevitarsis* population structure as inferred from nuDNA (EPIC-PCR and microsatellite DNA) markers using STRUCTURE 2.3.** *Culicoides brevitarsis* population structure as inferred using the Bayesian clustering algorithm implemented in the program STRUCTURE 2.3 [71], based on nuclear DNA (four EPIC and two microsatellite) markers. Bar graphs are from  $K = 2$  from 50,000 burn-in cycles and 500,000 runs. Populations are numbered from northern to southern regions, and combining (Kalumburu A and B (WA)) as one population (1), Douglas-Daly (2) and Katherine (3) from NT, and Lismore (4), Grafton (5), Bellingen (6), and Paterson (7) from NSW.

doi:10.1371/journal.pone.0146699.g004

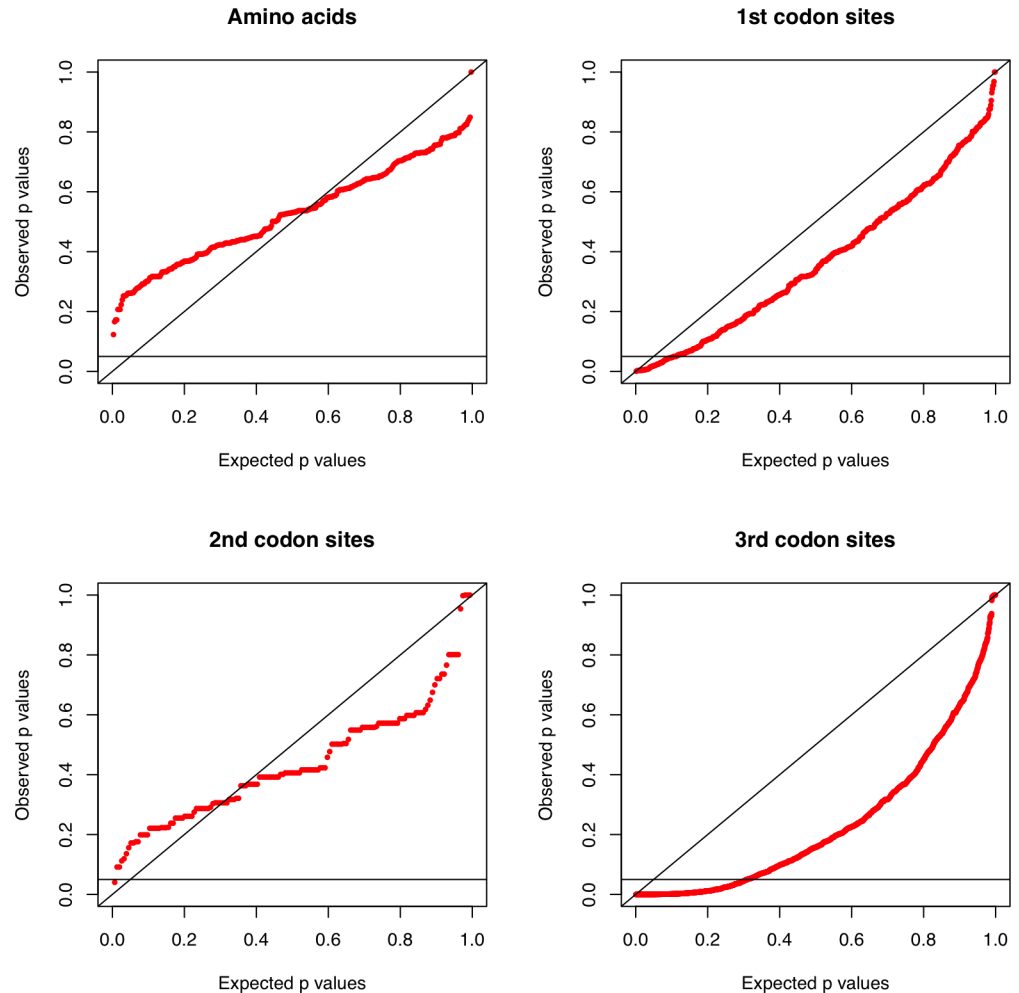
sites. Thirty-one per cent of the tests led to  $P$  values below 0.05 for third codon sites (5% was expected), and the lowest observed  $P$  value was  $1.33 \times 10^{-9}$ , which still is significant after a Holm-Bonferroni correction [69] to counteract the effect of multiple comparisons ( $P_{\text{cor}} = 2.76 \times 10^{-6}$ ). For first codon sites, the corresponding values were 10.7% and  $P_{\text{cor}} = 0.5395$ . These results show that it would be unwise to use third codon sites in a phylogenetic study that assumes time-reversibility.

The numbers of phylogenetically informative sites were 44, 49, and 13 for the MSAs of amino acids, first codon sites, and second codon sites, respectively, implying that it would be preferable to use MSAs of first and second codon sites to infer the phylogenetic tree. In order to increase the number of phylogenetically informative sites in the MSA used to infer the phylogenetic tree, we concatenated the MSAs for these two codon sites.

Using IQTREE, we found the best model of evolution to be TN+I+G4, implying that the accumulating nucleotide substitutions was best approximated by Tamura and Nei's [57] model of nucleotide substitution, with rate-heterogeneity across sites modelled assuming a proportion of invariable sites (I) and a discrete  $\Gamma$  distribution with 4 rate categories (G4). However, in practise the phylogenetic analyses were conducted under the TN+I+G4 and TN+G4 models to accommodate a known difficulty in estimating both I and  $\Gamma$  accurately (see page 114 in [82]).

Fig 6 presents the ML tree inferred from the concatenated alignment of first and second codon sites. The tree divides the Australian samples into two distantly related groups. The first group includes the Australian *C. brevitarsis*, which consistently (97%) formed a sister group to the Japanese *C. brevitarsis*. The second group includes *C. marksi*, *C. henryi*, *C. pallidothorax*, *C. bundyensis*, and *C. bunrooensis*, which are endemic Australian species. However, this group may also include two Japanese species (*C. verbosus* and *C. kibunensis*). Whether this is the case was impossible to determine from the data (the low bootstrap values for some of the edges and the presence of trifurcations in the tree indicate that the number of informative sites is too low to cast light on this issue). Notwithstanding this shortness of data, it is worth mentioning the evolutionary distance between Australian and Japanese samples of *C. brevitarsis* is as large as those between some of the other Australian endemic species (i.e., *C. marksi*, *C. henryi*, and *C. pallidothorax*). Looking at the phylogenetic tree more broadly, it is characterised by many low bootstrap values and large variation in the root-to-tip distances. Taken together, these features are consistent with the number of phylogenetically informative sites being too small, implying



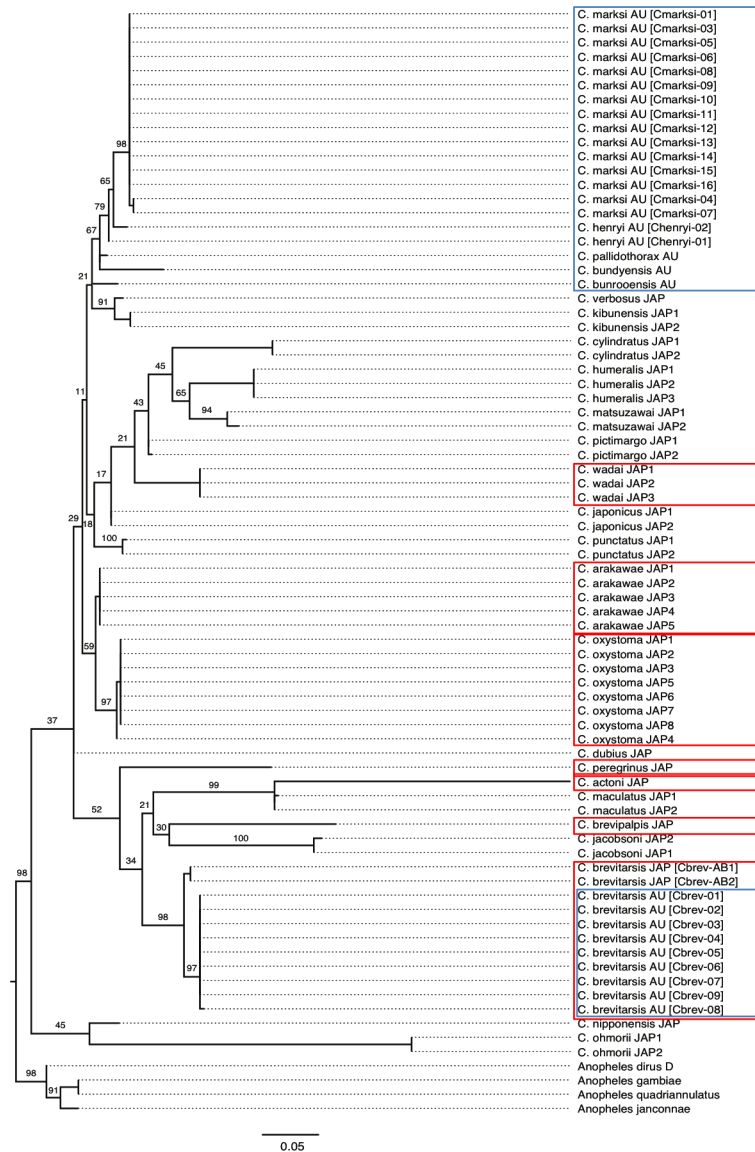


**Fig 5. PP-plots for  $P$  values distributions from matched-pairs tests of symmetry for mtDNA COI sequence data sets.** PP-plots with the distributions of  $P$  values from the matched-pairs tests of symmetry for the following data sets: amino acids, first codon sites, second codon sites, and third codon sites. A J-shaped distribution of dots, with more than 5% of the dots below 0.05 (i.e., the horizontal line) indicates violation of the assumption of evolution under globally SRH conditions. The PP-plot for third codon sites represents a clear example of violation of this phylogenetic assumption.

doi:10.1371/journal.pone.0146699.g005

that much longer sequences are needed to accurately resolve the phylogenetic relationships among these species.

With respect to the phylogenetic relationship among competent and potential vectors of BTV in Australia, our tree suggests that the five endemic species considered here (i.e., *C. marksi*, *C. henryi*, *C. pallidothorax*, *C. bundyensis*, and *C. bunroensis*) do not share a close common ancestor with BTV-competent vectors (i.e., *C. brevitarsis*, *C. wadai*, *C. fulvus*, and *C. actoni*) (note that we did not have access to Australian samples of *C. fulvus*, *C. wadai*, and *C. actoni* and, therefore, were unable to determine whether Japanese samples of *C. wadai* and *C. actoni* are the closest relatives of, respectively, the Australian samples of *C. wadai* and *C. actoni*). Although this does not guarantee that these endemic species cannot become vectors of BTV, it does imply that they are so evolutionarily different from known vectors of BTV in Australia, that they might not be potential vectors of BTV. To better address this issue, it will be



**Fig 6. Maximum Likelihood phylogenetic tree including bootstrap values of various *Culicoides* species from this study.** Maximum Likelihood tree, with bootstrap values of some of the internal edges inferred using the UFBoot method. The tree was inferred under the TN+I+G4 model of evolution. Samples are marked using their species name, geographical origin (AU: Australia; JAP: Japan), haplotype name (when available, listed in brackets), and a number (whenever needed, to distinguish some of the sequences). Australian *Culicoides* are within blue boxes, and known BTV competent Asian/South East Asian *Culicoides* hosts are in red boxes. The scale bar measures evolutionary distance in terms of average nucleotide substitutions per site. The tree was rooted using four species of *Anopheles*. The ML tree inferred under the TN+G4 model of evolution differed from this one at seven internal edges, but the overall conclusion remained the same (see text).

doi:10.1371/journal.pone.0146699.g006

necessary to consider much longer sequences (e.g., preferentially the whole mitochondrial genome; multi-loci sequence analysis) and more samples and species from, in particular, Australia.

## Origin(s) and hypothesised incursion routes of Australian *C. brevitarsis*

Our study also considered previously reported samples that were collected in limited numbers from neighbouring countries such as Solomon Islands ( $n = 4$ ; GenBank: KJ162971, KJ162970, KJ162969, KJ162967), Timor-Leste ( $n = 2$ ; GenBank: KJ162975, KJ162968), China ( $n = 3$ ; GenBank: KJ162966, KJ162972, KJ162973), and Japan ( $n = 2$ ; GenBank: AB360994, AB360995). In addition, we considered an Australian individual sampled from Grafton (GenBank: 162974), a site also used by us.

A comparison of Australian, Japanese, Chinese, Solomon Islands, and Timor-Leste *C. brevitarsis* mtDNA COI sequences, available through the studies of [61] and [51], showed that the Japanese samples, and to a lesser extent, the Chinese samples, are distantly related to the Australian, Solomon Islands, and Timor-Leste samples (S1–S3, S5 Tables). On the other hand, one of the two haplotypes from Timor-Leste [KJ162968] was identical across at least 646bp to an Australian [KJ162974] and all four Solomon Islands individuals [51]. This haplotype [KJ162968] was also identical to our Cbrev-01 and Cbrev-05 haplotypes between positions 324 and 646, and is thus consistent with the expected introduction scenario for this haplotype (i.e., Cbrev-01), with an incursion route from Timor-Leste to the northwest of Australia, based on the prevailing weather/wind patterns. In order to differentiate the most common haplotype previously found [51], the authors would need to have sequenced further along the COI gene, as was done by us for Cbrev-01 and Cbrev-05, which identified a single base pair difference at position 1144 (S4 Table).

## Discussion

### *Culicoides brevitarsis* mtDNA COI haplotypes

Our analysis of the mtDNA marker show that the *C. brevitarsis* samples from northern and southern Australian sites (i.e., the full geographical range of *C. brevitarsis* in Australia) is a single species with low nucleotide sequence and haplotype diversities across >800bp of this genome. There is no support for the notion that *C. brevitarsis* in Australia is a set of cryptic species. *Culicoides brevitarsis* has been recorded in Australia for over 100 years and is believed to be an introduced species [83] of Asian or South East Asian origin [39]. The spread of *C. brevitarsis* to its current southern limit is likely to be linked to European settlements, and its presence in northern Australia may also be associated with the introduction of ruminants [84] such as water buffalo (*Bubalus bubalis*) since around 1825 <<https://www.environment.gov.au/biodiversity/invasive-species/publications/factsheet-feral-water-buffalo-bubalus-bubalis>>. The inclusion of additional samples from South East Asia (e.g., Papua New Guinea, Timor-Leste, Indonesia, Malaysia, the Malay Archipelago), Micronesia, Melanesia, and Asia (e.g., India) would help to clarify what are the most likely origins of the Australian *C. brevitarsis* populations. However, the difficulty in obtaining good samples from these regions probably represents the single most challenging factor to future similar studies.

A total of nine mtDNA haplotypes were found among 70 *C. brevitarsis* individuals from three north Australian sampling sites (Kalumburu, Katherine, and Douglas-Daly) and four east Australian sampling sites (Lismore, Grafton, Bellingen, and Paterson) (Fig 1). With the exception of Cbrev-01, a haplotype found at all seven sampling sites (spanning distances of >3,000km), the remaining haplotypes could be split into northern haplotypes (Cbrev-03, Cbrev-04, and Cbrev-06) and eastern haplotypes (Cbrev-02, Cbrev-05, Cbrev-08, and Cbrev-09) (Fig 1). The detection of such distinct distributional differences may have

biosecurity implications to Australia such as detecting historical pathways of incursion (discussed later).

The mtDNA COI gene has been widely used to differentiate insect species (e.g., [1, 13, 85]). Differentiating dipteran species based on the mtDNA COI gene generally has had a low success rate due to overlapping ranges of genetic distances during intraspecific and interspecific comparisons [80]. Differentiating *Culicoides* species using genetic distances of the mtDNA COI gene may therefore also be of limited value due to overlapping intra- and inter-specific genetic distances (e.g., the lower range between *C. henryi* and *C. pallidothorax* is 0.096 while the maximum intra-specific genetic distance between individuals of *C. brevitarsis* is 0.082 (S5 Table); a difference of only 1.4%). To better understand the genetic relationship between *C. pallidothorax* and *C. henryi*, ecological, behavioural, and geographical information will be needed in addition to morphological and sequence data (e.g. integrative taxonomy, see [86, 87]). A high intraspecific distance (> 0.110) was also reported for *C. obsoletus* and *C. newsteadi*, while the mtDNA barcode method failed to identify two closely related pairs of species as four separate species (i.e., *C. festivipennis* and *C. clastrieri* (0.001); *C. salinarius* and *C. manchuriensis* (0.017)) due to low interspecific distances [35].

### *Culicoides marksii* mtDNA COI haplotypes

Analysis of 16 endemic *C. marksii* individuals from two populations detected greater haplotype diversity compared with all *C. brevitarsis* samples combined, and therefore adds support to the introduced species status of *C. brevitarsis*. *Culicoides marksii* is an important vector of many orbiviruses (e.g., Tables 1 and 2 in [88]). The species feeds readily on marsupials, and is distributed across the whole of Australia and into Papua New Guinea. A rare wing pattern variant of *C. marksii* has been found in the Northern Territory [27]. This, and the high level of genetic variation present in *C. marksii* populations, suggests that a future study should consider examining the likely presence of cryptic species within the currently recognised *C. marksii* species, as this might have been an important factor that contributed to previous unsuccessful artificial BTV infection experiments [88]. The high level of mtDNA haplotype diversity in *C. marksii* contrasted with that found in *C. brevitarsis*. Factors contributing to these differences between the two species are likely to include species-specific differences in flight patterns, dispersal ability, host-assisted or host-mediated dispersal, and preferred environmental/weather conditions and breeding sites (e.g., [41, 89–91]).

### *Culicoides brevitarsis* gene flow patterns inferred from nuDNA markers

Although the mtDNA genome is well suited for studying maternal gene flow patterns, inference of paternal gene flow patterns requires the inclusion of nuDNA markers (e.g., microsatellite and EPIC DNA markers). In this regard, EPIC DNA markers have not been used as widely as microsatellite DNA markers (e.g., [52]), but they are still highly suitable for inferring population genetic structures (e.g., [19]), especially in taxa with transposable elements (TEs) in their genomes, such as the Diptera (e.g., [53] and references therein). Here we demonstrated the usefulness of EPIC DNA markers, in addition to microsatellite DNA markers, in a study of gene flow patterns in Australian *C. brevitarsis*. EPIC and microsatellite DNA markers identified a population structure and gene flow patterns that are consistent with those detected using the mtDNA marker, with additional power to detect isolation by distance and genetic clusters in a species with limited polymorphism in its mtDNA. The discovery of isolation by distance using nuDNA markers but not mtDNA markers may imply sex-specific differences in dispersal patterns, feeding and/or mating behaviours in *C. brevitarsis*, although confirmation of this will require further work. Our STRUCTURE analysis of the nuDNA markers revealed the presence

of a northern cluster and a south cluster and also supported the discovery of isolation by distance, with north-western populations (e.g., Kalumburu) and south-eastern populations (e.g., Paterson) being genetically less similar to each other than they are to their neighbouring populations. Two potential genetic clusters of *C. brevitarsis* from Australia and its neighbouring regions were also identified based on multiple microsatellite DNA markers by Onyango et al. [92], although this study would have benefited also from the inclusion of the maternally inherited mtDNA marker, for inference of genetic structure and incursion pathways from the maternal lineage perspective, given that it is the female species that is the main arbovirus vector.

### Phylogeny of *Culicoides*

The alignment of 182 codons encoding some of the mtDNA CO I gene from *Culicoides* was characterised by compositional heterogeneity across the sequences at third codon sites, but not at first and second codon sites, implying that the mitochondrial genomes of many of the 25 species examined here must have evolved under non-stationary, reversible, and homogeneous (SRH) conditions, and that natural selection must have been sufficiently strong to hide evidence of this at the non-synonymous codon sites. This is consistent with reports on the effects of directional mutation pressure on non-synonymous codon sites of protein-coding genes in animal mitochondria [93]. Since compositional heterogeneity across sequences may bias phylogenetic estimates [94] and most model-based molecular phylogenetic methods assume that the data have evolved under globally SRH conditions, we decided to use first and second codon sites in our phylogenetic analysis of *Culicoides*, even though the number of phylogenetically informative sites was relatively small for these two codon sites. This type of consideration is not as uncommon as it used to be (for examples, see [95, 96] and papers cited therein).

The inferred phylogeny of the mtDNA COI marker from our *Culicoides* samples was characterised by large variation in the root-to-tip distances and bootstrap values ranging from 11% to 100%, which is consistent with the data containing too few phylogenetically informative sites. Therefore, it was difficult to have much confidence in the inferred phylogenetic tree (Fig 6). However, assuming the tree is a good representation of the underlying evolutionary history of *Culicoides*, it would appear that the Australian species are divided into two groups, with *C. brevitarsis* being more closely related to species found in Japan (e.g., *C. jacobsoni*, *C. brevipalpis*, *C. maculatus*, *C. peregrinus*, and *C. actoni*) than species found in Australia, and with the remaining Australian species (*C. marksi*, *C. henryi*, *C. pallidothorax*, *C. bundyensis*, and *C. bun-rooensis*) forming a closely related group of species.

### Biosecurity implications of *C. brevitarsis* northern and eastern mtDNA COI haplotypes

Dyce [39] suggested that Asian or South East Asian species such as *C. brevitarsis*, *C. oxystoma* and *C. peregrinus* may have arrived in Australia after early European settlement and the establishment of domesticated herbivores in the tropical north, with the proposed path to colonisation of *C. brevitarsis* being from Timor-Leste into the northern regions of WA, NT, and Queensland (QLD) prior to splitting into (i) a northern route (to colonise northern QLD and Papua New Guinea, and across to the Melanesian islands (e.g., Solomon Islands, Fiji, and maybe New Caledonia)), and (ii) a southern route (to colonise central QLD, northern and central NSW, and Lord Howe Island) (see Fig 2 of [97]). Using simulation studies, Eagles *et al.* [90] found that if the potential source of the Australian population were Timor-Leste, then, given the current climatic conditions and weather patterns, the Australian regions with the greatest risk of windborne incursion would be the top end of NT and the Kimberley region of

WA. Our analysis of *Culicoides* mtDNA haplotypes from Australia, Timor-Leste, and the Solomon Islands is consistent with the conclusions of Eagles *et al.* [90].

The detected distributional patterns of *C. brevitarsis* mtDNA haplotypes and nuclear genetic clusters imply that multiple independent incursions may have occurred into Australia after the arrival of suitable domesticated cattle hosts (i.e., since European settlement began in Australia). Similarly, Bellis *et al.* [98] also identified evidence to support multiple incursion pathways into Australia from surrounding regions from mtDNA marker analysis. Based on our comparisons of mtDNA haplotypes from neighbouring regions and throughout Australia, four hypotheses can be put forward to assist future in-depth studies of *C. brevitarsis* incursion pathways and population spread: (i) the presence of a single mtDNA haplotype (Cbrev-01) in Timor-Leste and across the distributional range of *C. brevitarsis* in Australia potentially indicates that populations in Timor-Leste may have been the source of Australian populations, and that members of these populations may have been carried by trade winds to northern Australia/north-western Australia before founding the first Australian populations of *C. brevitarsis*; (ii) a relatively long history of Cbrev-01 in northern Australia have enabled the spread of this haplotype to the southern-most Australian limit of this species' habitable range (Fig 1); (iii) the presence of other mtDNA haplotypes (e.g., Cbrev-03 in Kalumburu (in 2008 and 2009) and Katherine (in 2009)) suggests other possible incursion events, with members of likely source populations again being carried by trade winds from, for example, Timor-Leste or Indonesia [90]; and (iv) based on mtDNA haplotypes not present in north-western/northern Australia, at least one additional independent incursion event must have occurred, possibly from New Guinea (i.e., similar to that proposed for *C. dumdum*) [99] into northern QLD, which subsequently spread to our eastern sampling sites (Paterson, Grafton and Lismore).

The small number of specimens analysed from Timor-Leste and the Solomon Islands may explain the limited number of haplotypes detected in these regions. Increasing sample sizes from, for example, Timor-Leste, Papua New Guinea, Indonesia, and Solomon Islands will enable better testing of these hypotheses especially hypothesis (iv) listed above. Adding support to the proposed hypothesis of multi-incursion points into Australia of *C. brevitarsis* is the identification of other region-specific mtDNA haplotypes (i.e., Cbrev-04 and Cbrev-06 in the north and Cbrev-05, Cbrev-07, Cbrev-08, and Cbrev-09 in the east), although further confirmation is needed and will require extensive sampling and population genetic analysis of samples from especially northern QLD, which were not available for this study. Molecular analysis of the BTV serotypes (see [84] for a review) supports Dyce's [97] hypothesis of a northwest-southeast spread, with an origin in the Indian subcontinent, although multiple potential *Culicoides* incursion origins (e.g., from west (e.g., Lombok) to east (Fly (Western Province, PNG) into northern Australia (WA, NT, QLD) have been reported [90, 100]. Furthermore, differences in BTV virulence patterns (Table 3 in [84]) also lend support to the proposed multiple entry points of *C. brevitarsis* based on the presence of disjoint northern and eastern populations, as is evident from the region-specific mtDNA haplotypes, and may be reflected by geographic clines of genetic clusters based on nuclear DNA markers.

## Conclusions

Our results showed that *C. brevitarsis*, with a relatively well-established range across Australia and its neighbouring countries such as Indonesia, Timor-Leste, Papua New Guinea, and Solomon Islands, may have been introduced several times into Australia, as indicated by the heterogeneous patterns of geographical distribution of the mtDNA haplotypes in northern and eastern populations. The results highlight a need for further in-depth studies of incursion

pathways (e.g., significantly increase sampling sites, marker numbers, and sample over multiple seasons) of *C. brevitarsis* and other vectors of viruses in Australia and surrounding regions.

## Supporting Information

**S1 File. A multiple sequence (nucleotide) alignment (individual sequences in Fasta format) of the mtDNA COI partial gene used in this study.**

(PDF)

**S2 File. A multiple sequence (amino acids) alignment (individual sequences in Fasta format) of the mtDNA COI partial gene used in this study.**

(PDF)

**S1 Table. Uncorrected ('p')-pairwise mtDNA COI distances between Australian and Japanese *C. brevitarsis*.**

(DOCX)

**S2 Table. Uncorrected ('p')-pairwise mtDNA COI distances between Australian endemic *C. marksii*.**

(DOCX)

**S3 Table. Uncorrected ('p')-pairwise distances between *C. brevitarsis* from Australia, China, Japan, Solomon Islands and Timor-Leste based on 323bp of over-lapping gene region.** Upper-triangle values are standard errors.

(DOCX)

**S4 Table. Single nucleotide polymorphisms (SNPs) identified through sequence alignment of the mitochondrial DNA cytochrome oxidase I (mtDNA COI) partial genes in *Culicoides brevitarsis*.** SNPs in the consensus sequence were from KJ162970 (Solomon Island) and KP201844 (Cbrev-01). Unique SNPs as compared to the consensus are indicated, identical SNPs are represented by '.'; ambiguous bases A/T are indicated by 'W', and C/T as 'Y'. Start and end nucleotide positions are indicated by square boxes. The KJ162974, Cbrev-01 [KP201844] and Cbrev-05 [KP201848] are *C. brevitarsis* from the same location (Grafton, New South Wales) and were collected on the same date (14-Dec-07) and by the same person (H. McKenzie, New South Wales Department of Primary Industry). Grey-out regions represent non-overlapping mtDNA COI regions between current study, [51], and [61].

(XLSX)

**S5 Table. Uncorrected ("p")-distances between 25 *Culicoides* species across 554bp of partial mtDNA COI gene.**

(DOCX)

## Acknowledgments

We thank the reviewers for constructive comments on the manuscript and David Boyle, Peter Walker, Peter Durr, Owain Edwards, Andy Sheppard, and Glenn Bellis for constructive discussion on aspects of the project. *Culicoides* samples were provided by Luke Halling (Northern Australian Quarantine Strategy), Harry McKenzie (NSW DPI) and Glenn Bellis (AQISNT).

## Author Contributions

Conceived and designed the experiments: WTT PJK. Performed the experiments: WTT PJK LSJ. Analyzed the data: WTT PJK LSJ. Contributed reagents/materials/analysis tools: WTT PJK LSJ. Wrote the paper: WTT PJK LSJ.

## References

1. Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astrartes fulgerator*. *P Natl Acad Sci USA*. 2004; 101(41):14812–7. doi: [10.1073/Pnas.0406166101](https://doi.org/10.1073/Pnas.0406166101) PMID: [ISI:000224488000025](https://pubmed.ncbi.nlm.nih.gov/15100022448800025/).
2. Lehr MA, Kilpatrick CW, Wilkerson RC, Conn JE. Cryptic species in the *Anopheles* (Nyssorhynchus) *albitarsis* (Diptera: Culicidae) complex: Incongruence between random amplified polymorphic DNA-polymerase chain reaction identification and analysis of mitochondrial DNA COI gene sequences. *Ann Entomol Soc Am*. 2005; 98(6):908–17. doi: [10.1603/0013-8746\(2005\)098\[0908:Csitani\]2.0.Co;2](https://doi.org/10.1603/0013-8746(2005)098[0908:Csitani]2.0.Co;2) PMID: [ISI:000233210400016](https://pubmed.ncbi.nlm.nih.gov/151000233210400016/).
3. De Barro PJ, Liu SS, Boykin LM, Dinsdale AB. *Bemisia tabaci*: A Statement of Species Status. *Annu Rev Entomol*. 2011; 56:1–19. doi: [10.1146/annurev-ento-112408-085504](https://doi.org/10.1146/annurev-ento-112408-085504) PMID: [ISI:000286841900001](https://pubmed.ncbi.nlm.nih.gov/151000286841900001/).
4. Pagès N, Munoz-Munoz F, Talavera S, Sarto V, Lorca C, Nunez JI. Identification of cryptic species of *Culicoides* (Diptera: Ceratopogonidae) in the subgenus *Culicoides* and development of species-specific PCR assays based on barcode regions. *Vet Parasitol*. 2009; 165(3–4):298–310. doi: [10.1016/j.vetpar.2009.07.020](https://doi.org/10.1016/j.vetpar.2009.07.020) PMID: [19682796](https://pubmed.ncbi.nlm.nih.gov/19682796/).
5. Tabachnick WJ. *Culicoides variipennis* and bluetongue-virus epidemiology in the United States. *Annu Rev Entomol*. 1996; 41:23–43. PMID: [ISI:A1996TP40300003](https://pubmed.ncbi.nlm.nih.gov/151A1996TP40300003/).
6. Cêtre-Sossah C, Baldet T, Delecolle JC, Mathieu B, Perrin A, Grillet C, et al. Molecular detection of *Culicoides* spp. and *Culicoides imicola*, the principal vector of bluetongue (BT) and African horse sickness (AHS) in Africa and Europe. *Vet Res*. 2004; 35(3):325–37. doi: [10.1051/Vetres.2004015](https://doi.org/10.1051/Vetres.2004015) PMID: [ISI:000222441800007](https://pubmed.ncbi.nlm.nih.gov/151000222441800007/).
7. Klee J, Tay WT, Paxton RJ. Specific and sensitive detection of *Nosema bombi* (Microsporidia: Nosematidae) in bumble bees (*Bombus* spp.; Hymenoptera: Apidae) by PCR of partial rRNA gene sequences. *J Invertebr Pathol*. 2006; 91(2):98–104. PMID: [ISI:000235458200003](https://pubmed.ncbi.nlm.nih.gov/151000235458200003/).
8. Matsumoto Y, Yanase T, Tsuda T, Noda H. Characterization of Internal Transcribed Spacer (ITS1)-ITS2 Region of Ribosomal RNA Gene From 25 Species of *Culicoides* Biting Midges (Diptera: Ceratopogonidae) in Japan. *J Med Entomol*. 2009; 46(5):1099–108. PMID: [ISI:000269576700017](https://pubmed.ncbi.nlm.nih.gov/151000269576700017/).
9. Ritchie A, Blackwell A, Malloch G, Fenton B. Heterogeneity of ITS1 sequences in the biting midge *Culicoides impunctatus* (Goetghebuer) suggests a population in Argyll, Scotland, may be genetically distinct. *Genome*. 2004; 47(3):546–58. doi: [10.1139/G04-003](https://doi.org/10.1139/G04-003) PMID: [ISI:000222350000014](https://pubmed.ncbi.nlm.nih.gov/151000222350000014/).
10. O'Mahony EM, Tay WT, Paxton RJ. Multiple rRNA variants in a single spore of the microsporidian *Nosema bombi*. *J Eukaryot Microbiol*. 2007; 54(1):103–9. PMID: [ISI:000243869500020](https://pubmed.ncbi.nlm.nih.gov/151000243869500020/).
11. Behere GT, Tay WT, Russell DA, Batterham P. Molecular markers to discriminate among four pest species of *Helicoverpa* (Lepidoptera: Noctuidae). *Bulletin of Entomological Research*. 2008; 98:599–603. doi: [10.1017/S0007485308005956](https://doi.org/10.1017/S0007485308005956) PMID: [18631420](https://pubmed.ncbi.nlm.nih.gov/18631420/)
12. Pagès N, Monteys VSI. Differentiation of *Culicoides obsoletus* and *Culicoides scoticus* (Diptera: ceratopogonidae) based on mitochondrial cytochrome oxidase subunit I. *J Med Entomol*. 2005; 42(6):1026–34. doi: [10.1603/0022-2585\(2005\)042\[1026:Docoac\]2.0.Co;2](https://doi.org/10.1603/0022-2585(2005)042[1026:Docoac]2.0.Co;2) PMID: [ISI:000233211000016](https://pubmed.ncbi.nlm.nih.gov/151000233211000016/).
13. Tay WT, Beckett SJ, De Barro PJ. Phosphine resistance in Australian *Cryptolestes* species (Coleoptera: Laemophloeidae): Perspectives from mitochondrial DNA Cytochrome Oxidase I analysis. *Pest Manag Sci*. 2014. Epub 21 April 2014. doi: [10.1002/ps.3805](https://doi.org/10.1002/ps.3805)
14. Hadj-Henni L, De Meulemeester T, Mathieu B, Depaquit J, Augot D. Taxonomic assessment of *Culicoides brunnicans*, *C. santonicus* and *C. vexans* (Diptera: Ceratopogonidae) in France: Implications in systematics. *Infect Genet Evol*. 2015; 33:324–31. doi: [10.1016/j.meegid.2015.05.024](https://doi.org/10.1016/j.meegid.2015.05.024) PMID: [26005070](https://pubmed.ncbi.nlm.nih.gov/26005070/).
15. Tay WT, Soria MF, Walsh T, Thomazoni D, Silvie P, Behere GT, et al. A brave new world for an Old World pest: *Helicoverpa armigera* (Lepidoptera: Noctuidae) in Brazil. *PLoS One*. 2013; 8(11):e80134. doi: [10.1371/journal.pone.0080134](https://doi.org/10.1371/journal.pone.0080134) PMID: [24260345](https://pubmed.ncbi.nlm.nih.gov/24260345/)
16. Behere GT, Tay WT, Russell DA, Heckel DG, Appleton BR, Kranthi KR, et al. Mitochondrial DNA Analysis of Field Populations of *Helicoverpa armigera* (Lepidoptera: Noctuidae) and of its relationship to *H. zea*. *BMC Evol Biol*. 2007; 7:117. PMID: [17629927](https://pubmed.ncbi.nlm.nih.gov/17629927/)



17. Carew ME, Tay WT, Crozier RH. Polygyny via unrelated queens indicated by mitochondrial DNA variation in the Australian meat ant *Iridomyrmex purpureus*. *Insectes Sociaux*. 1997; 44(1):7–14. PMID: [ISI:A1997WP70200002](#).
18. Tay WT, Cook JM, Rowe DJ, Crozier RH. Migration between nests in the Australian arid-zone ant *Rhytidoponera* sp. 12 revealed by DGGE analyses of mitochondrial DNA. *Mol Ecol*. 1997; 6(5):403–11. PMID: [ISI:A1997WY86400001](#).
19. Behere GT, Tay WT, Russell DA, Kranthi KR, Batterham P. Population genetic structure of the cotton bollworm *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae) in India as inferred from EPIC-PCR DNA markers. *PLoS One*. 2013; 8(1):e53448. doi: [10.1371/journal.pone.0053448](#) PMID: [23326431](#)
20. Gongora J, Cuddahee RE, Do Nascimento FF, Palgrave CJ, Lowden S, Ho SYW, et al. Rethinking the evolution of extant sub-Saharan African suids (Suidae, Artiodactyla). *Zoologica Scripta*. 2011; 40(4):327–35.
21. Tay WT, Crozier RH. Mating behaviour of *Rhytidoponera* sp. 12 ants inferred from microsatellite analysis. *Mol Ecol*. 2001; 10(1):167–73. PMID: [ISI:000166647100015](#).
22. Tay WT, Seppä P, Pamilo P. Colony kin structure, reproductive dominance and colony founding strategies in the ant *Rhytidoponera* sp. 12 (Hymenoptera: Formicidae). *Myrmecological News*. 2011; 15:67–76.
23. Quezada-Euan JJG, Paxton RJ, Palmer KA, Itza WDM, Tay WT, Oldroyd BP. Morphological and molecular characters reveal differentiation in a Neotropical social bee, *Melipona beecheii* (Apidae: Meliponini). *Apidologie*. 2007; 38(3):247–58. PMID: [ISI:000248083300005](#).
24. Brokent A, Wirth WW. World species of biting midges (Diptera: Ceratopogonidae). *Bulletin of the American Museum of Natural History* 1997. p. 1–257.
25. Brokent A. World Species of Biting Midges (Diptera: Ceratopogonidae) 2012 [updated February 28, 2012; cited 18-Oct-2015]. Available from: <http://www.inhs.illinois.edu/files/8413/4219/9566/CeratopogonidaeCatalog.pdf>.
26. Harrup LE, Bellis GA, Balenghien T, Garros C. *Culicoides* Latreille (Diptera: Ceratopogonidae) taxonomy: current challenges and future directions. *Infect Genet Evol*. 2015; 30:249–66. doi: [10.1016/j.meegid.2014.12.018](#) PMID: [25535946](#); PubMed Central PMCID: PMCPCMC4330985.
27. Dyce AL, Bellis GA, Muller MJ. Pictorial Atlas of Australasian *Culicoides* Wings (Diptera: Ceratopogonidae). Wells A, editor. Canberra: ABRIS, Canberra; 2007. 84 p.
28. Meiswinkel R. Adult characters defining and separating the *Imicola* and *Orientalis* species complexes of the subgenus *Avaritia* Fox, 1955 (Culicoides, Diptera: Ceratopogonidae). *Vet Ital*. 2004; 40(3):345–51. PMID: [ISI:000234682200061](#).
29. Beveridge I, Kummerow EL, Wilkinson P, Copeman DB. An Investigation of Biting Midges in Relation to Their Potential as Vectors of Bovine Onchocerciasis in North Queensland. *J Aust Entomol Soc*. 1981; 20:39–45. PMID: [ISI:A1981LD99800007](#).
30. Holbrook FR, Tabachnick WJ, Brady R. Genetic variation in populations of *Culicoides variipennis* complex in the six New England states, USA. *Med Vet Entomol*. 1996; 10(2):173–80. doi: [10.1111/J.1365-2915.1996.Tb00725.X](#) PMID: [ISI:A1996UL34300010](#).
31. Sebastiani F, Meiswinkel R, Gomulski LM, Guglielmino CR, Mellor PS, Malacrida AR, et al. Molecular differentiation of the Old World *Culicoides imicola* species complex (Diptera, Ceratopogonidae), inferred using random amplified polymorphic DNA markers. *Mol Ecol*. 2001; 10(7):1773–86. PMID: [ISI:000170006100016](#).
32. Perrin A, Cetre-Sossah C, Mathieu B, Baldet T, Delecolle JC, Albina E. Phylogenetic analysis of *Culicoides* species from France based on nuclear ITS1-rDNA sequences. *Med Vet Entomol*. 2006; 20(2):219–28. doi: [10.1111/J.1365-2915.2006.00616.X](#) PMID: [ISI:000239629100008](#).
33. Gomulski LM, Meiswinkel R, Delecolle JC, Goffredo M, Gasperi G. Phylogenetic relationships of the subgenus *Avaritia* Fox, 1955 including *Culicoides obsoletus* (Diptera, Ceratopogonidae) in Italy based on internal transcribed spacer 2 ribosomal DNA sequences. *Syst Entomol*. 2005; 30(4):619–31. doi: [10.1111/J.1365-3113.2005.00286.X](#) PMID: [ISI:000232491700007](#).
34. Dallas JF, Cruickshank RH, Linton YM, Nolan DV, Patakakis M, Braverman Y, et al. Phylogenetic status and matrilineal structure of the biting midge, *Culicoides imicola*, in Portugal, Rhodes and Israel. *Med Vet Entomol*. 2003; 17(4):379–87. doi: [10.1111/J.1365-2915.2003.00454.X](#) PMID: [ISI:000186973700004](#).
35. Ander M, Troell K, Chirico J. Barcoding of biting midges in the genus *Culicoides*: a tool for species determination. *Med Vet Entomol*. 2013; 27(3):323–31. doi: [10.1111/J.1365-2915.2012.01050.X](#) PMID: [ISI:000322992400009](#).

36. Augot D, Ninio C, Akhouni M, Lehrter V, Couloux A, Jouet D, et al. Characterization of two cryptic species, *Culicoides stigma* and *C. parroti* (Diptera: Ceratopogonidae), based on barcode regions and morphology. *J Vector Ecol.* 2013; 38(2):260–5. PMID: [ISI:000326923500009](#). doi: [10.1111/j.1948-7134.2013.12039.x](#)
37. Nolan DV, Carpenter S, Barber J, Mellor PS, Dallas JF, Mordue AJ, et al. Rapid diagnostic PCR assays for members of the *Culicoides obsoletus* and *Culicoides pulicaris* species complexes, implicated vectors of bluetongue virus in Europe. *Vet Microbiol.* 2007; 124(1–2):82–94. doi: [10.1016/J.Vetmic.2007.03.019](#) PMID: [ISI:000249306600009](#).
38. Nolan DV, Dallas JF, Mordue AJ. Molecular taxonomy and population structure of a *Culicoides* midge vector. *Vet Ital.* 2004; 40(3):352–9. PMID: [ISI:000234682200062](#).
39. Dyce AL, editor Bluetongue vectors in Australia—Present and Future. *Arbovirus Research in Australia*; 1989 August 28–September 1, 1989; Brisbane, Australia: Proceedings Fifth Symposium.
40. Doyle KA, editor An overview and perspective on orbivirus disease prevalence and occurrence of vectors in Australia and Oceania. *Bluetongue, African horse sickness, and related orbiviruses*; 1992; Boca Raton: CRC Press, Inc.
41. Kettle DS. Bionomics and Control of *Culicoides* and *Leptoconops*—(Diptera, Ceratopogonidae = Heleidae). *Annu Rev Entomol.* 1962; 7:401–18. doi: [10.1146/Annurev.En.07.010162.002153](#) PMID: [ISI:A19624972A00009](#).
42. Campbell MM, Kettle DS. Number of Adult *Culicoides brevitarsis* Kieffer (Diptera-Ceratopogonidae) Emerging from Bovine Dung Exposed under Different Conditions in Field. *Aust J Zool.* 1976; 24(1):75–85. doi: [10.1071/Zo9760075](#) PMID: [ISI:A1976BH15300008](#).
43. Cannon LRG, Reye EJ. A larval habitat of the biting midge *Culicoides brevitarsis* Kieffer (Diptera:Ceratopogonidae). *J Entomol Soc Queensl.* 1966; 5:7–9.
44. Yanase T, Matsumoto Y, Matsumori Y, Aizawa M, Hirata M, Kato T, et al. Molecular Identification of Field-Collected *Culicoides* Larvae in the Southern Part of Japan. *J Med Entomol.* 2013; 50(5):1105–10. doi: [10.1603/Me11235](#) PMID: [ISI:000324374100021](#).
45. Kay BH, Boreham PFL, Dyce AL, Standfast HA. Blood Feeding of Biting Midges (Diptera-Ceratopogonidae) at Kowanyama, Cape-York Peninsula, North Queensland. *J Aust Entomol Soc.* 1978; 17:145–9. PMID: [ISI:A1978FA86300006](#).
46. Muller MJ, Murray MD. Blood-Sucking Flies Feeding on Sheep in Eastern Australia. *Aust J Zool.* 1977; 25(1):75–85. doi: [10.1071/Zo9770075](#) PMID: [ISI:A1977CV74900009](#).
47. Standfast HA, Muller MJ, Dyce AL, editors. An overview of bluetongue virus vector biology and ecology in the Oriental and Australasian regions of the western Pacific. *Bluetongue, African horse sickness, and related orbiviruses*; 1992; Paris, France: Proceedings of the Second International Symposium.
48. Carpenter S, Wilson A, Mellor PS. *Culicoides* and the emergence of bluetongue virus in northern Europe. *Trends Microbiol.* 2009; 17(4):172–8. doi: [10.1016/J.Tim.2009.01.001](#) PMID: [ISI:000265477600006](#).
49. Purse BV, Mellor PS, Rogers DJ, Samuel AR, Mertens PPC, Baylis M. Climate change and the recent emergence of bluetongue in Europe. *Nat Rev Microbiol.* 2005; 3(2):171–81. doi: [10.1038/Nrmicro1090](#) PMID: [ISI:000226691200015](#).
50. Wittmann EJ, Baylis M. Climate change: Effects on *Culicoides*-transmitted viruses and implications for the UK. *Vet J.* 2000; 160(2):107–17. PMID: [ISI:000089490900006](#).
51. Bellis GA, Dyce A, Gopurenko D, Yanase T, Garros C, Labuschagne K, et al. Revision of the *Culicoides (Avaritia) Imicola* complex Khamala & Kettle (Diptera: Ceratopogonidae) from the Australasian region. *Zootaxa.* 2013; 3768(4):401.
52. Tay WT, Behere GT, Heckel DG, Lee SF, Batterham P. Exon-Primed Intron-Crossing (EPIC) PCR markers of *Helicoverpa armigera* (Lepidoptera: Noctuidae). *Bulletin of Entomological Research.* 2008; 98:509–18. doi: [10.1017/S000748530800583X](#) PMID: [18826667](#)
53. Tay WT, Behere GT, Batterham P, Heckel DG. Generation of microsatellite repeat families by RTE retrotransposons in lepidopteran genomes. *BMC Evol Biol.* 2010; 10:144. Epub 2010/05/18. doi: [10.1186/1471-2148-10-144](#) PMID: [20470440](#); PubMed Central PMCID: PMC2887409.
54. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research.* 1997; 25:3389–402. PMID: [9254694](#)
55. Gordon K, Tay WT, Collinge D, Williams A, Batterham P. Genetics and molecular biology of the major crop pest genus *Helicoverpa*. In: Goldsmith MR, Marec F, editors. *Molecular Biology and Genetics of the Lepidoptera*. Contemporary Topics in Entomology CRC Press; 2010. p. 219–38.

56. Staden R, Beal KF, Bonfield JK. The Staden package, 1998. *Methods in Molecular Biology*. 2000; 132:115–30. PMID: [10547834](#)
57. Tamura K, Nei M. Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial-DNA in Humans and Chimpanzees. *Mol Biol Evol*. 1993; 10(3):512–26. PMID: [ISI:A1993LD11400002](#).
58. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol*. 2011; 28:2731–9. Epub 2011/05/07. doi: [10.1093/molbev/msr121](#) PMID: [21546353](#).
59. Posada D, Grandall KA. Modeltest: testing the model of DNA substitution. *Bioinformatics*. 1998; 14(9):817–8. PMID: [9918953](#)
60. Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res*. 2005; 33:D54–D8. doi: [10.1093/Nar/Gki031](#) PMID: [ISI:000226524300009](#).
61. Matsumoto Y, Yanase T, Tsuda T, Noda H. Species-specific mitochondrial gene rearrangements in biting midges and vector species identification. *Med Vet Entomol*. 2009; 23(1):47–55. doi: [10.1111/J.1365-2915.2008.00789.X](#) PMID: [ISI:000263331100008](#).
62. Clement M, Posada D, Crandall KA. TCS: a computer program to estimate gene genealogies. *Molecular Ecology*. 2000; 9(10):1657–9. PMID: [ISI:000089998600020](#).
63. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population-Structure. *Evolution*. 1984; 38(6):1358–70. doi: [10.2307/2408641](#) PMID: [ISI:A1984TY40400017](#).
64. Raymond M, Rousset F. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*. 1995; 86:248–9.
65. Rousset F. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol Ecol Resour*. 2008; 8(1):103–6. doi: [10.1111/J.1471-8286.2007.01931.X](#) PMID: [ISI:000253827100016](#).
66. Excoffier L, Smouse PE, Quattro JM. Analysis of Molecular Variance Inferred from Metric Distances among DNA Haplotypes—Application to Human Mitochondrial-DNA Restriction Data. *Genetics*. 1992; 131(2):479–91. PMID: [ISI:A1992HW75900021](#).
67. Schneider SD, Roessli D, Excoffier L. ARLEQUIN (version 2 000): A software for genetic data analysis. Geneva, Switzerland.: Genetics and Biometry Laboratory, University of Geneva; 2000.
68. Kettle DS, Edwards PB, Barnes A. Factors affecting numbers of *Culicoides* in truck traps in coastal Queensland. *Med Vet Entomol*. 1998; 12(4):367–77. doi: [10.1046/J.1365-2915.1998.00119.X](#) PMID: [ISI:000076933100004](#).
69. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. *Scand J Stat*. 1979; 6(2):65–70. PMID: [ISI:A1979JY78700003](#).
70. Slatkin M. A Measure of Population Subdivision Based on Microsatellite Allele Frequencies. *Genetics*. 1995; 139(1):457–62. PMID: [ISI:A1995PZ24800039](#).
71. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000; 155(2):945–59. PMID: [ISI:000087475100039](#).
72. Gouy M, Guindon S, Gascuel O. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Mol Biol Evol*. 2010; 27(2):221–4. doi: [10.1093/Molbev/Msp259](#) PMID: [ISI:000273704400003](#).
73. Ababneh F, Jermini LS, Ma CS, Robinson J. Matched-pairs tests of homogeneity with applications to homologous nucleotide sequences. *Bioinformatics*. 2006; 22(10):1225–31. doi: [10.1093/Bioinformatics/Btl064](#) PMID: [ISI:000237319300012](#).
74. Ababneh F, Jermini LS, Robinson J. Generation of the exact distribution and simulation of matched nucleotide sequences on a phylogenetic tree. *Journal of Mathematical Modelling and Algorithms*. 2006; 5:291–303.
75. Bryant D, Galtier N, Poursat MA. Likelihood calculation in molecular phylogenetics. In: Gascuel O, editor. *In Mathematics evolution and phylogeny*. New York: Oxford University Press. Inc.; 2005. p. 33–62.
76. Jayaswal V, Jermini LS, Robinson J. Estimation of Phylogeny Using a General Markov Model. *Evol Bioinform*. 2005; 1:62–80. PMID: [ISI:000207065900006](#).
77. Jermini LS, Jayaswal V, Ababneh F, Robinson J. Phylogenetic model evaluation. In: Keith J, editor. *Bioinformatics: Data, sequence analysis, and evolution*. 1. Totawa: Humana Press; 2008. p. 331–63.
78. Bowker AH. A Test for Symmetry in Contingency Tables. *J Am Stat Assoc*. 1948; 43(244):572–4. doi: [10.2307/2280710](#) PMID: [ISI:A1948XY35500008](#).

79. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast Approximation for Phylogenetic Bootstrap. *Mol Biol Evol.* 2013; 30(5):1188–95. doi: [10.1093/molbev/mst024](https://doi.org/10.1093/molbev/mst024) PMID: [ISI:000318165700018](https://pubmed.ncbi.nlm.nih.gov/24046721/).
80. Meier R, Shiyang K, Vaidya G, Ng PKL. DNA barcoding and taxonomy in Diptera: A tale of high intra-specific variability and low identification success. *Systematic Biol.* 2006; 55(5):715–28. doi: [10.1080/10635150600969864](https://doi.org/10.1080/10635150600969864) PMID: [ISI:000246721800001](https://pubmed.ncbi.nlm.nih.gov/16800001/).
81. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* 2005; 14(8):2611–20. doi: [10.1111/J.1365-294x.2005.02553.X](https://doi.org/10.1111/J.1365-294x.2005.02553.X) PMID: [ISI:000229961500029](https://pubmed.ncbi.nlm.nih.gov/1500029/).
82. Yang Z. *Computational Molecular Evolution*. Oxford: Oxford University Press; 2006. 357 p.
83. Murray MD, Dyce AL. Native Insects of Australia as Disease Vectors—a Review of Some Current Studies. *Aust Vet J.* 1970; 46(4):138–8. doi: [10.1111/J.1751-0813.1970.Tb01975.X](https://doi.org/10.1111/J.1751-0813.1970.Tb01975.X) PMID: [ISI:A1970G145400008](https://pubmed.ncbi.nlm.nih.gov/145400008/).
84. Daniels PW, Sendow I, Pritchard LI, Sukarsih, Eaton BT. Regional overview of bluetongue viruses in South-East Asia: viruses, vectors and surveillance. *Veter Ital Ser.* 2004; 40(3):94–100. PMID: [ISI:000234682200012](https://pubmed.ncbi.nlm.nih.gov/2200012/).
85. Conflitti IM, Pruess KP, Cywinska A, Powers TO, Currie DC. DNA Barcoding Distinguishes Pest Species of the Black Fly Genus *Cnephia* (Diptera: Simuliidae). *J Med Entomol.* 2013; 50(6):1250–60. doi: [10.1603/Me13063](https://doi.org/10.1603/Me13063) PMID: [ISI:000327448700009](https://pubmed.ncbi.nlm.nih.gov/2448700009/).
86. Schlick-Steiner BC, Steiner FM, Seifert B, Stauffer C, Christian E, Crozier RH. Integrative taxonomy: a multisource approach to exploring biodiversity. *Annu Rev Entomol.* 2010; 55:421–38. doi: [10.1146/annurev-ento-112408-085432](https://doi.org/10.1146/annurev-ento-112408-085432) PMID: [19737081](https://pubmed.ncbi.nlm.nih.gov/19737081/).
87. Yeates DK, Seago A, Nelson L, Cameron SL, Joseph L, Trueman JWH. Integrative taxonomy, or iterative taxonomy? *Syst Entomol.* 2011; 36(2):209–17. doi: [10.1111/j.1365-3113.2010.00558.x](https://doi.org/10.1111/j.1365-3113.2010.00558.x) PMID: [WOS:000288449300001](https://pubmed.ncbi.nlm.nih.gov/288449300001/).
88. Muller MJ, Standfast HA, St George TD, Cybinski DH, editors. *Culicoides brevitarsis* (Diptera: Ceratopogonidae) as a vector of arboviruses in Australia. *Arbovirus Research in Australia*; 1982; Brisbane: Proceedings 3rd Symposium.
89. Bishop AL, Kirkland PD, Mckenzie HJ, Spohr LJ, Barchia IM, Muller MJ. Distribution and Seasonal Movements of *Culicoides brevitarsis* Kieffer (Diptera, Ceratopogonidae) at the Southern Limits of Its Distribution in New-South-Wales and Their Correlation with Arboviruses Affecting Livestock. *J Aust Entomol Soc.* 1995; 34:289–98. PMID: [ISI:A1995TJ22300005](https://pubmed.ncbi.nlm.nih.gov/1995TJ22300005/).
90. Eagles D, Deveson T, Walker PJ, Zalucki MP, Durr P. Evaluation of long-distance dispersal of *Culicoides* midges into northern Australia using a migration model. *Med Vet Entomol.* 2012; 26(3):334–40. doi: [10.1111/J.1365-2915.2011.01005.X](https://doi.org/10.1111/J.1365-2915.2011.01005.X) PMID: [ISI:000307542900011](https://pubmed.ncbi.nlm.nih.gov/24900011/).
91. Lee DJ, Reye EJ. Australasian Ceratopogonidae (Diptera, Nematocera) Part VI. Australian species of *Culicoides*. *Proceedings of the Linnean Society of New South Wales.* 1953; 77:269–394.
92. Onyango MG, Beebe NW, Gopurenko D, Bellis G, Nicholas A, Ogugo M, et al. Assessment of population genetic structure in the arbovirus vector midge, *Culicoides brevitarsis* (Diptera: Ceratopogonidae), using multi-locus DNA microsatellites. *Vet Res.* 2015; 46:108. doi: [10.1186/s13567-015-0250-8](https://doi.org/10.1186/s13567-015-0250-8) PMID: [26408175](https://pubmed.ncbi.nlm.nih.gov/26408175/); PubMed Central PMCID: [PMCPMC4582633](https://pubmed.ncbi.nlm.nih.gov/PMC4582633/).
93. Jermini LS, Graur D, Lowe RM, Crozier RH. Analysis of Directional Mutation Pressure and Nucleotide Content in Mitochondrial Cytochrome-B Genes. *J Mol Evol.* 1994; 39(2):160–73. PMID: [ISI:A1994NX16800007](https://pubmed.ncbi.nlm.nih.gov/16800007/).
94. Jermini LS, Ho SYW, Ababneh F, Robinson J, Larkum AWD. The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. *Syst Biol.* 2004; 53(4):638–43. doi: [10.1080/10635150490468648](https://doi.org/10.1080/10635150490468648) PMID: [ISI:000223598500009](https://pubmed.ncbi.nlm.nih.gov/223598500009/).
95. Jayaswal V, Wong TKF, Robinson J, Poladian L, Jermini LS. Mixture models of nucleotide sequence evolution that account for heterogeneity in the substitution process across sites and across lineages. *Syst Biol.* 2014; 63:726–42. doi: [10.1093/sysbio/syu036](https://doi.org/10.1093/sysbio/syu036) PMID: [24927722](https://pubmed.ncbi.nlm.nih.gov/24927722/)
96. Cooper ED. Overly simplistic substitution models obscure green plant phylogeny. *Trends Plant Sci.* 2014; 19(9):576–82. doi: [10.1016/J.Tplants.2014.06.006](https://doi.org/10.1016/J.Tplants.2014.06.006) PMID: [ISI:000342269700007](https://pubmed.ncbi.nlm.nih.gov/269700007/).
97. Dyce AL, editor *Distribution of Culicoides (Avaritia) spp. (Diptera: Ceratopogonidae) west of the Pacific Ocean*. *Arbovirus Research in Australia*; 1982; Brisbane: Proceedings 3rd Symposium.
98. Bellis GA, Gopurenko D, Cookson B, Postle AC, Halling L, Harris N, et al. Identification of incursions of *Culicoides* Latreille species (Diptera: Ceratopogonidae) in Australasia using morphological techniques and DNA barcoding. *Austral Entomol.* 2015; 54(3):332–8. doi: [10.1111/aen.12131](https://doi.org/10.1111/aen.12131) PMID: [WOS:000360213900013](https://pubmed.ncbi.nlm.nih.gov/213900013/).

99. Bellis GA, Dyce AL. Intraspecific variation in three species of *Culicoides* in the *Orientalis* complex of the subgenus *Avaritia* within the Australasian zoogeographic region. *Arbovirus Research in Australia*. 2005; 9:33–6.
100. Eagles D, Walker PJ, Zalucki MP, Durr PA. Modelling spatio-temporal patterns of long-distance *Culicoides* dispersal into northern Australia. *Prev Vet Med*. 2013; 110(3–4):312–22. doi: [10.1016/j.pvetmed.2013.02.022](https://doi.org/10.1016/j.pvetmed.2013.02.022) PMID: [ISI:000320427100004](https://pubmed.ncbi.nlm.nih.gov/242710004/).