Molecular Phylogenetics and Evolution 195 (2024) 108061

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Machine learning approaches delimit cryptic taxa in a previously intractable species complex

Haley L.A. Heine [a], Shahan Derkarabetian [b,1], Rina Morisawa [a], Phoebe A. Fu [a], Nathaniel H.W. Moyes [a,2], Sarah L. Boyer [a,*]

[a] Biology Department, Macalester College, 1600 Grand Ave., St. Paul, MN 55105, USA
[b] Museum of Comparative Zoology, Harvard University, 26 Oxford St., Cambridge, MA 02138, USA

ARTICLE INFO

Keywords:
Opiliones
Cyphophthalmi
Aotearoa
New Zealand

ABSTRACT

Cryptic species are not diagnosable via morphological criteria, but can be detected through analysis of DNA sequences. A number of methods have been developed for identifying species based on genetic data; however, these methods are prone to over-splitting taxa with extreme population structure, such as dispersal-limited organisms. Machine learning methodologies have the potential to overcome this challenge. Here, we apply such approaches, using a large dataset generated through hybrid target enrichment of ultraconserved elements (UCEs). Our study taxon is the *Aoraki denticulata* species complex, a lineage of extremely low-dispersal arachnids endemic to the South Island of Aotearoa New Zealand. This group of mite harvesters has been the subject of previous species delimitation studies using smaller datasets generated through Sanger sequencing and analytical approaches that rely on multispecies coalescent models and barcoding gap discovery. Those analyses yielded a number of putative cryptic species that seems unrealistic and extreme, based on what we know about species' geographic ranges and genetic diversity in non-cryptic mite harvesters. We find that machine learning approaches, on the other hand, identify cryptic species with geographic ranges that are similar to those seen in other morphologically diagnosable mite harvesters in Aotearoa New Zealand's South Island. We performed both unsupervised and supervised machine learning analyses, the latter with training data drawn either from animals broadly (vagile and non-vagile) or from a custom training dataset from dispersal-limited harvesters. We conclude that applying machine learning approaches to the analysis of UCE-derived genetic data is an effective method for delimiting species in complexes of low-vagility cryptic species, and that the incorporation of training data from biologically relevant analogues can be critically informative.

1. Introduction

Ongoing destruction and disruption of the Earth's natural systems have accelerated extinction rates, posing a critical threat to global biodiversity (Urban, 2015; Román-Palacios and Wiens, 2020). As a majority of species remain undocumented in the scientific literature, there is an increased urgency to discover and describe biodiversity, especially in invertebrate taxa such as insects and arachnids where ~ 80 % of the estimated species diversity remains unknown or undescribed (Chapman, 2006). Cryptic species, typically defined as two or more distinct species which are not diagnosable morphologically, probably comprise a large portion of undiscovered diversity (Bickford et al., 2007; Struck et al., 2018). Because groups of cryptic species are or may have been classified as single nominal species, true species diversity, even within well-described taxa, is likely underestimated. In addition to developing a more accurate understanding of species diversity, identification of cryptic species may have critical impacts on conservation, biological control, and public health (Bickford et al., 2007).

Species are often described as the fundamental unit of biodiversity, or "the basic rank of classification" in the words of the International Commission of Zoological Nomenclature (Mallet et al., 2022). However, defining species is notoriously difficult and has at times been a topic of

* Corresponding author.
E-mail addresses: haleyheine@gmail.com (H.L.A. Heine), sderkarabetian@gmail.com (S. Derkarabetian), rina.morisawa1@gmail.com (R. Morisawa), phoebefu954@gmail.com (P.A. Fu), nathanielmoyes@gmail.com (N.H.W. Moyes), boyer@macalester.edu (S.L. Boyer).
[1] Present address: San Diego Natural History Museum, 1788 El Prado, Balboa Park, San Diego, CA 92101.
[2] Present address: Department of Zoology, #3051 - 6270 University Blvd., Vancouver, BC V6T 1Z4, Canada.

https://doi.org/10.1016/j.ympev.2024.108061
Received 2 October 2023; Received in revised form 5 March 2024; Accepted 11 March 2024
Available online 12 March 2024

heated debate in the scientific community (e.g. Wheeler and Meier, 1998). At present, some consensus exists; notably, the unified species concept of de Queiroz (2007) is grounded in the argument that various "competing" species concepts in fact share a common element of conceptualizing species as separately evolving metapopulation lineages - that is, a species is an inclusive population made of connected subpopulations and is separated from other metapopulations (see Levins, 1970; Hanski & Gaggiotti, 2004).

Even in the presence of agreement about species' ontology, the epistemological processes of identifying and delimiting species can pose extreme challenges. This is particularly true in the case of cryptic species, in which morphological information cannot be used for diagnosis, and delimitation efforts must therefore rely heavily or even exclusively on genetic data. Over the past generation, many novel methods have been developed for this purpose. In particular, researchers now have a wide variety of sophisticated multispecies coalescent (MSC) species delimitation methods available to them (e.g. Jones et al. 2014, Leaché et al. 2014, Yang and Ranalla 2014). These methods often assume a neutral coalescent - a model in which gene trees evolve within species without genetic differentiation, i.e. in a state of panmixia and without selection or recombination. Unfortunately, this assumption is violated in some systems, notably low-dispersal taxa, which have extremely high population structure, resulting in overestimation of species diversity by MSC delimitation (Jackson et al., 2017; Sukumaran & Knowles, 2017). This result has been demonstrated empirically across a variety of taxa that share similar characteristics (e.g. Niemiller et al., 2012; Barley et al., 2013; Hedin et al., 2015; Derkarabetian et al., 2022b), and even some that do not (e.g., Chambers and Hillis, 2020). Barcoding gap identification methods are prone to similar pitfalls (e.g. Fernández and Giribet, 2014).

Recently, machine learning (ML) approaches have been proposed as complementary methods for species delimitation (Derkarabetian et al., 2022a,b; Pei et al., 2018; Smith and Carstens, 2020). Broadly speaking, ML uses dimensionality reduction to identify latent variables that explain dataset structure. In the case of delimitation of cryptic species, these datasets consist of DNA sequences while the reduced-dimension clusters are species that are not detectable using morphological or ecological criteria. Unsupervised ML proceeds without any user input of *a priori* species groupings, while supervised ML can incorporate training data derived from closely related biologically and ecologically similar lineages in which species boundaries are already understood with confidence (e.g. Derkarabetian et al., 2022b).

Among arachnids, cryptic species have been identified in spiders (e.g. Duncan et al., 2010; Leavitt et al., 2015; Agnarsson et al., 2016; Tyagi et al., 2019), mites (e.g. Skoracka et al., 2015; Pfingstl et al., 2021), scorpions (e.g. Graham et al., 2019), whip spiders (e.g. Reveillion et al., 2020), pseudoscorpions (e.g. Ohira et al., 2018; Muster et al., 2021), and Opiliones - commonly known as harvesters or daddy long-legs (e.g. Arthofer et al., 2013; Clouse and Wheeler, 2014; Derkarabetian et al., 2022a,b; Martens, 2011). Opiliones in particular often display poor dispersal abilities and niche conservatism (Wiens and Graham, 2005), which can lead to non-ecological speciation, resulting in the morphological and ecological stasis that is the hallmark of cryptic species (Czekanski-Moir & Rundell, 2019). This greatly diminishes the utility of morphological and ecological species delimitation criteria in this group. Additionally, in Opiliones it can be difficult or impossible to directly test for reproductive isolation because sympatry between congeners is rare; they tend to show "nested" allopatry, where closely related species, and closely related populations within those species, are completely allopatric in distribution (Derkarabetian et al., 2011, 2019a). Given this combination of characteristics, it is unsurprising that many Opiliones are short-range endemics (Harvey, 2002), known from one or a small number of sites (e.g., Emata and Hedin, 2016). Unfortunately, the low gene flow present in these taxa can make species delimitation based entirely on genetic data difficult, as analyses may not be able to differentiate between high population structure within species and species-level divergences, resulting in the overestimation of species counts (Derkarabetian et al., 2019a; Fernández and Giribet, 2014). Recently, the use of unsupervised (Derkarabetian et al., 2019a) and supervised (Derkarabetian et al., 2022b) machine learning for species delimitation has shown promise as a robust approach to delimiting cryptic species in these challenging taxa. One such harvester taxon for which species delimitation has been problematic is the *Aoraki denticulata* species complex.

Aoraki denticulata (Forster, 1948) is a widespread species of mite harvester (Arachnida; Opiliones; Cyphophthalmi) endemic to the South Island of Aotearoa New Zealand. Ray Forster, who pioneered documentation of the Opiliones of the archipelago, originally described two subspecies: *A. denticulata denticulata* and *A. denticulata major*. *A. denticulata major,* distinguished from *A. denticulata denticulata* by its larger size, is restricted to localities in Arthur's Pass (Fig. 1) and has previously been demonstrated to be monophyletic (Boyer et al., 2007, 2022; Boyer and Giribet, 2007; Fernández and Giribet, 2014). In contrast, the subspecies *A. denticulata denticulata* has an extremely large range (Fig. 1), and multiple studies have demonstrated that it is paraphyletic with three morphologically distinct and geographically isolated lineages nested within: *A. denticulata major*, *A. longitarsa* (Forster, 1952), and *A. meridialis* Boyer, Hahn, & Ward 2022 (Fig. 1) (Boyer et al., 2007, 2022, Boyer and Giribet, 2007, 2009; Fernández and Giribet, 2014). Notably, one study based on cytochrome *c* oxidase subunit I (COI) found deep genetic divergences and high population structure within *A. denticulata denticulata*, with F_{ST} values greater than 0.8 in most

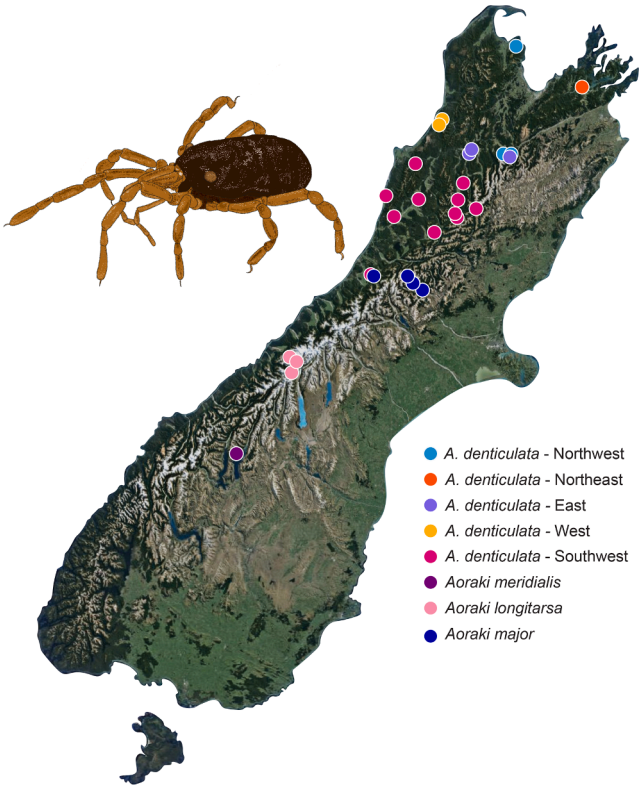

Fig. 1. Distribution of the *Aoraki denticulata* complex in the South Island of Aotearoa New Zealand. Points indicate collections included in the current study. Forster's subspecies *Aoraki denticulata denticulata* and *Aoraki denticulata major* are referred to simply as *A. denticulata* and *A. major* here and elsewhere. Animals from the orange set of localities identified as *A. denticulata* NE are described later in this paper as *A. tehoiereensis* n. sp.; animals from the broadly-distributed dark pink set of localities identified as *A. denticulata* E + W + SW are described later in this paper as *A. kawatiriensis* n. sp. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

population comparisons and average pairwise differences between populations as large as 19.2 % (Boyer et al., 2007; Boyer and Giribet, 2007). Despite finding that many genetic divergences within *A. denticulata denticulata* were larger than some interspecific divergences among other mite harvesters, researchers found no detectable morphological differences between populations. While this previous work strongly suggests the presence of cryptic species within the *Aoraki denticulata* complex, phylogeographic analysis of mitochondrial datasets has failed to fully resolve relationships among the group's deeply divergent genetic lineages (Boyer et al., 2007, 2022; Boyer and Giribet, 2007; Fernández and Giribet, 2014).

Previous attempts to delimit cryptic species within *A. denticulata* using MSC and barcoding gap discovery methods applied to mitochondrial DNA sequence data have likewise failed, for example at its most extreme in this system, delimiting up to 74 species from ~ 34 collecting sites (Fernández & Giribet, 2014). This is unsurprising given the nature of this system; conventional MSC methods tend to oversplit species level diversity in low-dispersal taxa because such populations violate the assumption of panmixia (Sukumaran & Knowles, 2017; Derkarabetian et al., 2022b). While previous phylogeographic work and species delimitation attempts in *A. denticulata* have utilized Sanger-generated mitochondrial DNA sequence data for a small number of loci, we present here data from hundreds of loci, generated through target capture and enrichment of ultraconserved elements (UCEs), orthologous regions of the genome with remarkable conservation across divergent taxa (up to 100 % identity). Over the past decade, UCEs have been proven to have immense phylogenetic utility across various animal taxa (e.g., Faircloth et al., 2012; Branstetter et al., 2017; Starrett et. 2017; Quattrini et al., 2018; Kulkarni et al., 2020) and across taxonomic levels including species and populations (e.g., Smith et al., 2014; Derkarabetian et al., 2022a).

In this study we attempt to resolve species limits in the *A. denticulata* complex by taking an integrative taxonomic approach, combining both molecular and morphological analyses, with a focus on the use of ML approaches to analyzing phylogenomic data. We applied ML analyses to our UCE-derived dataset, using both loci and SNPs. First we use multiple traditional and unsupervised ML approaches that do not require *a priori* hypotheses (i.e., clustering) to identify putative species level divergences, which are then used as *a priori* species in a supervised ML validation analysis that is biologically informed. This approach resulted in an estimate of a number of species within the previously intractable *A. denticulata* complex that is congruent with results based on consensus of morphological diagnosability and monophyly, and can therefore be considered robust and reliable.

2. Methods

2.1. Sampling

The majority of specimens were collected by SLB and collaborators over the past two decades under permits issued by Aotearoa New Zealand's Department of Conservation Te Papa Atawhai, and are all deposited at the Harvard Museum of Comparative Zoology (Supplementary Table 1). Specimens were collected in the field by leaf litter sifting followed by direct collection and immediate preservation in 95 % ethanol. Genetic data were generated from a total of 115 individuals: 113 from *Aoraki* and one from each of the outgroup genera *Neopurcellia* and *Karripurcellia.* Previous phylogenetic analysis of transcriptomic data identified *Neopurcellia* + *Karripurcellia* as the sister group of *Aoraki* (Baker et al., 2020), while recent analysis of UCE data found that *Neopurcellia* is sister to *Aoraki* + *Karripurcellia* (Giribet et al., 2022).

2.2. Sequence capture, phylogenetic analysis, and SNPs

For all specimens, genomic DNA was extracted from either a set of appendages or a whole body using the DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA) following the manufacturer's protocol with some modifications.

UCE library preparation and hybridization largely followed standard protocols from previous Opiliones studies (e.g., Derkarabetian et al. 2019b; Derkarabetian et al., 2022b) and from ultraconserved.org. UCEs were prepared over multiple plates that only differed in library preparation protocol; the DNA was either sheared to 500 bp using a Covaris S220 Focused-ultrasonicator and prepared using the KAPA Hyper Prep Kit (Roche Sequencing and Life Science, Indianapolis, IN) at half the manufacturer's recommended reaction, or DNA was fragmented using the KAPA Hyper Plus Prep Kit with a fragmentation time of 3 min. Target enrichment was performed using the MYbaits Arachnida 1.1 K version 1 kit (Arbor Biosciences, Ann Arbor, MI) arachnid-specific probeset (Starrett et al., 2017; Faircloth, 2017). DNA was quantified with PicoGreen at the University of Minnesota Genomics Center or with a Qubit fluorometer. Illumina sequencing was performed at the Bauer Core Facility at Harvard University on an Illumina NovaSeq 6000 with 150 bp paired-end reads.

Raw data reads were processed using the Phyluce pipeline (v1.7.1, Faircloth, 2016). Reads were cleaned and adapters were removed using the Illumiprocessor wrapper (Faircloth, 2011). Contigs were assembled using ABySS (Simpson et al., 2009) and Trinity (Grabherr et al., 2011) and were matched with probes using minimum coverage and minimum identity values of 65. UCE loci were aligned using mafft (Katoh and Standley, 2013) and were trimmed using trimAI (Capella-Gutiérrez et al., 2009). Loci with 50 % taxon coverage were imported into Geneious 11.0.14.1(https://www.geneious.com; Kearse et al., 2012) for manual inspection and sequences that were determined to be non-homologous were trimmed. Specimens that were unsuccessfully sequenced were removed. A phylogeny was then constructed using RAxML version 8.2.12 (Stamatakis, 2014) using 500 bootstrap replicates with the GTR + GAMMA model on an unpartitioned, concatenated dataset. We also ran a weighted ASTRAL analysis using ASTER (Zhang amd Mirarab, 2022) where individual gene trees were estimated using IQ-TREE v.2.2 (Nguyen et al., 2015) with models determined via ModelFinder Plus (Kalyaanamoorthy et al., 2017) and 1000 ultrafast bootstrap replicates (Hoang et al., 2018). We extracted SNPs from our UCE matrix as in Derkarabetian et al. (2019a).

2.3. Mitochondrial data

A dataset was created using a combination of previously published Sanger-generated mitochondrial cytochrome oxidase subunit I (COI) sequences for *Aoraki denticulata* (detailed in Boyer et al., 2022) and sequences derived as UCE by-catch (e.g. Hedin et al., 2019) from newly sequenced samples. COI sequences were pulled from UCE assemblies via local BLAST search against published *Aoraki* COI data in Geneious 11.0.14.1. An ultrametric tree was constructed in BEAST 1.10.4 (Suchard et al., 2018) using the same parameters described in Fernández & Giribet (2014), with 10 % burnin and run for 10,000,000 generations, sampling every 1000th generation.

2.4. Traditional species delimitation

For conventional species delimitation analyses using our COI dataset, we performed one distance-based and two coalescent-based methods. First, using the ultrametric tree described previously, the Generalized Mixed Yule Coalescent (GMYC) (Pons et al., 2006) method was performed in R version 4.2.0 (R Core Team, 2020) with the packages *paran* version 1.5.2 (Dinno, 2018) and *splits* version 1.0–20 (Ezard et al., 2009) using a single threshold. Next, we used a NEXUS tree generated in Geneious in the bPTP web server (species.h-its.org/) (Zhang et al., 2013) which was run for 100,000 generations, with thinning at 100, and a burnin of 10 %. Finally, we used the Assemble Species by Automatic Partitioning (ASAP) web server using the Jukes-Cantor substitution model (Puillandre et al., 2021).

2.5. Unsupervised machine learning (UML) species delimitation

For clustering-based analyses, we used both traditional clustering approaches and novel techniques utilizing unsupervised machine learning (UML) algorithms as outlined and described in Derkarabetian et al. (2019a). For these analyses, we used our SNP dataset derived from our UCE data. First, we used the R package adegenet (Jombart, 2008) to conduct principal component analysis (PCA) and used discriminant analysis of principal components (DAPC) at default parameters to determine the optimal number of genetic clusters.

We utilized three UML algorithms to aid in species delimitation: Random Forest (RF; Breiman, 2001), t-distributed Stochastic Neighbour Embedding (t-SNE; van der Maaten and Hinton, 2008), and Variational Autoencoders (VAE; Kingma and Welling, 2013). These algorithms as unsupervised approaches are essentially dimensionality reduction, allowing high dimensional data (i.e.,100 s of SNPs) to be reduced and visualized as a low-dimensional representation, in this case, two-dimensional. From these two-dimensional representations, we then performed hierarchical clustering using the mclust R package for the RF and t-SNE analyses (Scrucca et al., 2016). The RF and t-SNE approaches were implemented in R using a modified script from Derkarabetian et al. (2019a) (https://github.com/shahanderkarabetian/UML-Tutorial) using default settings. Parameters used are discussed in Derkarabetian et al. (2019a). The VAE was run using default settings as in Derkarebetian et al. (2019a) using the "sp_deli" script (https://github.com/sokrypton/sp_deli). The VAE was run five times, with the favored representation being the run with the lowest average loss score after the first 50 % of epochs (generations) were removed as "burnin".

2.6. Supervised machine learning (SML) validation

We also ran a supervised machine learning (SML) approach using CLADES (Pei et al., 2018). We created a dataset that included only samples within the *A. denticulata* complex (*A. denitculata denticulata* + *A. denticulata major* + *A. longitarsa* + *A. meridialis*) and only included the loci that had sequences from at least 59/63 samples. Our *a priori* species hypotheses were based on the clustering results from the UML approaches. We ran this dataset against two training datasets, the "general" training set from Pei et al (2018) and the "custom" training dataset created in Derkarabetian et al., (2022b), which is based on UCE data from the Opiliones genus *Metanonychus* (Derkarabetian et al., 2019a). As detailed in Derkarabetian et al., (2022b), the custom training set was created because the focal taxon of that study shares similar biological and ecological characteristics with other Opiliones (low dispersal ability, high ecological constraints, high population genetic structure), leading to similar underlying genetic structure and patterns at the population and species levels. As in Derkarabetian et al., (2022b), we applied this training dataset to *Aoraki* here, as the biological characteristics between *Aoraki* and *Metanonychus* are more similar to each other than they are to the the vast majority of other taxa, which is represented by the "general" training set of Pei et al. (2018) that was created to represent general patterns across animals.

2.7. Morphology

All specimens were initially observed using light microscopy. In Cyphophthalmi, only males possess characteristics diagnostic of species; therefore, selected male individuals were mounted, coated with gold–palladium alloy, and imaged with a JEOL 6610 LV Scanning Electron Microscope (SEM). We also re-examined all SEM images taken for previous studies of *Aoraki* (Boyer et al., 2007, 2022; Boyer and Giribet, 2007). We made qualitative observations of features known to distinguish species: the shape of tergite VIII, the arrangement of scopular hairs emerging from the anal plate, and the shape of tarsus IV and its adenostyle. In order to make quantitative observations of external morphology, we chose 65 male individuals from across the geographic range of *A. denticulata* and measured the length and width of the prosoma, and also the space between the ventral-most point of the two tubercles formed by the pointed lobes of tergite VIII. Measurements were taken either with JEOL JSM-6610 software, or using a Minitool microscale ruler under an Olympus SZX10 light microscope. To test for statistically significant differences between the cryptic species described here, we performed Hotelling's T-squared distribution test using the Hotelling package in R (Curran, 2021). We also dissected the spermatopositor (male reproductive organ) from one to three individuals from each of the five putative species and compared the number and positioning of microtrichi using an Olympus BX60 light microscope.

3. Results

3.1. Phylogenetic analysis

Raw reads are available from the Sequence Read Archive under BioProject PRJNA1051163. The total number of UCEs in our 50 % matrix was 847 with a maximum alignment length of 270,347 bp. Higher-level phylogenetic relationships retrieved using both RAxML (Fig. 2) and ASTRAL (Fig. S1) reflect previous results of analysis of mitochondrial data (Boyer et al., 2022), which identified three major lineages within *Aoraki:* species that have an entirely smooth male tarsus IV (*A. crypta* + *A. grandis* + *A. healyi* + *A. inerma*), species that have a heavily granulose male tarsus IV (*A. tumidata* + *A. westlandica*), and species that have a lightly granulose male tarsus IV (the *A. denticulata* complex: *A. denticulata denticulata* + *A. denticulata major* + *A. longitarsa* + *A. meridialis*).

Most morphologically distinctive species of *Aoraki* were retrieved as monophyletic groups with 100 % support. Within the *A. denticulata* complex, we confirm the monophyly of the three morphologically distinct lineages: *A. longitarsa*, *A. meridialis*, and *A. denticulata major*. *A. denticulata denticulata* is paraphyletic, consisting of three different geographically distinct lineages (Fig. 2). *A. longitarsa* and *A. meridialis* form a clade sister to a lineage comprised of *A. denticulata denticulata* E + W + SW. *A. denticulata major* is sister to *A. denticulata denticulata* NE, and those two form a clade sister to *A. denticulata denticulata* NW (Fig. 2). This topology is similar to but not fully congruent with the findings of Fernández and Giribet (2014): their analysis of mtDNA identified a widespread "Northern" lineage that is paraphyletic in our results; the placement of *A. longitarsa* differs in our results; and their analysis did not include *A. meridialis,* which was discovered and described after the publication of their study. Outside of the *A. denticulata* complex, we find that *A. inerma* (Forster, 1952) is paraphyletic, with *A. crypta* (Forster, 1952) nested within it; genetic divergences within the *A. crypta* + *A. inerma* lineage are comparable to what is seen within other morphologically uniform species (e.g. *A. tumidata, A. healyi, A. grandis*) (Fig. 2).

3.2. Phylogenomic species delimitation analyses

GMYC and bPTP analyses supported 29 species: two within *A. denticulata major,* one within *A. longitarsa,* one within *A. meridialis* and 24 within *A. denticulata denticulata.* The ASAP analysis yielded fewer species: one species within *A. denticulata major, A. longitarsa,* and *A. meridialis* and 19 within *A. denticulata denticulata*. Our standard clustering analyses using PCA and DAPC recovered an optimal of K = 8: one species each for *A. meridialis, A. longitarsa,* and *A. denticulata major* and five geographically distinctive species within *A. denticulata denticulata* which we term the Northwest (NW), Northeast (NE), East (E), West (W), and Southwest (SW) lineages (Figs. 1, 2, 3).

Among our other UML analyses, VAE and t-SNE both supported an optimum of K = 8 with five clusters within *A. denticulata denticulata* while RF supported an optimum of K = 10 with seven clusters within *A. denticulata denticulata.* Supervised ML analyses using the "general" training data set favored eight species within the *A. denticulata* complex,

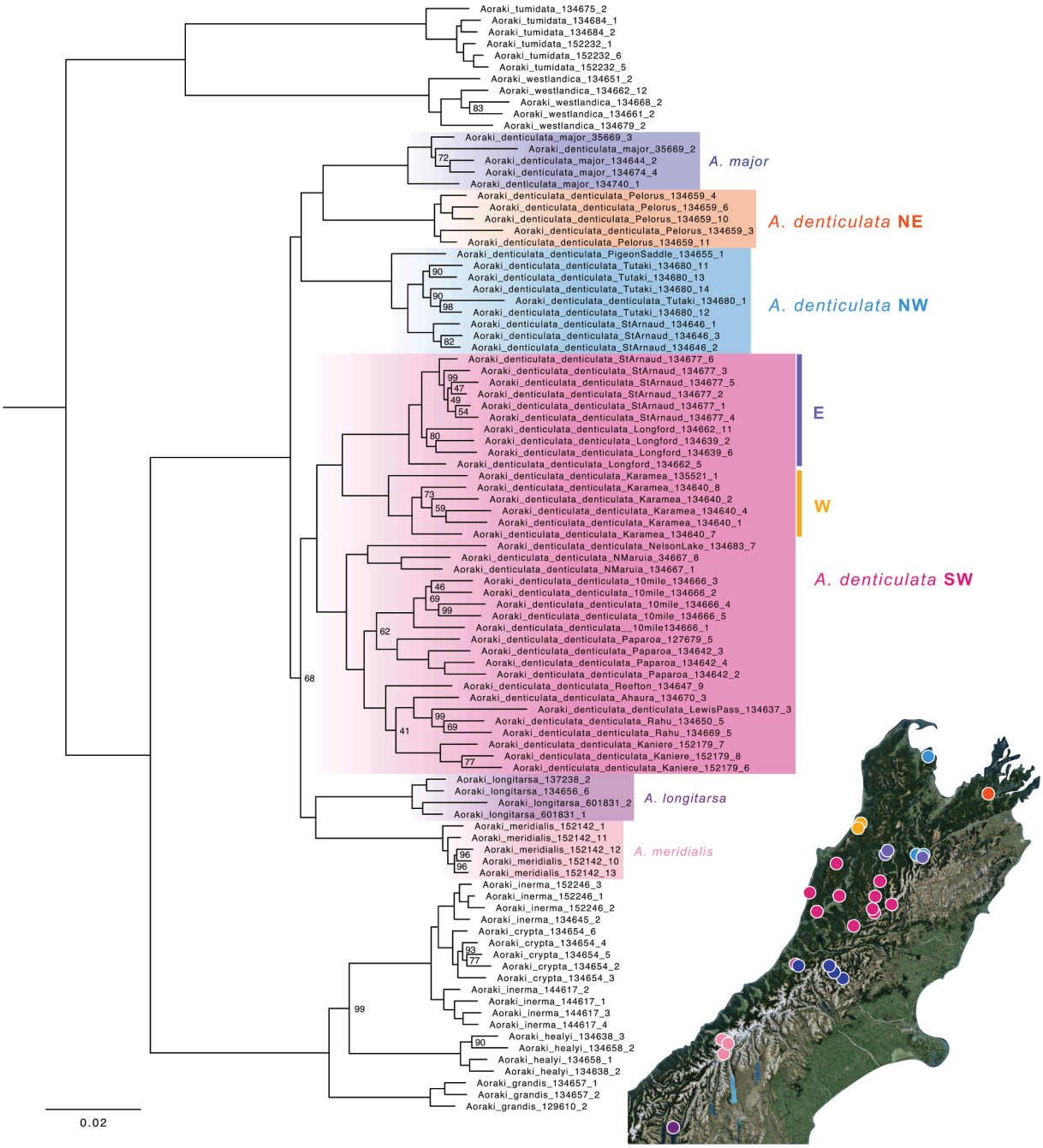

Fig. 2. Result of phylogenetic analysis of UCE-derived genomic data in RAxML. Bootstrap support is 100 % unless otherwise indicated. The orange lineage identified as *A. denticulata* NE is described later in this paper as *A. tehoiereensis* n. sp.; the broadly-distributed dark pink lineage lineage identified as *A. denticulata* E + W + SW is described later in this paper as *A. kawatiriensis* n. sp. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in agreement with standard clustering and UML analyses. The analysis using the "custom" training data set favored six species, lumping the E, W, and SW lineages of *A. denticulata*, reflecting the expected hypotheses for the minimum number of morphologically distinctive monophyletic species (Fig. 2).

3.3. Morphological analyses

Results from Hotelling's T-squared distribution test found statistically significant differences between the SW population and the E, NE, and NW populations. ANOVA analyses found a significant p-value for length measurements between populations (<0.005) but not for width measurements. Despite statistically significant differences in size between populations, there is considerable overlap in length and width measurements for all populations, making diagnosis on the basis of length or width impossible. Qualitative analyses of external morphology, including that of the adenostyle, anal plate, scopular hairs, and tubercles of tergite VIII, found no diagnosable features between putative cryptic species (Fig. 4). Additionally, all spermatopositors examined had the same number of microtrichii, meaning genitalia can not be used in the diagnosis of these cryptic species.

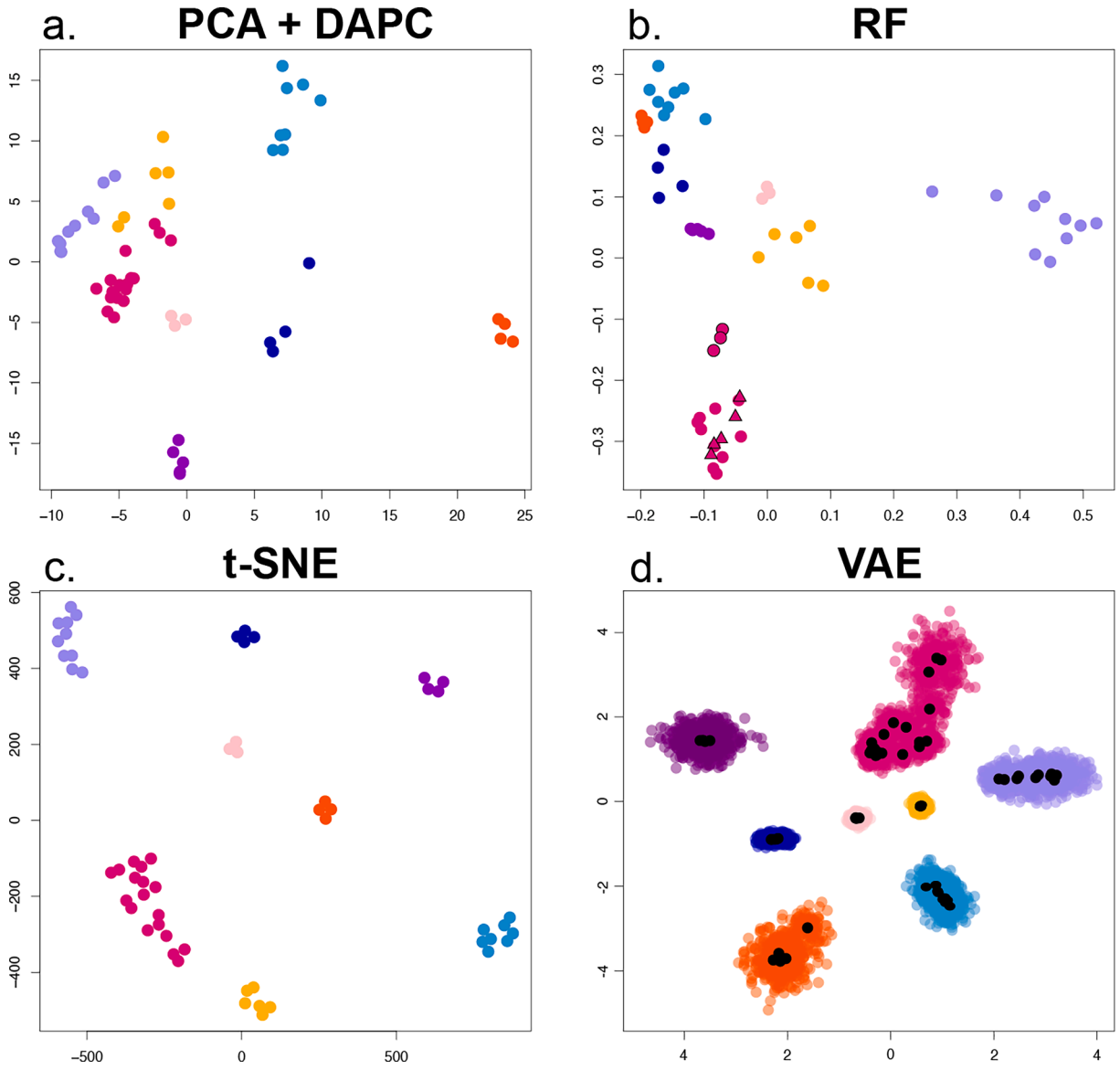

Fig. 3. Results of unsupervised machine learning analyses. a) PCA + DAPC, b) Random Forest, c) t-SNE, d) Variational Auto-Encoder. Colors correspond to mapped localities in Fig. 1 and branches on phylogeny in Fig. 2.

4. Discussion

4.1. Species delimitation in the Aoraki denticulata complex

Previous authors have remarked that a thorough study of the morphology of the *Aoraki denticulata* complex might yield previously unnoticed characteristics diagnostic of species (Boyer et al., 2007; Boyer and Giribet, 2007). Somewhat to our surprise, we did not discover any such characteristics of the course of this study; therefore, the criterion of morphological diagnosability suggests that only four species exist in the complex: *A. major, A. denticulata, A. meridialis,* and *A. longitarsa*. However, the fully-resolved phylogeny that we were able to derive from our UCE dataset confirmed the presence of cryptic species: it is clear that the morphologically distinct entity named *Aoraki denticulata* consists of three monophyletic groups, none of which is sister to the other (Fig. 2). Additionally, the three *A. denticulata* lineages occupy non-overlapping geography (Fig. 1). Therefore, integration of morphological, phylogenetic, and geographic information suggests six species in the *A. denticulata* complex.

In order to further explore species boundaries in *Aoraki denticulata,* we turned to results of delimitation analyses. As seen in the previous implementation of the SML approach in a similar context (Derkarabetian et al., 2022b), the "custom" training data set, derived from a taxon with similar biological and ecological characteristics to the test taxon, resulted in a more conservative estimate of species numbers than did SML using a "general" training data set (Fig. 5). This resulting hypothesis, which splits *Aoraki denticulata denticulata* into three cryptic species, reflects the minimum number of species identified through integration of morphology, phylogeny, and geography, but is an underestimate compared to all other delimitation analyses, including the UML approaches which favored splitting *A. denticulata denticulata* into five cryptic species. The difference in numbers is due to the lumping of *A. denticulata denticulata* E + W + SW lineages (Fig. 2). Considered as a single unit, this group of lineages is geographically more widespread than any other species identified in our analyses. From a topological perspective, this group shows deeper branches among lineages compared to the other species, which included both morphologically diagnosable species (*A. denticulata major*, *A. longitarsa*, *A. meridialis*) and cryptic species (*A. denticulata denticulata* Northwest and Northeast), that have longer branches leading to the population-level divergences

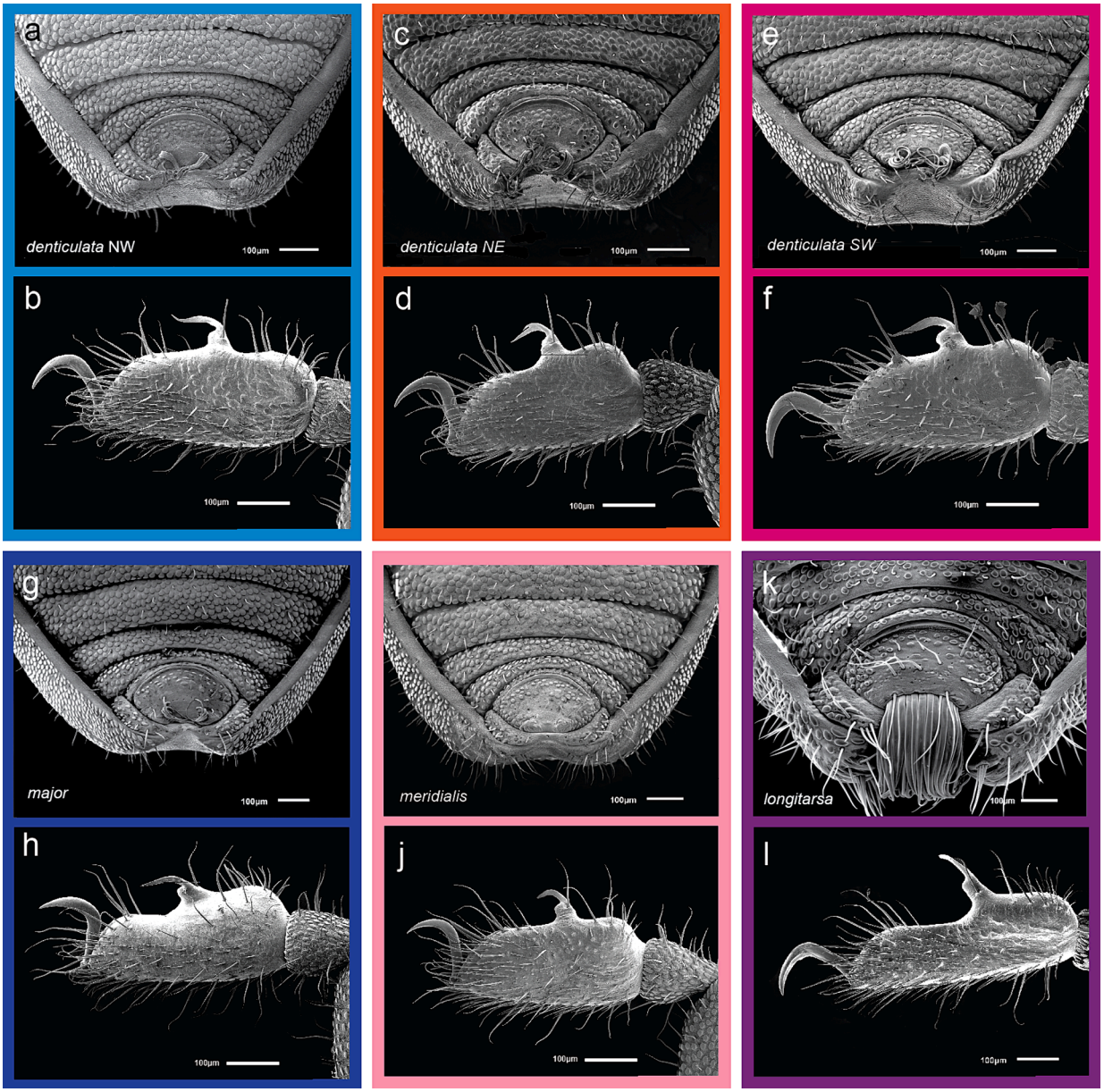

Fig. 4. Comparison of anatomy of the fourth tarsus and ventral posterior regions of morphologically distinctive species and cryptic species within the *A. denticulata* complex. For each, we give the MCZ Invertebrate Zoology Collection number as well as the SEM stub number. a) *Aoraki denticulata denticulata* NW IZ-134680 (M43-1), b) *A. denticulata denticulata* NE IZ-134680 (M43-8), c) *A. denticulata denticulata* NE IZ134659 (18–12), d) *A. denticulata denticulata* NE IZ134659 (18–11), e) *A. denticulata denticulata* SW IZ-134669 (M45-1), f) *A. denticulata denticulata* SW IZ-134650 (11–4), g) *A. denticulata major* IZ-152201 (M45-9), h) *A. denticulata major* IZ-127676 (M43-11), i) *A. meridialis* IZ-152142 (M35-6), j) *A. meridialis* IZ-152142 (M35-7), k) *A. longitarsa* IZ-134656 (15–5), l) *A. longitarsa* IZ-134656 (15–4).

(Fig. 2).

Previous attempts to delimit species within the *A. denticulata* complex based on phylogeographic analysis of small numbers of Sanger-sequenced loci failed due to lack of well-supported resolution among lineages. Most recently, Boyer et al. (2022) analyzed COI and 16S data and found strong support for the monophyly of the *A. denticulata* complex and for *A. denticulata major, A. longitarsa,* and *A. meridialis*. However, little resolution was recovered among lineages of *A. denticulata denticulata*. Fernández and Giribet (2014) performed a phylogeographic analysis of the complex using COI and 16S as well as the nuclear loci 28S and H3. They identified three lineages of interest within the species complex, including a well supported group (termed the "Alps" clade in their work) within which the NE *A. denticulata denticulata* lineage is sister to *A. denticulata major,* as found in the current study (Fig. 2). However, they found that the lineages we call E, W, and NW form a monophyletic group (termed the "Northern" lineage) in both maximum likelihood and Bayesian analyses (albeit with < 70 % posterior probability). Their Bayesian analysis also identified a "Southern" lineage, which corresponds to our SW lineage with *A. longitarsa* nested within; however, it was not recovered as monophyletic in their ML analysis. The dataset presented in the current study is orders of magnitude larger than those developed by previous authors, and as such has the power to solve the "evolutionary puzzle" (per Fernández and Giribet) that stymied those researchers.

4.2. Machine learning in species delimitation

We demonstrated the successful application of machine learning approaches to species delimitation, specifically applying these methods to a historically recalcitrant species complex that includes cryptic

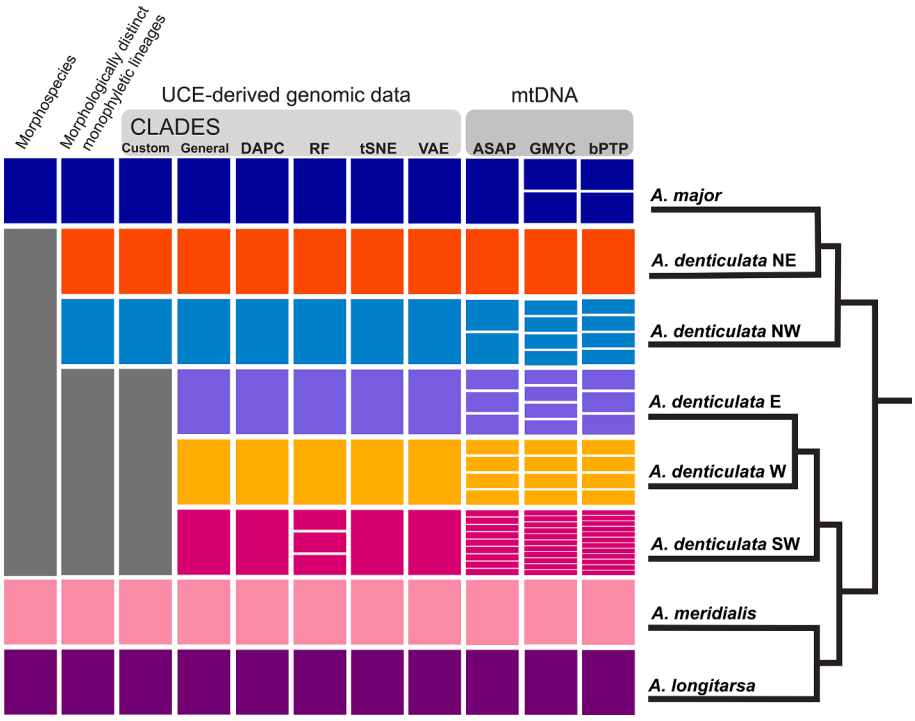

Fig. 5. Summary of results of different delimitation analyses, grouped by approach and data type.

species. In particular, the supervised approach using a more relevant, biologically informed custom training data set provided a more conservative number of species relative to the unsupervised clustering approaches and a more general training data set - a result consistent with the number of species identified through integration of morphology and phylogeny. While the putative eight species identified in unsupervised approaches perhaps each represent distinct and separately evolving lineages or metapopulations (de Quieroz 2007), we favor the six species hypothesis validated by the supervised machine learning approach. In the context of species delimitation, and its downstream applications to taxonomy and conservation, it is best to take an integrative approach that seeks congruence across multiple data types and analytical approaches. This leads to conservatism in terms of the number of species that are identified and delimited, but this is preferable to incorrectly delimiting species that may not exist (Carstens et al., 2013). The six-species hypothesis reflects the minimum number of monophyletic species identified through integration of morphology and phylogenetic structure (i.e., morphological species criteria and phylogenetic species criteria). Because it reflects a consensus among multiple data types and analytical approaches, this hypothesis can be considered robust.

Determining the line at which to draw species boundaries can be problematic in species complexes that possess the biological and ecological characteristics that force delimitation to rely solely or largely on genetic data (i.e., cryptic species). It is clear in these systems that species numbers are underestimated, but genetic data does not easily or consistently provide a realistic result, and many commonly used approaches fail by overestimating species numbers (e.g. Hedin et al., 2015; Derkarabetian et al., 2022b). Unsupervised ML approaches first demonstrated in Opiliones (Derkarabetian et al., 2019a) have already shown great utility in species delimitation across a diverse array of taxa including both invertebrates, like arachnids (Hedin et al., 2020; Newton et al., 2020; Giribet et al., 2021) and mollusks (Moles et al., 2021), and vertebrate taxa (Mussmann et al., 2020; Martin et al., 2021; Chan et al., 2022; DeRaad et al., 2022; Obiol et al., 2023). Other unsupervised approaches have also been incorporated in species delimitation very recently, like the use of Self-Organizing maps, including an integrative version that can combine multiple data types (Pyron et al., 2023; Pyron, 2024). These approaches also demonstrated an ability to help alleviate the issues of oversplitting with genetic data. The use of supervised ML with a custom training data set, here implemented through the program CLADES, is much more recent, but is certainly a promising avenue to incorporate organismal and taxonomic expertise into the genetic species delimitation pipeline. Here, this involves creating custom training data sets derived from related taxa with similar biological and ecological characteristics, hence similar modes of speciation, theoretically leading to similar underlying genetic patterns across the population–species boundary. The success of the supervised ML approach here underscores the importance of considering natural history and creating more biologically informed methods for delimiting species. The agreement seen in our study between the more traditional approaches and the newer machine learning genetic species delimitation analyses is a promising sign for these algorithms in species delimitation. Previous genetic analyses erratically and drastically oversplit species in this system; however, our machine learning analyses provided more conservative estimates based purely on genetic data. Moving forward, these results could be deemed more trustworthy in similar systems where other data is limited or non-existent, such as other cryptic species complexes.

4.3. Future directions

Our results increase the number of *Aoraki* species known from the South Island of Aotearoa New Zealand and clarify the geographic distribution of diversity within the genus. The South Island is bisected north to south by an area of low species richness and high biotic turnover known as the Beech Gap due to a paucity of the iconic beech trees that make up more than half of New Zealand's standing native forests. The Beech Gap roughly corresponds to the narrow "waist" of the island (Fig. 1), and is hypothesized to have been established during the Last Glacial Maximum (LGM) approximately 20 ka (Marske and Boyer 2022 and references therein). We currently know of eleven *Aoraki* species from the South Island, with only two of those known from south of the Beech Gap - a pattern that may have been driven by the presence of the exceptionally large forest refugium that persisted north of the Beech Gap through the LGM (Newnham et al. 2013).

Beyond *Aoraki,* the archipelago is home to two additional cyphophthalmid genera: *Neopurcellia* Forster, 1948, and *Rakaia* Hirst 1925. *Neopurcellia* is a monotypic genus that encompasses two genetically distinct lineages that occupy non-overlapping geographic ranges (Tardelli Canedo et al., 2021). *Rakaia,* the geographically most widespread and morphologically most diverse of the genera (Boyer & Giribet, 2009), also includes phylogenetically distinct lineages that are morphologically cryptic (Morisawa, 2020). The approach we have used to understand species boundaries in *Aoraki* could be applied to both of these genera in order to fully identify and map all species of mite harvesters from across Aotearoa New Zealand.

Our results highlight the rarity of species co-occurrence among closely related low-dispersal species. While there is close geographic proximity (e.g., within 1–2 km) between genetic lineages of *A. denticulata*, none of the lineages that are validated as species by SML co-occur at a single locality. Indeed, across more than one thousand collections of Cyphophthalmi from across New Zealand, there are only 34 recorded instances of syntopy (i.e. collection of multiple species at the same geographic point and at the same time) (Shu et al., 2023). Similarly, within the well-studied mite harvester fauna of the Australian Wet Tropics (e.g. Jay et al., 2016, Oberski et al., 2018), only a handful of the hundreds of collections include syntopic species. In approximately half of the cases of syntopy known from Aotearoa New Zealand, coexisting species are members of different genera; syntopic congeneric species are sometimes sister taxa and other times more distantly related. These patterns could (for example) reflect niche conservatism combined with competition, or active avoidance of congeners that is selected through reinforcement. Fully documenting such biogeographic patterns and developing hypotheses to explain them is not possible without reliable delimitation of species boundaries.

Students of biogeography know that biodiversity is not distributed evenly in space at scales from local to global; studies that seek to understand the mechanisms that drive biogeographic patterns should rely on robust and repeatable methodologies for identifying biologically relevant and comparable units of biodiversity. Additionally, different organisms with varying biological characteristics, such as dispersal ability, may show differing biogeographic patterns across the same geographic space. For example in Gondwanan taxa the phylogenetic complexity of biogeographic patterns may be correlated with dispersal ability and microhabitat constraints (e.g., Derkarabetian et al., 2021). Taxa with the lowest dispersal ability and highest microhabitat specificity (leading to morphological and ecological conservatism) also tend to be taxa with potential cryptic species (Czekanski-Moir & Rundell, 2019). In systems that include cryptic species, integrative delimitation that draws on maximally informed quantitative methods constitutes an ideal approach.

5. Taxonomy

Family Pettalidae Shear, 1980.
Genus *Aoraki* Boyer et al., 2007; Boyer and Giribet, 2007.

- *Aoraki crypta* (Forster, 1948) new synonymy

We found that *Aoraki inerma* is paraphyletic, with *A. crypta* nested within the species, confirming suspicions raised by previous analysis of mitochondrial data and detailed examination of morphology (Boyer et al., 2022). The two do not differ in either tarsus IV or the male anal plate, associated tergites, and scopulae (Figures 13 and 14 in Boyer et al., 2022), and were distinguished by Forster only by the degree of ventral curvature of the body, a character that may be affected by collection and preservation conditions. In addition, *A. crypta* is known only from Mt. Te Aroha, a locality that falls within the range of *A. inerma*, which is widespread across the northern half of the North Island (Fig. 2 in Boyer et al., 2022). Forster described the two species in the same publication with *A. crypta* appearing earlier in the text; *A. inerma* is therefore considered a junior synonym of *A. crypta* **new synonymy.**

- *Aoraki major* (Forster, 1948) new combination

Forster distinguished *A. denticulata major* from *A. denticulata denticulata* by size alone. He reported the length of the *A. denticulata major* holotype male as 2.63 mm, and width at widest point 1.50 mm. Detailed examination of collections from across both described subspecies confirms that males in the well-supported *A. denticulata major* clade are consistently $>$ 2.3 mm in length, while males of all other *A. denticulata* lineages are $<$ 2.3 mm in length. In a recent catalogue of Cyphophthalmi Giribet (2020) elevated many other subspecies from Aotearoa New Zealand to species status but refrained from taking this action with *A. denticulata major* due to the lack of phylogenetic resolution within the *A. denticulata* species complex. Here we elevate *A. denticulata* major to *A. major* **new combination**.

- *Aoraki denticulata* (Forster, 1948) new combination

Type material: Holotype: Male (MONZ vial DM 2/53 and slide 4/13), Starvation Ridge, Nelson, South Island, New Zealand. Paratypes: Males, females (MONZ vial DM 2/54, 2/99).

Diagnosis: Males are differentiated from all other Aotearoa New Zealand Cyphophthalmi by the absence of a ventral process on the trochanter of the palp (characteristic of *Aoraki*); the femora of all legs at least twice as long as wide; the anal plate not indented and bearing two small scopulae originating from the ventral surface; male tarsus IV broader proximally with a pronounced rounded protuberance at the base of the spur of the adenostyle; and a body length of $<$ 2.3 mm. *Aoraki denticulata* is the only one of the three cryptic species (*A. denticulata, A. tehoeiriensis, A. kawatiriensis*) found in Abel Tasman National Park or Tutaki, and can also be found in St. Arnaud near lake Rotoiti. Within COI, a site change of C at position 560 and A at position 404 (rather than C) is diagnostic (position corresponds to COI alignment available on doi.org/10.7910/DVN/M1703Y).

Description: Forster (1948) described *A. denticulata* this way (figure numbers Forster's): "Colour: Body dark chocolate-brown, appendages light yellow–brown. Body: Entire surface closely and uniformly granulate. Carapace widening behind stink gland mounds, as wide as abdomen. Stink gland mounds slightly wider than high, bluntly conical and directed slightly back, set a little more than their diameter from the lateral margin of the carapace and almost six diameters apart. Tergites separated by faint straight transverse grooves; median longitudinal groove faint, but deepening on tergites VII and VIII. Tergite VIII not modified to form a pair of tubercles, but with a broad median groove. A pair of scapulae arise from the anal plate as in Fig. 57. Corona analis as in Fig. 57. Posterior portion of the abdomen flexed down. Arculi genitales present as ovoid swellings of the inner margin of coxae IV, which are contiguous anterior to the genital opening. Stomotheca longer than wide in the ratio of 14:11. Chelicerae: As in Fig. 64. Basal segment uniformly granulate; transverse ridge sharp and directed back; ventral swelling well developed.

Second segment smooth and slender. Teeth on inner margins of fingers as in Fig. 65. Pedipalps: As in Fig. 63. Ventral process absent from trochanter, but a number of disto-dorsal and disto-ventral denticulations are present. Femur sparsely denticulate. Pedipalp otherwise smooth. Legs: All segments except tarsi uniformly granulate. Tarsus IV as in Figs. 60, 61, wider proximally and narrowing to the spur, which is present as a strong, forwardly directed, spine; the cuticle on the anterior portion at the base of the spur is distended to form a rounded tubercle.

The duct of the tarsal gland opens a little way above this tubercle on the under surface of the spur." Forster reported the length of the holotype male as 2.00 and width at widest point 1.12.

Other Material Examined: Coordinates for all specimens examined for all species: see Supplementary Table 1. Here we list only those

specimens examined for *A. denticulata*. All specimens were collected in the Tasman District in South Island, New Zealand. IZ-134680 Tutaki Conservation Area, Nelson Lakes National Park, east of Gowanbridge, −41.826769 172.47633, 27 February 2011, A. Schomann, J. Pedersen, 77 specimens including males, females, and juveniles; IZ-134655, Pigeon Saddle, Abel Tasman National Park, along Totaranui Rd, −40.8325790148 172.9689690191, 3 July 2004, S. L. Boyer, G. Giribet, 2 specimens; IZ-134646, St. Arnaud, Rotoiti Lake, Brunner Peninsula, −41.8081239983 172.8336219862, 30 Jan 2003, S. L. Boyer, G. Giribet, 5 specimens. Species range is represented in Figs. 1 and 2 as *Aoraki denticulata* - NW.

Comments: This species retains Forster's *denticulata* specific epithet because the type locality (Starvation Ridge, Nelson) lies within the geographic range.

- *Aoraki tehoiereensis* Heine & Boyer, new species

Type material: Holotype male from Pelorus Bridge Scenic Reserve, Marlborough, New Zealand, −41.297725°, 173.571684°, 19 January 2006, by J. M. Baker and S. L. Boyer, Harvard MCZ IZ-167485. Paratype males and females with identical collecting information, MCZ IZ-134659.

Etymology: The name of the species is derived from the Maori name for Pelorus River, Te Hoiere, which in turn refers to the name of the voyaging canoe (waka) used by Matua Hautere during a long journey up the river.

Diagnosis: Morphologically indistinguishable from *Aoraki denticulata* (Forster, 1948). Geography is diagnostic, as *A. tehoiereensis* is known from only one collecting locality at the Pelorus Bridge Reserve and it does not exist sympatrically with any other of the cryptic species identified here. Within COI, a site change of C (rather than A or G) at position 814 is diagnostic (position corresponds to COI alignment available at https://doi.org/10.7910/DVN/M1703Y).

Description: Follows the description of *A. denticulata* (Forster, 1948). Color: Body dark chocolate-brown, appendages light yellow–brown. Body (Supplementary Figures 4, 5, 6): Entire surface closely and uniformly granulate. Length of male holotype is 2.0 mm and width at widest point is 1.1 mm. Carapace widening behind ozophores, as wide as abdomen. Ozophores slightly wider than high, bluntly conical and directed slightly back, set a little more than their diameter from the lateral margin of the carapace and almost six diameters apart. Tergites separated by faint straight transverse grooves; median longitudinal groove faint, but deepening on tergites VII and VIII. Tergite VIII not modified to form a pair of tubercles, but with a broad median groove. A pair of scopulae arise from the anal plate. Posterior portion of the abdomen flexed down. Stomotheca longer than wide. Ventral process absent from trochanter of palp. Femur sparsely denticulate. Palp otherwise smooth. Legs: All segments except tarsi uniformly granulate. Tarsus IV wider proximally and narrowing to the adenostyle, which is present as a strong, forwardly directed spine; the cuticle on the anterior portion at the base of the spur is distended to form a rounded tubercle. The duct of the tarsal gland opens a little way above this tubercle on the under surface of the spur. (Supplementary Fig. 2).

Other Material Examined: There is a single known collection of this species. Locality is represented in Figs. 1 and 2 as *Aoraki denticulata* - NE.

- *Aoraki kawatiriensis* Heine & Boyer, new species

Type material: Holotype male from Riwaka River Scenic Reserve along Kawatiri river, Tasman, New Zealand, −41.7838709615°, 172.3690329585°, 22 January 2006, by J. M. Baker, S. L. Boyer, Harvard MCZ IZ-167486. Paratype males and females with identical collecting information, MCZ IZ-134662,

Etymology: The name refers to the Kawatiri River (also known as the Buller River), which flows through the range of the species. The name is derived from the Maori words for "deep and swift," effectively describing the river with the greatest flood discharge in Aotearoa New Zealand. **Diagnosis**: Morphology same as for *Aoraki denticulata* (Forster, 1948). Within COI, position 404 is C (rather than A as in *A. denticulata*) and position 814 is A or G (as opposed to C as in *A. tehoiereensis*) (position corresponds to COI alignment available at https://doi.org/10.7910/DVN/M1703Y). *A. kawatiriensis* does not exist sympatrically with any other of its cryptic species. Coordinates for known collecting localities can be found in Supplementary Table 1.

Description: Color: Body dark chocolate-brown, appendages light yellow–brown. Body: Entire surface closely and uniformly granulate. Length of male holotype is 2.0 mm and width at widest point is 1.1 mm. Carapace widening behind ozophores, as wide as abdomen. Ozophores slightly wider than high, bluntly conical and directed slightly back, set a little more than their diameter from the lateral margin of the carapace and almost six diameters apart. Tergites separated by faint straight transverse grooves; median longitudinal groove faint, but deepening on tergites VII and VIII. Tergite VIII not modified to form a pair of tubercles, but with a broad median groove. A pair of scopulae arise from the anal plate as in Fig. 57. Posterior portion of the abdomen flexed down. Stomotheca longer than wide. Ventral process absent from trochanter of palp. Femur sparsely denticulate. Palp otherwise smooth. Legs: All segments except tarsi uniformly granulate. Tarsus IV wider proximally and narrowing to the adenostyle, which is present as a strong, forwardly directed spine; the cuticle on the anterior portion at the base of the spur is distended to form a rounded tubercle. The duct of the tarsal gland opens a little way above this tubercle on the under surface of the spur. (Supplementary Fig. 3).

Other Material Examined: Here we list only those specimens examined for *A. kawatiriensis*. All specimens were collected in South Island, New Zealand. Tasman: IZ-134677, East of St. Arnaud, off Route 63, −41.803061 172.845780, 25 January 2006, S. L. Boyer, J. M. Baker (7 males, 12 females); IZ-134639, Riwaka River Scenic Reserve near Longford, along Kawatiri river, −41.783870960 172.3690329585, 30 Jan 2003, S. L. Boyer, C. D'Haese, G. Giribet, 21 specimens. IZ-134683, Warbeck Scenic Reserve, near Maruia Saddle, −42.032969 172.293518, 27 February 2011, A. Schomann, J. Pedersen (25 specimens) West Coast: IZ-135521, South Terrace, Karamea, −41.272778 172.13325, 19 December 2009, S. Vélez (1 female); IZ-134640, Karamea Bluffs near summit, −41.51919703 172.0225239918, 1 February 2003, S. L. Boyer, C. D'Haese, G. Giribet (20 specimens); IZ-134669, Rahu Scenic Reserve, −42.3322530370 172.1712720115, 26 January 2006, S. L. Boyer, J. M. Baker (5 males, 13 females, 3 juveniles); IZ-134667, North of Maruia, Hwy 65, −42.188781172.22034, 26 January 2006, S. L. Boyer, J. M. Baker (11 males, 8 females); IZ-134670, Ahaura Kopara Rd, −42.481138 171.834264, 27 January 2006, S. L. Boyer, J. M. Baker (2 males, 2 juveniles); IZ-127679, Paparoa Range NW of Blackball, Croesus Track near Garden Gully Hut, −42. 29,163,333 171.459167, 11–12 November 2011, A. Solodovnikov, L. Vilhelmsen (11 specimens); IZ-134642, Truman Track, Paparoa National Park, −42.0928269904, 171.3409659639, 2 February 2003, S. L. Boyer, C. D'Haese, G. Giribet (4 specimens); IZ-134666, 10 Mile Creek, Hwy 6, −41.835833 171.676528, 24 January 2006, S. L. Boyer, J. M. Baker (4 males, 5 females); IZ-134637, Lewis Pass Rd. between Rough Creek and Jackson Creek, −42.3824444 172. 2888889, 26 February 2011, A. Schomann, J. Pedersen (7 males, 3 females); IZ-134647, Approx. 10 km SW of Reefton, −42.1563919913 171.7926470283, 30 Jan 2003, S. L. Boyer, C. D'Haese, G. Giribet (23 specimens); IZ-134650, Rahu Scenic Reserve, −42.3322530370 172.1712720115, 31 January 2003 S. L. Boyer, C. D'Haese, G. Giribet (8 specimens); IZ-152179, Dorothy Creek Walk, Lake Kaniere, −42.844667 171.16475, 18 January 2019, C. M. Baker, S. L. Boyer, R. Morisawa, E. Pessereau, P. Tardelli Canedo (19 males, 29 females, 3 juveniles). Species range is represented in Figs. 1 and 2 as *Aoraki denticulata* - E + W + SW.

CRediT authorship contribution statement

Haley L.A. Heine: Writing – review & editing, Writing – original draft, Visualization, Formal analysis. **Shahan Derkarabetian:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Rina Morisawa:** Investigation, Formal analysis, Data curation. **Phoebe A. Fu:** Investigation, Formal analysis, Data curation. **Nathaniel H.W. Moyes:** Investigation, Formal analysis, Data curation. **Sarah L. Boyer:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data are available at https://doi.org/10.7910/DVN/M1703Y.

Acknowledgements

We thank the New Zealand Department of Conservation Te Papa Atawhai for their support around logistics and permitting over the past 20 years of work on Aotearoa New Zealand Opiliones by SLB and collaborators. Gonzalo Giribet has provided invaluable perspective and support throughout these endeavors, including hosting some of the work of this study in his lab. Prashant Sharma took many of the SEM images examined for this study. Newly-collected specimens were obtained in 2019 with help from Caitlin Baker, Eliza Pessereau, and Pietro Tardelli Canedo. Analysis of a preliminary UCE dataset for this study was performed by Ciara Moore and Maddie Tuffield as part of their work in the Spring 2020 iteration of Biology 476: Research in Biodiversity and Evolution at Macalester College. Anatomical measurements and qualitative comparisons of morphology were performed with help from Boyer Lab members Sophia Anderson and Yoonjin Shu. We graciously thank Rodrigo Monjaraz-Ruedas for help with ASTRAL analyses. Shanta Hejmadi served on the undergraduate honors thesis committee of HLAH and provided helpful feedback. For help with loans of specimens we thank Phil Sirvid (Museum of New Zealand Te Papa Tongarewa) and Adam Baldinger (Harvard Museum of Comparative Zoology), who also provided outstanding curatorial support for this project. Three anonymous reviewers provided helpful feedback that considerably improved our work. This work was supported by Macalester College, the Arnold and Mabel Beckman Foundation, and US National Science Foundation grant DEB-1732643 to SLB.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2024.108061.

References

Agnarsson, I., LeQuier, S.M., Kuntner, M., Cheng, R.-C., Coddington, J.A., Binford, G., 2016. Phylogeography of a good Caribbean disperser: Argiope argentata (Araneae, Araneidae) and a new 'cryptic' species from Cuba. Zookeys 25–44.

Arthofer, W., Rauch, H., Thaler-Knoflach, B., Moder, K., Muster, C., Schlick-Steiner, B.C., Steiner, F.M., 2013. How diverse is mitopus morio ? Integrative taxonomy detects cryptic species in a small-scale sample of a widespread harvestman. Mol Ecol 22, 3850–3863.

Baker, C.M., Boyer, S.L., Giribet, G., 2020. A well-resolved transcriptomic phylogeny of the mite harvestman family pettalidae (Arachnida, Opiliones, Cyphophthalmi) reveals signatures of Gondwanan vicariance. J Biogeogr 47, 1345–1361.

Barley, A.J., White, J., Diesmos, A.C., Brown, R.M., 2013. The challenge of species delimitation at the extremes: diversification without morphological change in Philippine sun skins: Species delimitation in evolutionary radiations. Evolution 67, 3556–3572.

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., Ingram, K.K., Das, I., 2007. Cryptic species as a window on diversity and conservation. Trends Ecol. Evol. 22, 148–155.

Boyer, S.L., Baker, J.M., Giribet, G., 2007. Deep genetic divergences in Aoraki denticulata (Arachnida, Opiliones, Cyphophthalmi): a widespread 'mite harvestman' defies DNA taxonomy. Mol. Ecol 16, 4999–5016.

Boyer, S.L., Dohr, S.R., Tuffield, M.S., Shu, Y., Moore, C.D., Hahn, K.M., Ward, R.S., Nguyen, P., Morisawa, R., 2022. Diversity and distribution of the New Zealand endemic mite harvestman genus *Aoraki*. Invertebr. Syst. 36, 372–387.

Boyer, S.L., Giribet, G., 2007. A new model Gondwanan taxon: Systematics and biogeography of the harvestman family Pettalidae (Arachnida, Opiliones, Cyphophthalmi), with a taxonomic revision of genera from Australia and New Zealand. Cladistics 23, 337–361.

Boyer, S.L., Giribet, G., 2009. Welcome back New Zealand: regional biogeography and Gondwanan origin of three endemic genera of mite harvestmen (Arachnida, Opiliones, Cyphophthalmi). J. Biogeogr. 36, 1084–1099.

Branstetter, M.G., Longino, J.T., Ward, P.S., Faircloth, B.C., 2017. Enriching the ant tree of life: enhanced UCE bait set for genome-scale phylogenetics of ants and other Hymenoptera. Methods Ecol. Evol. 8, 768–776.

Breiman, L., 2001. No title found. Mach. Learn. 45, 5–32.

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25 (15), 1972–1973.

Carstens, B.D., Pelletier, T.A., Reid, N.M., Satler, J.D., 2013. How to fail at species delimitation. Mol. Ecol. 22 (17), 4369–4383.

Chambers, E.A., Hillis, D.M., 2020. The multispecies coalescent over-splits species in the case of geographically widespread taxa. Syst. Biol. 69, 184–193.

Chan, K.O., Hutter, C.R., Wood Jr, P.L., Su, Y.C., Brown, R.M., 2022. Gene flow increases phylogenetic structure and inflates cryptic species estimations: a case study on widespread philippine puddle frogs (Occidozyga laevis). Syst. Biol. 71 (1), 40–57.

Chapman, A. D. 2006. Numbers of living species in Australia and the world. Australian Govt., Dept. of the Environment and Heritage, Canberra, ACT.

Clouse, R.M., Wheeler, W.C., 2014. Descriptions of two new, cryptic species of metasiro (Arachnida: opiliones: cyphophthalmi: neogoveidae) from South Carolina, USA, including a discussion of mitochondrial mutation rates. Zootaxa 3814, 177.

Curran, J. 2021. Package 'Hotelling.' https://github.com/jmcurran/Hotelling.

Czekanski-Moir, J.E., Rundell, R.J., 2019. The ecology of nonecological speciation and nonadaptive radiations. Trends Ecol. Evol. 34, 400–415.

De Queiroz, K., 2007. Species concepts and species delimitation. Syst. Biol. 56, 879–886.

DeRaad, D.A., McCormack, J.E., Chen, N., Peterson, A.T., Moyle, R.G., 2022. Combining species delimitation, species trees, and tests for gene flow clarifies complex speciation in scrub-jays. Syst. Biol. 71, 1453–1470.

Derkarabetian, S., Benavides, L.R., Giribet, G., 2019b. Sequence capture phylogenomics of historical ethanol-preserved museum specimens: Unlocking the rest of the vault. Mol. Ecol. Res. 19 (6), 1531–1544.

Derkarabetian, S., Castillo, S., Koo, P.K., Ovchinnikov, S., Hedin, M., 2019a. A demonstration of unsupervised machine learning in species delimitation. Mol. Phylogenet. Evol. 139, 106562.

Derkarabetian, S., Ledford, J., Hedin, M., 2011. Genetic diversification without obvious genitalic morphological divergence in harvestmen (Opiliones, Laniatores, Sclerobunus robustus) from montane sky islands of western North America. Mol. Phylogenet. Evol. 61, 844–853.

Derkarabetian, S., Baker, C.M., Giribet, G., 2021. Complex patterns of Gondwanan biogeography revealed in a dispersal-limited arachnid. J. Biogeography 48 (6), 1336–1352.

Derkarabetian, S., Paquin, P., Reddell, J., Hedin, M., 2022a. Conservation genomics of federally endangered Texella harvester species (Arachnida, Opiliones, Phalangodidae) from cave and karst habitats of central Texas. Conserv. Genet. 23, 401–416.

Derkarabetian, S., Starrett, J., Hedin, M., 2022b. Using natural history to guide supervised machine learning for cryptic species delimitation with genetic data. Front. Zool. 19, 8.

Dinno, A., 2018. paran: Horn's Test of Principal Components/Factors. R. package version 1.5.2. https://cran.r-project.org/web/packages/paran/index.html.

Duncan, R.P., Rynerson, M.R., Ribera, C., Binford, G.J., 2010. Diversity of loxosceles spiders in Northwestern Africa and molecular support for cryptic species in the *Loxosceles rufescens* lineage. Mol. Phylogenet. Evol. 55, 234–248.

Emata, K.N., Hedin, M., 2016. From the mountains to the coast and back again: Ancient biogeography in a radiation of short-range endemic harvestmen from California. Mol. Phylogenet. Evol. 98, 233–243.

Ezard, T., Fujisawa, T., Barraclough, T., 2009. splits: SPecies; LImits by Threshold Statistics. R package version 1.0-20. https://rdrr.io/rforge/splits/.

Faircloth, B.C., 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics 32, 786–788.

Faircloth, B.C., 2017. Identifying conserved genomic elements and designing universal bait sets to enrich them. Methods Ecol. Evol. 8, 1103–1112.

Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 61, 717–726.

Faircloth, B. C. 2011. Illumiprocessor - software for Illumina read quality filtering. Brant Faircloth.

Fernández, R., Giribet, G., 2014. Phylogeography and species delimitation in the New Zealand endemic, genetically hypervariable harvestman species, aoraki denticulata (Arachnida, opiliones, cyphophthalmi). Invert. Systemat. 28, 401.
Forster, R.R., 1948. The sub-order cyphophthalmi Simon in New Zealand. Dominion Museum Records Entomol. 1, 79–119.
Forster, R.R., 1952. Supplement to the sub-order cyphophthalmi. Dominion Museum Records Entomol. 1, 179–211.
Giribet, G., 2020. An updated catalogue of the suborder Cyphophthalmi (Arachnida: Opiliones). Rev. Ibérica Aracnol. 37, 61–100.
Giribet, G., Baker, C.M., Sharma, P.P., 2021. A revised phylogeny of the New Caledonian endemic genus Troglosiro (Opiliones: Cyphophthalmi: Troglosironidae) with the description of four new species. Invertebr. Syst. 35 (1), 59–89.
Giribet, G., Shaw, M., Lord, A., Derkarabetian, S., 2022. Closing a biogeographic gap: a new pettalid genus from South Australia (Arachnida : Opiliones : Cyphophthalmi : Pettalidae) with a UCE-based phylogeny of cyphophthalmi. Invertebr. Syst. 36, 1002–1016.
Grabherr, M.G., et al., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29 (7), 644–652.
Graham, M.R., Myers, E.A., Kaiser, R.C., Fet, V., 2019. Cryptic species and co-diversification in sand scorpions from the Karakum and Kyzylkum deserts of Central Asia. Zool. Scr. 48, 801–812.
Hanski, I. and O. E. Gaggiotti. 2004. Metapopulation biology: Past, present, and fugture. Pages 3-22 in Ecology, genetics, and evolution of metapopulations (I. Hanski and O. E. Gaggiotti, eds.). Elsevier, Amsterdam, Netherlands.
Harvey, M.S., 2002. Short-range endemism in the Australian fauna: some examples from non-marine fauna. Invert. Systemat. 16, 555–570.
Hedin, M., Carlson, D., Coyle, F., 2015. Sky island diversification meets the multispecies coalescent - divergence in the spruce-fir moss spider (Microhexura montivaga, Araneae, Mygalomorphae) on the highest peaks of southern Appalachia. Mol. Ecol. 24, 3467–3484.
Hedin, M., Derkarabetian, S., Alfaro, A., Ramírez, M.J., Bond, J.E., 2019. Phylogenomic analysis and revised classification of atypoid mygalomorph spiders (Araneae, Mygalomorphae), with notes on arachnid ultraconserved element loci. PeerJ 7, e6864.
Hedin, M., Foldi, S., Rajah-Boyer, B., 2020. Evolutionary divergences mirror Pleistocene paleodrainages in a rapidly-evolving complex of oasis-dwelling jumping spiders (Salticidae, Habronattus tarsalis). Mol. Phylogenet. Evol. 144, 106696.
Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35, 518–522.
Jackson, N.C., Carstens, B.C., Morales, A.E., O'Meara, B.C., 2017. Species delimitation with gene flow. Syst. Biol. 66, 799–812.
Jay, K.R., Popkin-Hall, Z.R., Coblens, M.J., Oberski, J.T., Sharma, P.P., Boyer, S.L., 2016. New species of Austropurcellia, cryptic short-range endemic mite harvestmen (Arachnida, Opiliones, Cyphophthalmi) from Australia's Wet Tropics biodiversity hotspot. ZooKeys 586, 37–93.
Jombart, T., 2008. adegenet : a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405.
Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. Nat. Methods 14, 587–589.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30 (4), 772–780.
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.
Kingma, D. P., and M. Welling. 2013. Auto-Encoding Variational Bayes, https://doi.org/10.48550/ARXIV.1312.6114.arXiv.
Kulkarni, S., Wood, H., Lloyd, M., Hormiga, G., 2020. Spider-specific probe set for ultraconserved elements offers new perspectives on the evolutionary history of spiders (Arachnida, Araneae). Mol. Ecol. Resour. 20, 185–203.
Leavitt, D.H., Starrett, J., Westphal, M.F., Hedin, M., 2015. Multilocus sequence data reveal dozens of putative cryptic species in a radiation of endemic Californian mygalomorph spiders (Araneae, Mygalomorphae, Nemesiidae). Mol. Phylogenet. Evol. 91, 56–67.
Levins, R. 1970. Extinction. Pages 7-107 in Some mathematical problems in biology (M. Desternhaber, ed.). American Mathematical Society, Providence, RI, USA.
Mallet, J., F. Seixas, and Y. Thawornwattana. 2022. Species, Concepts of. Pages 1-15 in The Encyclopedia of Biodiversity (S. Levin, ed.). Elsevier, Amsterdam, Netherlands.
Martens, J., 2011. The centetostoma scabriculum complex—a group of three cryptic species (Arachnida: Opiliones: Nemastomatidae). Zootaxa 2783, 35.
Martin, B.T., Chafin, T.K., Douglas, M.R., Placyk, J.S., Birkhead, R.D., Phillips, C.A., Douglas, M.E., 2021. The choices we make and the impacts they have: Machine learning and species delimitation in north American box turtles (Terrapene spp.). Mol. Ecol. Resour. 21, 2801–2817.
Moles, J., Derkarabetian, S., Schiaparelli, S., Schrödl, M., Troncoso, J.S., Wilson, N.G., Giribet, G., 2021. An approach using ddRADseq and machine learning for understanding speciation in Antarctic Antarctophilinidae gastropods. Sci. Rep. 11, 8473.
Morisawa, R. 2020. Sequencing Ultraconserved Elements in the New Zealand Mite Harvestman Genus *Rakaia*. Undergraduate Honors Thesis. Macalester College, St. Paul, USA.
Mussmann, S.M., Douglas, M.R., Oakey, D.D., Douglas, M.E., 2020. Defining relictual biodiversity: Conservation units in speckled dace (Leuciscidae: Rhinichthys osculus) of the Greater Death Valley ecosystem. Ecol. Evol. 10 (19), 10798–10817.
Muster, C., Spelda, J., Rulik, B., Thormann, J., von der Mark, L., Astrin, J.J., 2021. The dark side of pseudoscorpion diversity: The German Barcode of life campaign reveals high levels of undocumented diversity in European false scorpions. Ecol. Evol. 11, 13815–13829.
Newton, L.G., Starrett, J., Hendrixson, B.E., Derkarabetian, S., Bond, J.E., 2020. Integrative species delimitation reveals cryptic diversity in the southern Appalachian Antrodiaetus unicolor (Araneae: Antrodiaetidae) species complex. Mol. Ecol. 29, 2269–2287.
Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32 (1), 268–274.
Niemiller, M.L., Near, T.J., Fitzpatrick, B.M., 2012. Delimiting species using multilocus data: Diagnosing cryptic diversity in the southern cavefish, Typhlichthys subterraneus (Teleostei: Amblyopsidae): Species delimitation in cavefish. Evolution 66, 846–866.
Oberski, J.T., Sharma, P.P., Jay, K.R., Coblens, M.J., Lemon, K.A., Johnson, J.E., Boyer, S.L., 2018. A dated molecular phylogeny of mite harvestmen (Arachnida Opiliones Cyphophthalmi) elucidates ancient diversification dynamics in the Australian wet tropics. Mol. Phylogenet. Evol. 127, 813–822.
Obiol, J.F., Herranz, J.M., Paris, J.R., Whiting, J.R., Rozas, J., Riutort, M., González-Solís, J., 2023. Species delimitation using genomic data to resolve taxonomic uncertainties in a speciation continuum of pelagic seabirds. Mol. Phylogenet. Evol. 179, 107671.
Ohira, H., Sato, K., Tsutsumi, T., Kaneko, S., Choi, H.-J., 2018. DNA barcoding suggested the existence of cryptic species and high biodiversity of South Korean pseudoscorpions (Arachnida, Pseudoscorpiones). J. Asia-Pac. Biodivers. 11, 399–407.
Pei, J., Chu, C., Li, X., Lu, B., Wu, Y., 2018. CLADES: A classification-based machine learning method for species delimitation from population genetic data. Mol. Ecol. Resour. 18, 1144–1156.
Pfingstl, T., Lienhard, A., Baumann, J., Koblmüller, S., 2021. A taxonomist's nightmare – Cryptic diversity in Caribbean intertidal arthropods (Arachnida, Acari, Oribatida). Mol. Phylogenet. Evol. 163, 107240.
Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D., Vogler, A.P., 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed Insects. Syst. Biol. 55, 595–609.
Puillandre, N., Brouillet, S., Achaz, G., 2021. ASAP: assemble species by automatic partitioning. Mol. Ecol. Resour. 21, 609–620.
Pyron, R.A., O'Connell, K.A., Duncan, S.C., Burbrink, F.T., Beamer, D.A., 2023. Speciation hypotheses from phylogeographic delimitation yield an integrative taxonomy for seal salamanders (Desmognathus monticola). Syst. Biol. 72, 179–197.
Pyron, R. A. 2024. Unsupervised machine learning for species delimitation, integrative taxonomy, and biodiversity conservation. bioRxiv preprint: https://doi.org/10.1101/2023.06.12.544639.
Quattrini, A.M., Faircloth, B.C., Dueñas, L.F., Bridge, T.C.L., Brugler, M.R., Calixto-Botía, I.F., DeLeo, D.M., Forêt, S., Herrera, S., Lee, S.M.Y., Miller, D.J., Prada, C., Rádis-Baptista, G., Ramírez-Portilla, C., Sánchez, J.A., Rodríguez, E., McFadden, C. S., 2018. Universal target-enrichment baits for anthozoan (Cnidaria) phylogenomics: New approaches to long-standing problems. Mol. Ecol. Resour. 18, 281–295.
Reveillion, F., Wattier, R., Montuire, S., Carvalho, L.S., Bollache, L., 2020. Cryptic diversity within three South American whip spider species (Arachnida, Amblypygi). Zool. Res. 41, 595–598.
Román-Palacios, C., and J. J. Wiens. 2020. Recent responses to climate change reveal the drivers of species extinction and survival. Proceedings of the National Academy of Sciences 117:4211–4217. Proceedings of the National Academy of Sciences.
Scrucca, L., Fop, M., Murphy, T.B., Raftery, A.E., 2016. mclust 5: clustering, classification and density estimation using gaussian finite mixture models. R. J. 8, 289–317.
Shu, Y.N., Moyes, P.A.F., Anderson, S.F., Boyer, S.L., 2023. Patterns of co-occurrence in New Zealand mite harvestmen (Arachnida, Opiliones, Cyphophthalmi). Integr. Comp. Biol. 62, S288.
Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J.M., Birol, Í., 2009. ABySS: A parallel assembler for short read sequence data. Genome Res 19, 1117–1123.
Skoracka, A., Magalhães, S., Rector, B.G., Kuczyński, L., 2015. Cryptic speciation in the Acari: a function of species lifestyles or our ability to separate species? Exp. Appl. Acarol. 67, 165–182.
Smith, M.L., Carstens, B.C., 2020. Process-based species delimitation leads to identification of more biologically relevant species. Evolution 74, 216–229.
Smith, B.T., Harvey, M.G., Faircloth, B.C., Glenn, T.C., Brumfield, R.T., 2014. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Syst. Biol. 63, 83–95.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
Starrett, J., Derkarabetian, S., Hedin, M., Bryson, R.W., McCormack, J.E., Faircloth, B.C., 2017. High phylogenetic utility of an ultraconserved element probe set designed for Arachnida. Mol. Ecol. Resour. 17, 812–823.
Struck, T.H., Feder, J.L., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V.I., Kistenich, S., Larsson, K.-H., Liow, L.H., Nowak, M.D., Stedje, B., Bachmann, L., Dimitrov, D., 2018. Finding evolutionary processes hidden in cryptic species. Trends Ecol. Evol. 33, 153–163.
Suchard M. A., P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, and A. Rambaut. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10 *Virus* Evolution 4, vey016.
Sukumaran, J., Knowles, L.L., 2017. Multispecies coalescent delimits structure, not species. Proc. Natl. Acad. Sci. U.S.A. 114, 1607–1612.

Tardelli Canedo, P., Baker, C.M., Morisawa, R., Pessereau, E.J., Boyer, S.L., 2021. Phylogeography of Neopurcellia salmoni, a widespread mite harvestman from the South Island of New Zealand, with the first report of male polymorphism in the suborder Cyphophthalmi (Arachnida: Opiliones). J. Arachnol. 49.

Tyagi, K., Kumar, V., Kundu, S., Pakrashi, A., Prasad, P., Caleb, J.T.D., Chandra, K., 2019. Identification of Indian spiders through DNA barcoding: Cryptic species and species complex. Sci Rep 9, 14033.

Urban, M.C., 2015. Accelerating extinction risk from climate change. Science 348, 571–573.

van der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 2579–2605.

Wheeler, Q. D. and Meier, R., eds. 1998. Species Concepts and Phylogenetic Theory: A Debate. Columbia University Press, New York, NY, USA.

Wiens, J.J., Graham, C.H., 2005. Niche conservatism: Integrating evolution, ecology, and conservation biology. Annu. Rev. Ecol. Evol. Syst. 36, 519–539.

Zhang, J., Kapli, P., Pavlidis, P., Stamatakis, A., 2013. A general species delimitation method with applications to phylogenetic placements. Bioinformatics 29 (22), 2869–2876.

Zhang, C., Mirarab, S., 2022. Weighting by gene tree uncertainty improves accuracy of quartet-based species trees. Mol. Biol. Evol. 39 (12), msac215.