

Research Article


Cite this article: De Vivo M, Chen W-Y, Huang J-P (2023). Testing the efficacy of different molecular tools for parasite conservation genetics: a case study using horsehair worms (Phylum: Nematomorpha). *Parasitology* **150**, 842–851. <https://doi.org/10.1017/S0031182023000641>

Received: 21 February 2023
 Revised: 26 June 2023
 Accepted: 27 June 2023
 First published online: 7 July 2023

Keywords:
 COXI; ddRADseq; nematomorpha; parasite conservation; population size; Taiwan

Corresponding author:
 Mattia De Vivo;
 Email: mattiadevivopatalano@gmail.com

Testing the efficacy of different molecular tools for parasite conservation genetics: a case study using horsehair worms (Phylum: Nematomorpha)

Mattia De Vivo^{1,2,3} , Wei-Yun Chen³ and Jen-Pan Huang³

¹Department of Life Science, National Taiwan Normal University, Taipei, Taiwan; ²Biodiversity Program, Taiwan International Graduate Program, Taipei, Taiwan and ³Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

Abstract

In recent years, parasite conservation has become a globally significant issue. Because of this, there is a need for standardized methods for inferring population status and possible cryptic diversity. However, given the lack of molecular data for some groups, it is challenging to establish procedures for genetic diversity estimation. Therefore, universal tools, such as double-digest restriction-site-associated DNA sequencing (ddRADseq), could be useful when conducting conservation genetic studies on rarely studied parasites. Here, we generated a ddRADseq dataset that includes all 3 described Taiwanese horsehair worms (Phylum: Nematomorpha), possibly one of the most understudied animal groups. Additionally, we produced data for a fragment of the cytochrome c oxidase subunit I (COXI) for the said species. We used the COXI dataset in combination with previously published sequences of the same locus for inferring the effective population size (N_e) trends and possible population genetic structure.

We found that a larger and geographically broader sample size combined with more sequenced loci resulted in a better estimation of changes in N_e . We were able to detect demographic changes associated with Pleistocene events in all the species. Furthermore, the ddRADseq dataset for *Chordodes formosanus* did not reveal a genetic structure based on geography, implying a great dispersal ability, possibly due to its hosts. We showed that different molecular tools can be used to reveal genetic structure and demographic history at different historical times and geographical scales, which can help with conservation genetic studies in rarely studied parasites.

Introduction

In recent years, there has been a growing concern regarding the conservation of parasites that do not harm humans or do not impede conservation attempts (e.g. Dougherty *et al.*, 2016; Carlson *et al.*, 2020). This is because parasites are integral parts of the ecosystems from several points of view, e.g. species diversity, biomass, evolution of host immunity and importance in food webs (Dougherty *et al.*, 2016). Moreover, they are affected by anthropogenic environmental changes and can go extinct together with their hosts (Carlson *et al.*, 2020). However, tools for parasite conservation have rarely been tested (Carlson *et al.*, 2020). Particularly, molecular tools (e.g. DNA barcoding and reconstruction of demographic histories through molecular data) have not been used extensively in non-medically relevant parasites and they can be useful for estimating general population parameters [e.g. genetic diversity and trends of effective population size (N_e); Criscione *et al.*, 2005; Selbach *et al.*, 2019; Strobel *et al.*, 2019; Carlson *et al.*, 2020]. Genetic diversity may be linked to local adaption or life history, and its estimation may help to understand ecological traits or even infer potential population bottlenecks (van Schaik *et al.*, 2015; Tobias *et al.*, 2017; Radačovská *et al.*, 2022). N_e is known to greatly influence the genetic diversity and viability of populations (Criscione *et al.*, 2005; Pérez-Pereira *et al.*, 2022), but estimating it in parasites may come with some caveats depending on the studied system (Criscione *et al.*, 2005; Strobel *et al.*, 2019). Nevertheless, estimates of N_e can be helpful for considering the taxa's conservation status (Pérez-Pereira *et al.*, 2022). Unfortunately, the current lack of genetic resources makes it difficult for researchers to enact timely conservation plans, putting some species at risk of extinction (Carlson *et al.*, 2020).

Methods for estimating N_e trends from genetic data of non-model organisms are being developed (e.g. Liu and Fu, 2020). This has been possible thanks to protocols that allow researchers to generate genome-wide data without reference sequences (Peterson *et al.*, 2012; Nunziata and Weisrock, 2018). For example, restriction-site associated DNA sequencing (RADseq) approaches have been shown to work well with non-model organisms and became popular in population and conservation genetic studies due to their relatively low cost, versatility and ability to generate data for up to tens of thousands of loci (Peterson *et al.*, 2012; Nunziata and Weisrock, 2018). Furthermore, the produced data can be used with

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (<http://creativecommons.org/licenses/by-nc/4.0>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.

bioinformatic tools for estimating population trends at least 20 generations before a decline (Nunziata and Weisrock, 2018) and genetic structure (e.g. Dalapiccola *et al.*, 2021). This is interesting from a molecular conservation parasitology perspective, given that non-medically relevant parasites often lack reference data (Selbach *et al.*, 2019). Alternatively, even single-locus methods for *Ne* estimation are available, allowing researchers to use previously released data (e.g. sequences from GenBank) for demographic inferences, although having more loci leads to better parameter estimation (Ho and Shapiro, 2011).

In this study, we used multiple protocols and methods for generating molecular data to study population genetic structure (i.e. potential genetic differentiation based on host or geography: e.g. van Schaik *et al.*, 2015 and Dalapiccola *et al.*, 2021) and the changes of *Ne* over time in freshwater horsehair worms, also called 'gordiids' (phylum Nematomorpha, class Gordioida; Sup. Info) in Taiwan. Such worms are regarded as one of the least studied animal group, especially from a molecular perspective (Bolek *et al.*, 2015; Tobias *et al.*, 2017) and they have interesting ecological and conservation characteristics (Sato *et al.*, 2014; Bolek *et al.*, 2015; Sup. Info). The methods used in this study can be applied to other parasitic groups for conservation and biodiversity-oriented studies. Specifically, a double-digest restriction-site-associated DNA sequencing (ddRADseq) dataset including all 3 known horsehair species in Taiwan (Sup. Info) was generated. Additionally, DNA sequence data from the mitochondrial cytochrome c oxidase subunit I (COXI) region were generated and used for evaluating population and genetic status in combination with previously released data (Chiu *et al.*, 2011, 2016, 2017, 2020). This mitochondrial marker was used due to the lack of any other previously released molecular data on Taiwanese horsehair worms. Although the use of mitochondrial data for genetic diversity studies can be controversial, COXI is regarded as a good proxy for evaluating the conservation status of species population for the first time (Petit-Marty *et al.*, 2021 and references therein). Additionally, such marker was also utilized in a previous study about hairworms' population genetic structure (Tobias *et al.*, 2017) and its availability in public repositories is expected to increase, given its use in barcoding (Selbach *et al.*, 2019). With both datasets, population genetic structure and demographic history were estimated and compared. More focus was given to *Chordodes formosanus*, since they are relatively easier to sample and are more common (Chiu, 2017). A country-wide ddRADseq dataset was generated for the said species, allowing us to estimate *Ne* and population genetic structure at Taiwan main island-level. Additionally, the Taiwanese and Japanese populations' trends were compared using the COXI dataset.

Materials and methods

Sampling

Gordiids were sampled over a span of 14 years, from 2007 to 2021 (Fig. 1; Sup. Table 1). *Chordodes formosanus*, the most common species in Taiwan (Chiu, 2017), was sampled country-wide, while specimens of *Acutogordius taiwanensis* and *Gordius chiaschanus* were collected from fewer localities (Fig. 1; Sup. Table 1). Specimens were collected either by hand in bodies of water (i.e. streams and man-made ponds), by sampling the hosts and then immersing them until the adult worm came out, or by collecting the dead gordiids on the ground, which is the most common method for finding them in Taiwan (Chiu, 2017). The samples were fixed in 95% alcohol and stored at -20°C . Species recognition was based on DNA barcoding, given that the majority of the samples were too ruined for morphological identification. A total of 38 *C. formosanus* specimens, 10

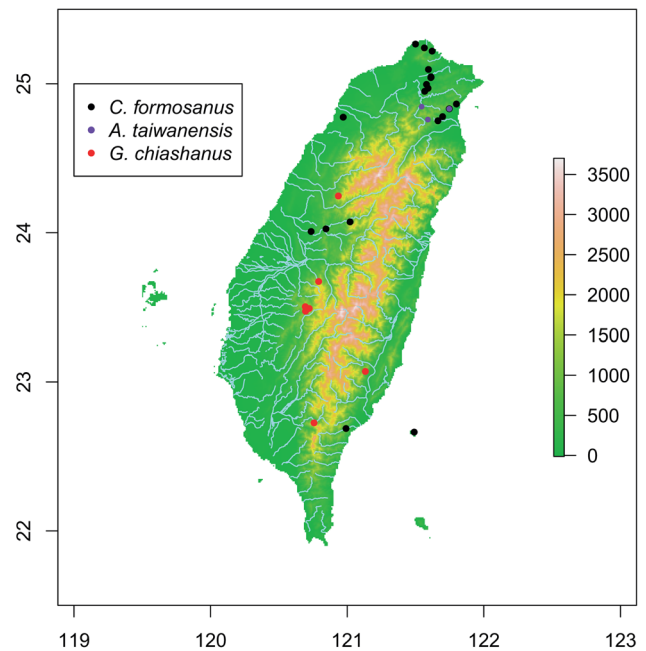


Figure 1. Sampling localities for *Chordodes formosanus* (black), *Acutogordius taiwanensis* (purple) and *Gordius chiaschanus* (red), with a map showing elevation and river network in Taiwan. Localities from previous studies have been included. The purple dot with black borders shows an area of sympatry between *C. formosanus* and *A. taiwanensis*. The legend on the right shows elevation in metres.

A. taiwanensis and 3 *G. chiaschanus* were collected (see SRA BioProject PRJNA914055); however, DNA extraction and sequencing was not successful on all of them (Sup. Table 1), probably due to the low quality of some of the specimens.

DNA extraction, COXI amplification and ddRADseq protocol

For DNA extraction, we used a modified version of the Qiagen DNeasy Animal Tissue Protocol (Qiagen, Hilden, Germany). We modified the original protocol by using double-distilled water (ddH_2O) instead of elution buffer for dilution, and we eluted $60\ \mu\text{L}$ of DNA, given the low concentration of DNA. Additionally, we left the tissue to incubate overnight at 60°C .

The amplification of the COXI sequences through polymerase chain reaction (PCR) was done using the universal primer set LCO1490 and HC02198 (Folmer *et al.*, 1994) and was performed in a total volume of $25\ \mu\text{L}$ using PuReTaq Ready-To-Go PCR Beads (Cytiva, formerly GE Healthcare, Marlborough, United States). The PCR was initiated at 95°C for 5 min, followed by 40 cycles at 95°C for 1 min, 50°C for 1 min and 72°C for 1 min, with a final extension at 72°C for 10 min. The PCR products were purified following the NautiaZ Gel/PCR DNA Purification Mini Kit protocol (Nautia Gene, Taipei, Taiwan). Sanger Sequencing was done by the DNA Sequencing Core Facility of the Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan.

An edited version of the Peterson *et al.* (2012) protocol was used for generating the ddRADseq library. We modified the original method by using between 100 and 140 ng of genomic DNA per sample, $4\ \mu\text{L}$ of ddH_2O for bead clean-up and $60\ \mu\text{L}$ of ddH_2O for elution before measuring DNA concentration with Qubit. Single-end sequencing of 100-nucleotide base pair (bp) lengths were carried out using an Illumina HiSeq2500 system.

COXI analyses

The COXI forward and reversed sequences from each individual were visualized with MEGA X (ver. 10.1.8; Kumar *et al.*, 2018).

The alignments were done with MAFFT 7.471 (Kato and Standley, 2013) using the L-INS-i algorithm. The sequences were trimmed manually and consensus sequences were saved from the alignments.

COXI sequences at least 400 bp long for each taxon from previous studies (Chiu et al. 2011, 2016, 2017, 2020; Sup. Table 2) were downloaded from GenBank (Clark et al., 2016: last assessed 4th November 2022), aligned with the newly generated sequences using MAFFT 7.471 with the L-INS-i algorithm and manually trimmed while being visualized with MEGA X. In the case of *A. taiwanensis*, an additional sequence of a hairworm from Myanmar (MF983649) was included for population genetic structure analyses; said sequence has been shown to fall inside the known COXI variability of *A. taiwanensis* in previous studies (Chiu et al., 2020) and therefore it should represent a sequence from the said species. In total, 433 bp were recovered for 85 specimens of *C. formosanus* (36 newly generated), 414 bp for 31 *A. taiwanensis* (3 newly generated) and 382 bp for 26 individual of *G. chiashanus* (1 newly generated). The newly generated sequences are available in GenBank (OQ121045-OQ121047 for *A. taiwanensis*, OQ121048-OQ121083 for *C. formosanus*, OQ121084 for *G. chiashanus*).

Possible population genetic structure, hypothetical haplotypes (i.e. unsampled sequences) and genetic diversity indices [nucleotide diversity (π) and Tajima's D; Nei and Li, 1979; Tajima, 1989] were inferred by a haplotype network in PopArt (Leigh and Bryant, 2015), generated with the TCS 1.21 algorithm (Clement et al., 2000). Estimation of N_e trends was calculated by running Coalescent Bayesian Skyline plot analyses on BEAST2 (version 2.6.3; Bouckaert et al., 2019). Such analyses allow the estimation of trends from mitochondrial data, even if only one locus is available, and recover well with Pleistocene trends (Ho and Shapiro, 2011). Furthermore, previous studies were able to recover demographic trends with less than 300 bp of mitochondrial sequences (Ho and Shapiro, 2011 and references therein); therefore, the alignments should be long enough for allowing trends' estimations. The most likely nucleotide substitution model for each species was inferred using ModelTest, implemented in raxmlGUI (Edler et al., 2021) according to Akaike information criterion (AIC; Akaike, 1998). A Relaxed Clock Log model, with a clock rate of roughly 0.0013 per million years, was set. We estimated this clock rate by aligning full COXI sequences from the only available *Chordodes* and *Gordius* mitogenomes from GenBank (MG257764 and MG257767), dividing the number of variable sites by the total number of sites and then dividing again by 2 before dividing by 110, which represents the estimated age in million years of the oldest known gordiid fossil (Poinar and Buckley, 2006). Note that estimates for the appearance of crown group Nematomorpha range from around the mid-Cambrian to the mid-Cretaceous (Howard et al., 2022), but that these estimates include the possible divergence time of the *Nectonema* lineage from gordiids too. Therefore, we prefer to refer to the fossil datum, since it is the earliest possible divergence point for the Chordodidae/Gordiidae lineage. Additionally, the *C. formosanus* Japanese and Taiwanese samples were split for estimations, while the Myanmar sequence was removed for *A. taiwanensis*.

ddRADseq analyses

Two different pipelines for processing the ddRAD data were used for *C. formosanus*: ipyrad (version 0.9.84: Eaton and Overcast, 2020) and Stacks (version 2.6.2: Rochette et al., 2019). For the ipyrad pipeline, we performed de-novo assembly with a Phred Qscore offset of 43, 90% clust threshold, 0.5 maximum of heterozygotes in consensus, and minimum sample locus at 80% of all

the samples, while keeping other settings at their default values. Samples with less than 300 loci were discarded for further analyses. For Stacks, the protocol listed in Rivera-Colón and Catchen (2022) was followed and all the sites (variant and invariant) were outputted in the Variant Call Format (VCF) file. In total, 27 individuals were retained for *C. formosanus* from both pipelines. In the case of *A. taiwanensis* and *G. chiashanus*, only 2 individuals per species had at least 200,000 reads after trimming and therefore we only used Stacks for their assemblies, with the 'r' flag set at 1 instead of 0.8.

To check the genetic population genetic structure of *C. formosanus*, tools from ipyrad and R were used with both the ipyrad and Stacks output. The Stacks VCF file was filtered to remove the loci without SNPs and the SNPs with more than 20% missing data, using an edited version of previously released R scripts (Dalapiccola et al., 2021). Additionally, the filtered VCF file was converted to a HDF5 database using the converter inside the ipyrad-analysis toolkit (Eaton and Overcast, 2020) to make it compatible with the tools implemented inside ipyrad. After this, *k-means* principal component analyses (PCA) were run 100 times in the ipyrad-analysis tools suite. Moreover, the *snappclust* clustering method implemented into the 'adegenet' R package (Beugin et al., 2018) was used. The best number of clusters (*k*) was calculated by computing the AIC and the Bayesian information criterion (BIC; Schwarz, 1978), while picking the value for which both criteria were lower. The *snappclust* analyses were run with both the ipyrad full output STRUCTURE file and a STRUCTURE file generated from the filtered Stacks VCF thanks to PGDSpider (Lischer and Excoffier, 2012). Furthermore, for analysing possible admixture between geographically separated *C. formosanus* individuals, RADpainter and fineRADstructure (Malinsky et al., 2018) were used and the results were plotted with R. The RADpainter file from the Stacks output was used for such software, while the 'hapsFromVCF' command generated an input file with the ipyrad-outputted VCF.

For population demographic analyses of all the taxa, a site-frequency spectra (SFS) file was generated using the easySFS script (<https://github.com/isaacovercast/easySFS>) by converting the Stacks VCF file with the invariant sites, choosing the number of individuals that maximize the number of segregating sites. Note that easySFS takes into consideration the fact that missing data are present, and therefore we used the full Stacks VCF output for SFS conversion. After that, Stairway Plot 2 (version 2.1.1: Liu and Fu, 2020) was used per species following the software's manual. Given the current knowledge on Taiwanese horsehair worms (Chiu, 2017; Chiu et al., 2020), generation time was set at 1 per year. As for the mutation rate, given the lack of whole-genome data for horsehair worms, a spontaneous mutation of rate of 2×10^{-9} per site per generation, which is one of the nematode *Pristionchus pacificus* (Weller et al., 2014), was set. We argue that, even if nematodes and hairworms have been split for hundreds of millions of years (Howard et al., 2022), mutation rates in invertebrates tend to be around the same order of magnitude (10^{-9} ; e.g. Konrad et al., 2019) and therefore should not influence the results in a significant way.

Results

Coxi analyses

In total, there were 54 different mitochondrial haplotypes in *C. formosanus*, 9 of which were hypothetical and 28 of which were made up by a single individual (Fig. 2A). *Acutogordius taiwanensis* presented 9 haplotypes, none of which were hypothetical and 5 of which were made up by a single individual; however, 5 haplotypes were separated from the most common

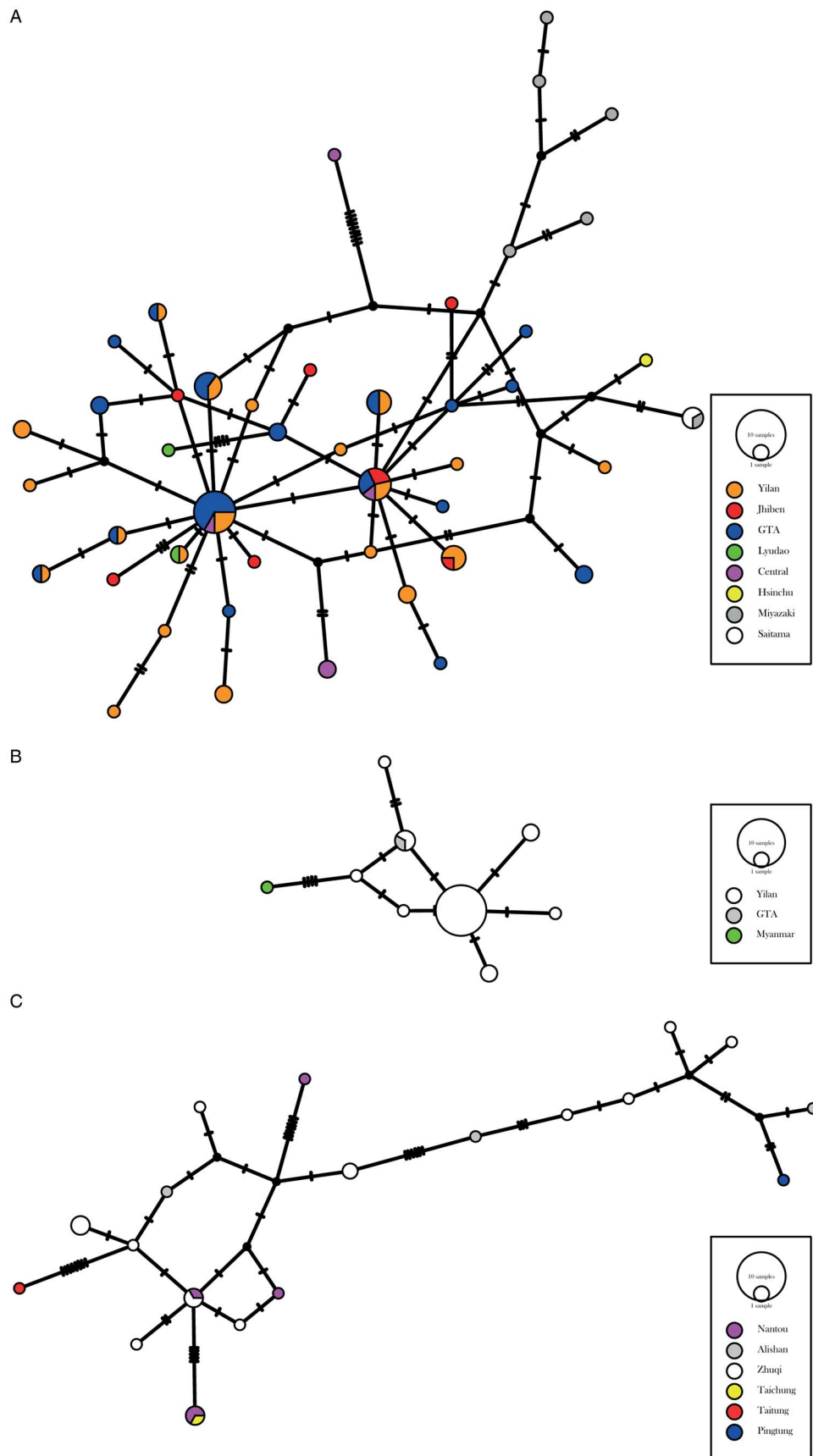


Figure 2. Haplotype networks based on COXI fragments. (A) *Chordodes formosanus*; (B) *Acutogordius taiwanensis*; (C) *Gordius chiashanus*.

one by just 1 mutation (Fig. 2B). 24 haplotypes, with 5 hypotheticals, were counted for *G. chiashanus*, and 15 of these were made up by a single individual (Fig. 2C).

Pi was 297585 for *C. formosanus*, around 0.001 for *A. taiwanensis* and 67063.1 for *G. chiashanus*. *C. formosanus* had 54 segregating sites and 30 parsimony-informative ones, while *A. taiwanensis* had 11 and

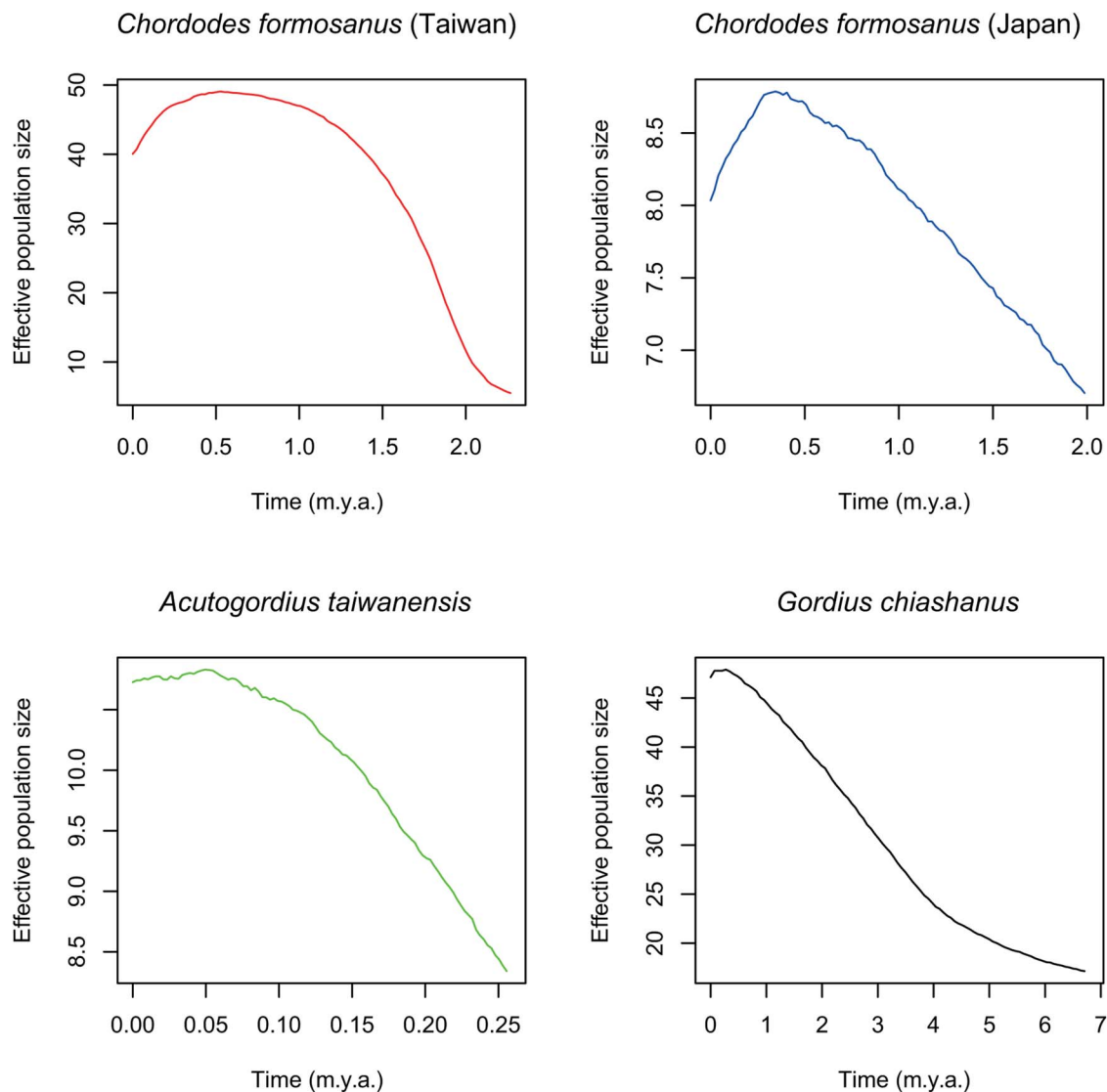


Figure 3. Bayesian Sky Plots based on COXI fragments per each species/population.

4 and *G. chiashanus* had 30 and 23, respectively. Tajima's D for *C. formosanus* was around 3.889×10^7 , while it was around -2.692 for *A. taiwanensis* and around 1.209×10^7 for *G. chiashanus*; this statistic was significant ($P = 0$) for both *C. formosanus* and *G. chiashanus*, but not for *A. taiwanensis* ($P \approx 1$) (Sup. Table 3).

For the Coalescent Bayesian Skyline plot analyses, the number of variable sites was 49 for the Taiwanese samples of *C. formosanus*, while it was 12 for the Japanese ones. After removing the Myanmar sequence, 7 variable sites were retained for *A. taiwanensis*. We did not change the alignment file for *G. chiashanus*, and therefore 30 variable sites remained (Sup. Data). The best substitution model for the Taiwanese samples of *C. formosanus* was TrN + G4, while TIM1 was the best for the Japanese ones. HKY + I and TPM3uf + I were the best models for *A. taiwanensis* and *G. chiashanus*, respectively (Sup. Table 4). The Bayesian Skyline Plot results show how both the populations of *C. formosanus* are having a small decline, while *A. taiwanensis*'s population tends to be stable (Fig. 3). *G. chiashanus*'s population size seemed to be increasing until recently (Fig. 3).

ddRADseq analyses

The ipyrad trimming step led to an average read number of 2195578.39 per *C. formosanus* sample (including samples that

were discarded because of the low number of reads or low number of retained loci). The ipyrad final output for *C. formosanus* had a SNP matrix size of 2248, with 17.82% missing sites, while the total sequence matrix size was 52233, with 17.59% missing sites. The lowest number of loci in assembly per an individual was 331, while the highest one was 548 and the average was around 473.889 loci per samples (Sup. Data).

For the Stacks output, an average read number of 1672668.836 per sample was retained (including samples that were discarded because of low number of reads). The optimal settings for *C. formosanus* were obtained with both the number of allowed mismatches (M) and the number of allowed mismatches between individual loci and the catalogue of loci (n) set to 6. For *A. taiwanensis* and *G. chiashanus*, the best results were obtained with M and n set to 12 and 9, respectively. The number of R80 loci was 1876 for *C. formosanus*, 1388 for *A. taiwanensis* and 2484 for *G. chiashanus*. After removing the SNPs with 20% of the missing data, 4243 SNPs were retained for PCA and *snappclust* analyses for *C. formosanus* (Sup. Data).

Both PCA and *snappclust* analyses, conducted using both the output VCFs (ipyrad and filtered Stacks), failed to trace any structure whatsoever in *C. formosanus*, with both the AIC and BIC computed by *snappclust* agreeing that $k = 1$ was the optimal number of clusters for both datasets (Fig. 4; Sup. Figs 1 and 2). The

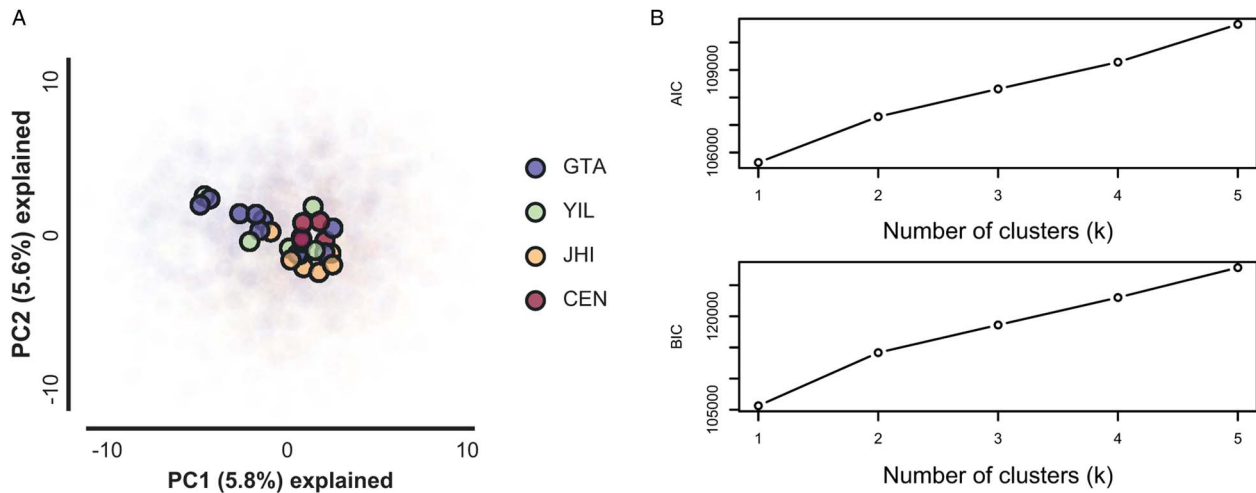


Figure 4. (A) PCA results for Stacks ddRAD-seq data for *Chordodes formosanus*. (B) AIC and BIC values from *snappclust* analyses for *C. formosanus* based on the same data as the PCA. The PCA and *snappclust* results based on *ipyrad* are available as Supplementary Figs 1 and 2.

same pattern was observed using *fineRADstructure* with both input files, given that no cluster formed based on shared coancestry (Fig. 5; Sup. Fig. 3).

The Stairway Plot output for *C. formosanus* shows several population drops during the Pleistocene and an estimated N_e of around 1000 individuals in more recent times (Fig. 6; Sup. Fig. 4). However, only 2 drops were detected for both *A. taiwanensis* and *G. chiashanus*, and their estimated N_e in recent times were larger (around 300,000 for the former and 150,000 for the latter; Sup. Figs 5 and 6).

Discussion

COXI and ddRADseq were used for evaluating population genetic structure and trends of N_e in the considered gordiids. The trends for the most sampled species, *C. formosanus*, were similar in the Pleistocene according to both datasets, although the Stairway Plot based on ddRAD was able to detect more changes in the last 100,000 years. For the other 2 species, the scale of the recorded events with COXI differed between each other and *C. formosanus*, while the ddRAD datasets were too small for detecting genetic structure and recent trends. In all the 3 species, no differentiation based on geography was found according to COXI. The results associated with each species and datasets are discussed below.

COXI diversity analyses

Both *C. formosanus* and *G. chiashanus* exhibit high levels of intraspecific diversity and do not exhibit a geographical genetic structure based on mitochondrial data. High mitochondrial diversity without population genetic structure in Nematomorpha species has been shown in taxa from New Zealand (Tobias *et al.*, 2017), hinting at a high dispersal ability or a multigenerational dispersal process. In the case of the previous study, the authors were surprised to discover a lack of genetic structure in *Euchordodes nigromaculatus*, because the species was known to parasitize wētās (family: Anostomatidae), which are definitive hosts that do not disperse well and exhibit a strong population genetic structure. Additionally, the known paratenic hosts (i.e. used for transport before the final infection and not for development: Criscione *et al.*, 2005; Bolek *et al.*, 2015) should not be able to disperse greatly in the mountain range inhabited by *E. nigromaculatus* (Tobias *et al.*, 2017). However, new studies have revealed that this New Zealand species can also infect cockroach *Celatoblatta quinque maculata*, a common insect in the mountains of central Otago (Doherty *et al.*, 2022),

which is the area considered by Tobias *et al.* (2017) for the study. This suggests that this cockroach shows more dispersal ability than the wētās. Therefore, it could be the case that dispersal is facilitated by definitive hosts. Note that dispersal by definitive hosts does not exclude the possibility of a multigenerational process, as highlighted by Tobias *et al.* (2017). Given the scarce knowledge of paratenic hosts used by Taiwanese horsehair worms, we should not exclude them as possible vectors also.

In the case of *G. chiashanus*, only one definitive host (a millipede species from the genus *Spirobolus*) and only one potential paratenic host [the mayfly *Ephemera orientalis* (Chiu *et al.*, 2020), which is known to be present in different East Asian countries (Lee *et al.*, 2008)] are known. The mayfly does not seem to be commonly infected (Chiu *et al.*, 2020), which raises questions about both the paratenic and definitive host ranges of *G. chiashanus*: such ranges may include other species that are able to disperse in middle elevation areas, given the high diversity shown by the haplotype network. Another hairworm species with terrestrial adult ecology, *Gordius terrestris*, prefers earthworms as natural paratenic hosts (Anaya *et al.*, 2021) and *G. chiashanus* might do the same. All the possible explanations for dispersal listed here (by paratenic hosts, by definitive hosts, by multigenerational process) do not exclude each other, and it is likely that hairworm dispersal is a complex process that may include all the options listed above.

While *A. taiwanensis* showed less intraspecific genetic diversity than the other 2 considered taxa, it must be noted that almost all the samples for this species came from a single region of Taiwan, Yilan County (Chiu *et al.*, 2017; this study). Meanwhile, the COXI sequences of *C. formosanus* were generated from individuals from multiple localities in Taiwan and Japan (Chiu *et al.* 2011, 2016, 2017; this study), and the ones for *G. chiashanus* came from multiple middle elevation mountain localities in central-south Taiwan (Chiu *et al.*, 2020; this study). These differences in sampling areas can be explained by the varying difficulties encountered when sampling the adult horsehair worms (Sup. Info). Sampling is usually regarded as a significant issue while studying Nematomorpha, due to the lack of standardized collecting methods and the short lifespan of the adults (Bolek *et al.*, 2015). Furthermore, biological differences among species may make some taxa easier to collect compared to others (Sup. Info). As a result of sampling difficulties, it is likely that the genetic diversity of *A. taiwanensis* is extremely underestimated, as already suggested by Chiu (2017) and Chiu *et al.* (2020), and the species may have gone unnoticed in many areas. Alternatively, *A. taiwanensis* may be the most vulnerable Taiwanese hairworm species to human actions (Chiu

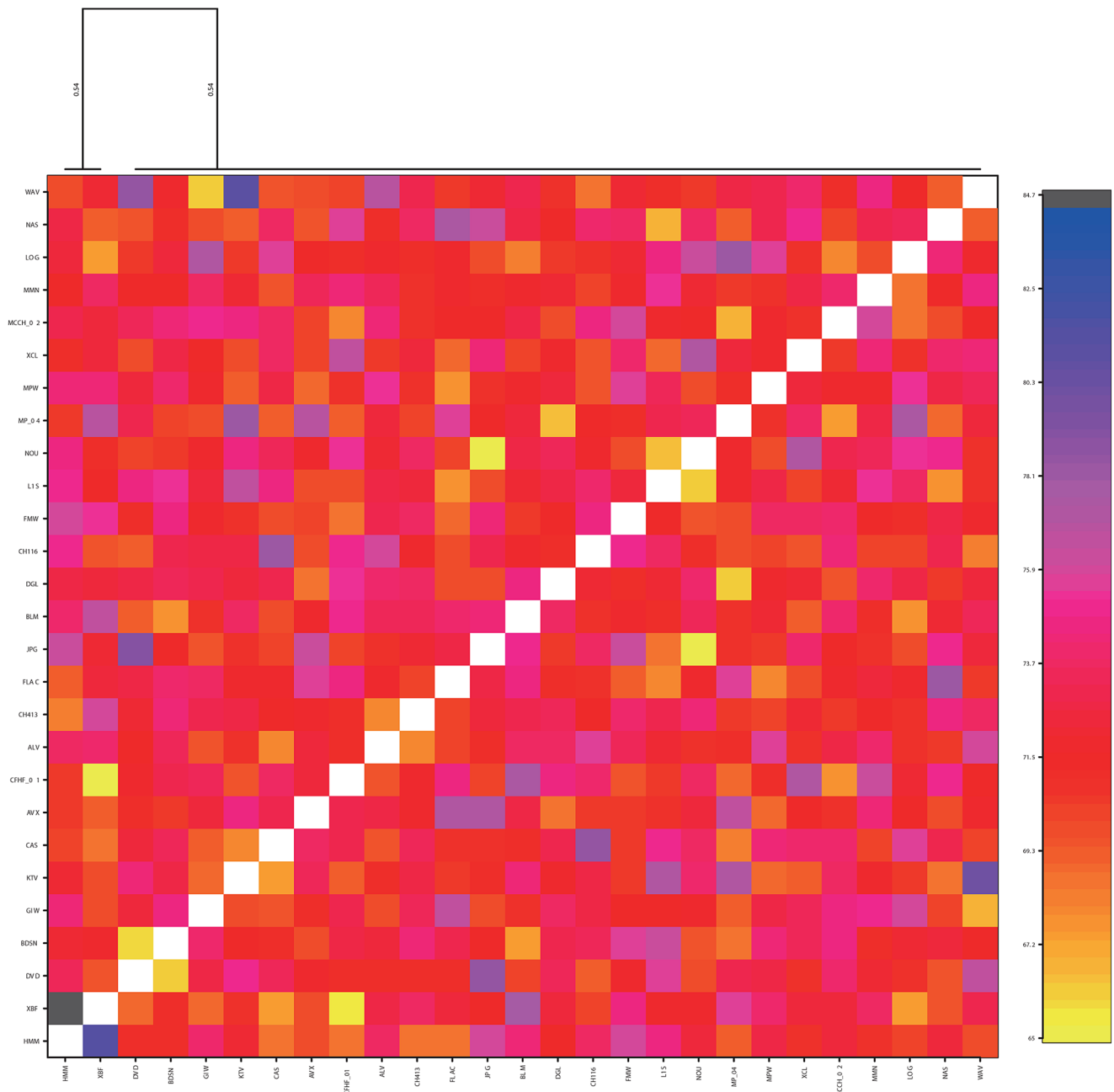


Figure 5. Co-ancestry matrix for *Chordodes formosanus* based on Stacks ddRAD-seq data. The one from ipyrad data is available as Supplementary Fig. 3.

et al., 2016); previous studies highlighted the possibility of hair-worm extirpation caused by anthropic activities (Sato *et al.*, 2014). Therefore, *A. taiwanensis* may have gone extinct in some areas in Taiwan due to human disturbance.

ddRADseq diversity analyses for *Chordodes formosanus* in Taiwan

Even with genome-wide data, *C. formosanus* did not show any sign of population genetic structure according to geographical origins (Fig. 4). Together with its known distribution and COXI haplotype, the results suggest a panmictic population with great dispersal ability. However, gordiids disperse poorly by themselves when they are at the larval stage (Bolek *et al.*, 2015; Chiu *et al.*, 2016; Chiu, 2017) and the river network in Taiwan prevents non-flying organisms associated with freshwater from dispersing, given its fragmentation (Shih *et al.*, 2006). That would block potential dispersal by river flow at a country-wide level.

Furthermore, the presence of this species in Lyudao (Chiu *et al.*, 2011), a volcanic island that has never been connected to the main island of Taiwan (Yang *et al.*, 1996), implies sea crossing. Previous modelling efforts noticed a very high similarity between the mantis hosts' range and the range of *C. formosanus*, hinting at dispersal by definitive hosts, although the dispersal by paratenic ones cannot be excluded (De Vivo and Huang, 2022; Sup. Info). Therefore, dispersal by active dispersing flying insect hosts (paratenic and/or final ones) with or without a multigenerational process can be a possible explanation for this lack of geographic structure shown in our results.

Demographic history inferences

The demographic analyses using the COXI datasets were able to trace historical demography through 2 million years for *C. formosanus* (with the Taiwanese population going a little bit deeper in time). The same approach revealed N_e trends up to roughly 6

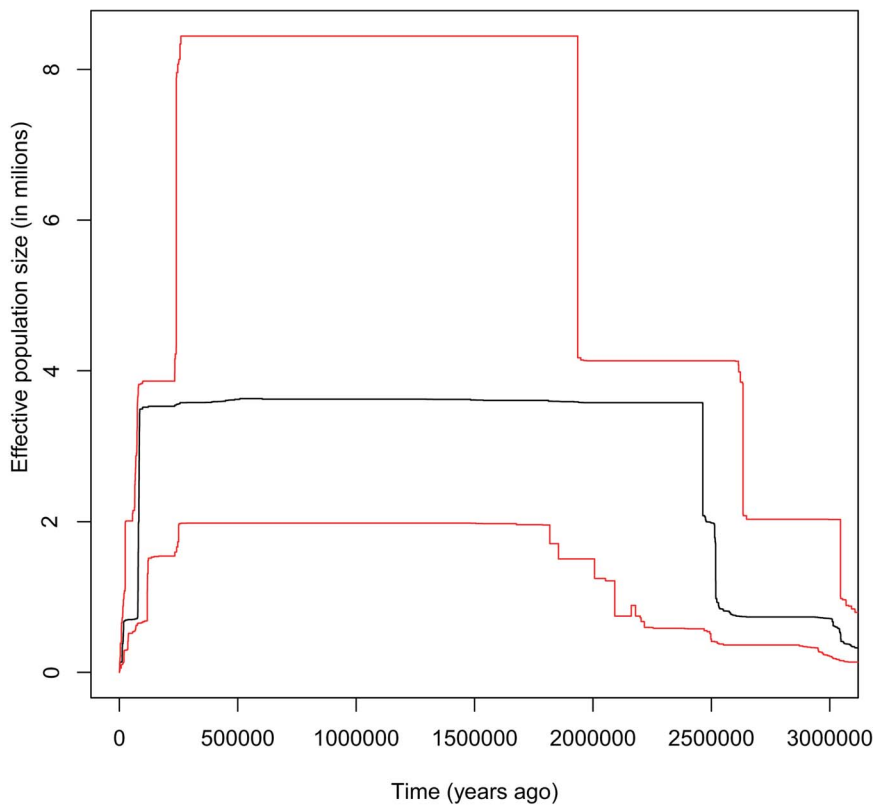
***C. formosanus*'s demographic history**

Figure 6. Stairway plot for *Chordodes formosanus*. The original version and the ones for *Acutogordius taiwanensis* and *Gordius chiashanus* are available as Supplementary Figs 4–6.

million years ago for *G. chiashanus* and 250,000 years of demographic history for *A. taiwanensis*. This difference in scale may have been caused by the lower variability of the *A. taiwanensis* sequences (7 variable sites in total), which in turn was caused by a limited geographic sampling. In general, it is known that mitochondrial sequence data usually tend to recover demographic histories in the Pleistocene (Ho and Shapiro, 2011); however, such data are less informative with recent history compared to RADseq (Nunziata and Weisrock, 2018). The COXI-based Bayesian Skyline Plots revealed recent drops in N_e for both Taiwanese and Japanese populations of *C. formosanus* over the last 250,000 years. Several drops in the same period were found by Stairway Plot using the ddRADseq data. Additionally, both approaches found a demographic increase a little bit before 2 million years ago for the Taiwanese population. That said, Stairway Plot analyses with ddRADseq data were able to detect more events in the last 100,000 years for the Taiwanese population of *C. formosanus*, compared to the Coalescent Bayesian Sky Plot based on COXI sequences, probably due to the differences in number of loci between the 2 datasets (e.g. Cristofari *et al.*, 2018) and the different mutation rate in mitochondrial genes (Ho and Shapiro, 2011).

The demographic histories reconstructed using genome-wide SNP data and the Stairway Plot were very large for *A. taiwanensis* and *G. chiashanus*, but at the same time, the software failed to recover any possible trends in recent years. This was caused by the very small sample size, which were 2 diploid individuals per species. Specifically, the number of historical events that can be estimated using Stairway Plot, and any programs that take site frequency spectrum as inputs and calculate composite likelihoods, is constrained to the number of site frequency categories. As a result, for a folded site frequency spectrum, as implemented here, the use of 2 individuals will only result in 2 site frequency categories, and thus only 2 historical demographic episodes can

be estimated (Liu and Fu, 2020). The 2 species also have extremely large N_e estimated based on the Stairway Plot results. It is known that sample size can impact the estimated demographic parameters using RADseq data (Nunziata and Weisrock, 2018) and therefore we attribute the high N_e estimates to the sampling size, which makes the inferences unreliable.

For *C. formosanus*, however, more events were recorded from the Stairway Plot analyses. That said, it is surprising that this species, which is regarded as the most common in Taiwan, is showing signs of decline – although the current N_e should still be large enough to sustain the population (Pérez-Pereira *et al.*, 2022). The recent demographic decrease might be the result of extirpation events, possibly caused by changes in land use in Taiwan, hypothesized after niche modelling in De Vivo and Huang (2022). It is known that land-use changes can influence helminths, although the effects seem to depend on both host's and parasite's ecology (e.g. Chakraborty *et al.*, 2019; Portela *et al.*, 2020).

Future directions

In this study, we evaluated the use of different genetic and analytical tools for estimating demographic history and population genetic structure, which informed us about potential declines and dispersal traits. From a conservation perspective, the negative trend shown by *C. formosanus* would deserve more attention in the future, given its confirmation by both Coalescent Skyline and Stairway Plots. *A. taiwanensis* should also be monitored, due to its known smaller geographical range compared to the other 2 evaluated species and potential sensitivity to anthropic changes (Sup. Info). For *G. chiashanus*, additional samples should be collected for testing the inferred decline shown by COXI Coalescent Skyline Plots with genome-wide data too and see if recent (last 100,000 years) declines happened. In our datasets,

sampling bias for some taxa was present and can influence the results. For possible future studies on the conservation genetics of parasites, we give 3 suggestions that can help with evaluating a parasite's conservation status:

- (i) consider the target species' ecologies, given that they can highly influence their population genetic structure (van Schaik *et al.*, 2015; Radačovská *et al.*, 2022), *Ne* estimates (Criscione *et al.*, 2005; Strobel *et al.*, 2019) as well as the sampling strategy, given that some parasites have life cycles that can influence when and how to collect them (e.g. van Schaik *et al.*, 2015);
- (ii) try to get the geographically broadest and biggest sample size possible; for example, in this study a small sample size in the ddRADseq dataset did not allow us to estimate enough recent trends for 2 taxa, while the COXI dataset for *A. taiwanensis* was too biased to one area, which did not allow us to properly estimate the diversity of the species;
- (iii) define the molecular and bioinformatic tools used; a direct estimate of *Ne* for parasites can be tricky to calculate (e.g. Strobel *et al.*, 2019; Carlson *et al.*, 2020). Given this, we suggest focusing on trends instead of raw numbers. Additionally, previous studies showed possible limits of single-locus analyses (Ho and Shapiro, 2011). Therefore, we suggest using as many loci as possible (see Peterson *et al.*, 2012 and Nunziata and Weisrock, 2018 for potential protocols). However, single-locus data are often the kinds of data that are the most available for parasites (Selbach *et al.*, 2019). Therefore, if budget and resources are limited, Sanger sequencing can be pursued (but see Radačovská *et al.*, 2022 for caveats with mitochondrial sequences in polyploid species). The use of depositories such as GenBank can be useful for retrieving previously released sequences and therefore increase both sample size and loci used while utilizing software such as BEAST2 that can use such data. That said, it is crucial to check for a possible population genetic structure before demographic trend analyses, since it influences the results (Ho and Shapiro, 2011).

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0031182023000641>

Acknowledgements. We would like to thank all the people who helped us collect samples all around Taiwan. We would also like to thank Ming-Chung Chiu for helping us with designing the sampling strategy and discussing the paper and the biology of the involved species, Hsiang-Yun Lin for sampling assistance, Justin Pelofsky for revising the grammar of this article, Xiaoming Liu for helping us with Stairway Plot 2's settings, the DNA Sequencing Core Facility of the Institute of Biomedical Sciences, Academia Sinica for Sanger sequencing, the NGS High Throughput Genomics Core at Biodiversity Research Center, Academia Sinica, for providing Illumina sequencing and the anonymous reviewers who gave us suggestions on the manuscripts.

Authors' contributions. MDV extensively sampled the specimens around the country, performed the bioinformatic analyses and wrote the original draft. MDV and WYC performed molecular work. MDV and JPH conceived and designed the study. All the authors designed the methodology, corrected the original draft and approved the manuscript's content.

Financial support. MDV was supported by a scholarship from Taiwan International Graduate Program (TIGP) and from TIGP Research Performance Fellowship 2022. JPH was supported by a grant from Ministry of Science and Technology, Taiwan (MOST 108-2621-B-001-001-MY3) and internal research support from Academia Sinica.

Competing interests. None.

Ethical standards. Not applicable.

Data availability. The COXI sequences produced in this study are available in GenBank (accession numbers: OQ121045- OQ121084). The ddRAD Stacks-trimmed reads are available in Sequence Read Archive (BioProject number: PRJNA914055). Supplementary Data (i.e. logs and outputs) for this article can be found in Zenodo (<https://zenodo.org/record/7659651>, doi: 10.5281/zenodo.7659651). A preprint of this article, based on the manuscript's first version, is available on BioRxiv (<https://www.biorxiv.org/content/10.1101/2023.02.21.529467v1>, doi:10.1101/2023.02.21.529467).

References

- Akaike H (1998) Information theory and an extension of the maximum likelihood principle. In Parzen E, Tanabe K and Kitigawa G (eds), *Selected Papers of Hirotugu Akaike*. New York, USA: Springer, pp. 199–213.
- Anaya C, Hanelt B and Bolek MG (2021) Field and laboratory observations on the life history of *Gordius terrestris* (phylum Nematomorpha), a terrestrial nematomorph. *The Journal of Parasitology* **107**, 48–58.
- Beugin MP, Gayet T, Pontier D, Devillard S and Jombart T (2018) A fast likelihood solution to the genetic clustering problem. *Methods in Ecology and Evolution* **9**, 1006–1016.
- Bolek MG, Schmidt-Rhaesa A, De Villalobos LC and Hanelt B (2015) Phylum Nematomorpha. In Thorp J and Rogers DC (eds), *Thorp and Covich's Freshwater Invertebrates*. Cambridge, Massachusetts, USA: Academic Press, pp. 303–326. <https://doi.org/10.1016/B978-0-12-385026-3.00015-2>.
- Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, Matschiner M, Mendes FK, Müller NF, Ogilvie HA, du Plessis L, Popinga A, Rambaut A, Rasmussen D, Siveroni I, Suchard MA, Wu CH, Xie D, Zhang C, Stadler T and Drummond AJ (2019) BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology* **15**, e1006650.
- Carlson CJ, Hopkins S, Bell KC, Doña J, Godfrey SS, Kwak ML, Lafferty KD, Moir ML, Speer KA, Strona G, Torchin M and Wood CL (2020) A global parasite conservation plan. *Biological Conservation* **250**, 108596.
- Chakraborty D, Reddy M, Tiwari S and Umapathy G (2019) Land use change increases wildlife parasite diversity in Anamalai Hills, Western Ghats, India. *Scientific Reports* **9**, 11975.
- Chiu MC (2017) Biodiversity of the Taiwanese horsehair worms and the host morphological development manipulated by infection (PhD thesis). National Taiwan University, Taipei, Taiwan.
- Chiu MC, Huang CG, Wu WJ and Shiao SF (2011) A new horsehair worm, *Chordodes formosanus* sp. n. (Nematomorpha, Gordiida) from *Hierodula* mantids of Taiwan and Japan with redescription of a closely related species, *Chordodes japonensis*. *ZooKeys* **160**, 1–22.
- Chiu MC, Huang CG, Wu WJ and Shiao SF (2016) Annual survey of horsehair worm cysts in northern Taiwan, with notes on a single seasonal infection peak in chironomid larvae (Diptera: Chironomidae). *The Journal of Parasitology* **102**, 319–326.
- Chiu MC, Huang CG, Wu WJ and Shiao SF (2017) A new orthopteran-parasitizing horsehair worm, *Acutogordius taiwanensis* sp. n., with a redescription of *Chordodes formosanus* and novel host records from Taiwan (Nematomorpha, Gordiida). *ZooKeys* **683**, 1–23.
- Chiu MC, Huang CG, Wu WJ, Lin ZH, Chen HW and Shiao SF (2020) A new millipede-parasitizing horsehair worm, *Gordius chiashanus* sp. nov., at medium altitudes in Taiwan (Nematomorpha, Gordiida). *ZooKeys* **941**, 25–48.
- Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J and Sayers EW (2016) GenBank. *Nucleic Acids Research* **44**, D67–D72.
- Clement M, Posada D and Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**, 1657–1659.
- Criscione CD, Poulin R and Blouin MS (2005) Molecular ecology of parasites: elucidating ecological and microevolutionary processes. *Molecular Ecology* **14**, 2247–2257.
- Cristofari R, Liu X, Bonadonna F, Cherel Y, Pistorius P, Le Maho Y, Raybaud V, Stenseth NC, Le Bohec C and Trucchi E (2018) Climate-driven range shifts of the king penguin in a fragmented ecosystem. *Nature Climate Change* **8**, 245–251.
- Dalapiccola J, do Prado JR, Percequillo AR and Knowles LL (2021) Functional connectivity in sympatric spiny rats reflects different dimensions of Amazonian forest-association. *Journal of Biogeography* **48**, 3196–3209.

- De Vivo M and Huang JP** (2022) Modeling the geographical distributions of *Chordodes formosanus* and its mantis hosts in Taiwan, with considerations for their niche overlaps. *Ecology and Evolution* **12**, e9546.
- Doherty JF, Filion A and Poulin R** (2022) Infection patterns and new definitive host records for New Zealand gordiid hairworms (phylum Nematomorpha). *Parasitology International* **90**, 102598. doi: 10.1016/j.parint.2022.102598
- Dougherty ER, Carlson CJ, Bueno VM, Burgio KR, Cizauskas CA, Clements CF, Seidel DP and Harris NC** (2016) Paradigms for parasite conservation. *Conservation Biology* **30**, 724–733.
- Eaton DA and Overcast I** (2020) ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics (Oxford, England)* **36**, 2592–2594.
- Eidler D, Klein J, Antonelli A and Silvestro D** (2021) raxmlGUI 2.0: a graphical interface and toolkit for phylogenetic analyses using RAxML. *Methods in Ecology and Evolution* **12**, 373–377.
- Folmer O, Black M, Hoeh W, Lutz R and Vrijenhoek R** (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3**, 294–299.
- Ho SYW and Shapiro B** (2011) Skyline-plot methods for estimating demographic history from nucleotide sequences. *Molecular Ecology Resources* **11**, 423–434.
- Howard RJ, Giacomelli M, Lozano-Fernandez J, Edgecombe GD, Fleming JF, Kristensen RM, Ma X, Olesen J, Sørensen MV, Thomsen PF, Wills MA, Donoghue PCJ and Pisani D** (2022) The Ediacaran origin of Ecdysozoa: integrating fossil and phylogenomic data. *Journal of the Geological Society* **179**. doi: 10.1144/jgs2021-107
- Katoh K and Standley DM** (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772–780.
- Konrad A, Brady MJ, Berghorsson U and Katju V** (2019) Mutational landscape of spontaneous base substitutions and small indels in experimental *Caenorhabditis elegans* populations of differing size. *Genetics* **212**, 837–854.
- Kumar S, Stecher G, Li M, Knyaz C and Tamura K** (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **35**, 1547–1549.
- Lee SJ, Hwang JM and Bae YJ** (2008) Life history of a lowland burrowing mayfly, *Ephemera orientalis* (Ephemeroptera: Ephemeridae), in a Korean stream. *Hydrobiologia* **596**, 279–288.
- Leigh JW and Bryant D** (2015) POPART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* **6**, 1110–1116.
- Lischer HE and Excoffier L** (2012) PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics (Oxford, England)* **28**, 298–299.
- Liu X and Fu YX** (2020) Stairway Plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biology* **21**, 280. doi: 10.1186/s13059-020-02196-9
- Malinsky M, Trucchi E, Lawson DJ and Falush D** (2018) RADpainter and fineRADstructure: population inference from RADseq data. *Molecular Biology and Evolution* **35**, 1284–1290.
- Nei M and Li WH** (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences* **76**, 5269–5273.
- Nunziata SO and Weisrock DW** (2018) Estimation of contemporary effective population size and population declines using RAD sequence data. *Heredity* **120**, 196–207.
- Pérez-Pereira N, Wang J, Quesada H and Caballero A** (2022) Prediction of the minimum effective size of a population viable in the long term. *Biodiversity Conservation* **31**, 2763–2780.
- Peterson BK, Weber JN, Kay EH, Fisher HS and Hoekstra HE** (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* **7**, e37135.
- Petit-Marty N, Vázquez-Luis M and Hendriks IE** (2021) Use of the nucleotide diversity in COI mitochondrial gene as an early diagnostic of conservation status of animal species. *Conservation Letters* **14**(1), e12756. doi: 10.1111/conl.12756
- Poinar G and Buckley R** (2006) Nematode (Nematoda: Mermithidae) and hairworm (Nematomorpha: Chordodidae) parasites in early Cretaceous amber. *Journal of Invertebrate Pathology* **93**, 36–41.
- Portela AAB, dos Santos TG and dos Anjos LA** (2020) Changes in land use affect anuran helminths in the South Brazilian grasslands. *Journal of Helminthology* **94**, e206.
- Radačovská A, Čisovská Bazsalovicsová E, Šoltys K, Štefka J, Minárik G, Gustinelli A, Chugunova JK and Králová-Hromádová I** (2022) Unique genetic structure of the human tapeworm *Dibothriocephalus latus* from the Alpine lakes region – a successful adaptation? *Parasitology* **149**, 1106–1118.
- Rivera-Colón AG and Catchen J** (2022) Population genomics analysis with RAD, reprised: stacks 2. In Verde C and Giordano D (eds), *Marine Genomics*. Humana, New York, USA: Springer, pp. 99–149. doi: 10.1007/978-1-0716-2313-8_7
- Rochette NC, Rivera-Colón AG and Catchen JM** (2019) Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology* **28**, 4737–4754.
- Sato T, Watanabe K, Fukushima K and Tokuchi N** (2014) Parasites and forest chronosequence: long-term recovery of nematode parasites after clear-cut logging. *Forest Ecology and Management* **314**, 166–171.
- Schwarz G** (1978) Estimating the dimension of a model. *Annals of Statistics* **6**, 461–464.
- Selbach C, Jorge F, Dowe E, Bennett J, Chai X, Doherty JF, Eriksson A, Filion A, Hay E, Herbison R, Lindner J, Park E, Presswell B, Ruehle B, Sobrinho PM, Wainwright E and Poulin R** (2019) Parasitological research in the molecular age. *Parasitology* **146**, 1361–1370.
- Shih HT, Hung HC, Schubart CD, Chen CA and Chang HW** (2006) Intraspecific genetic diversity of the endemic freshwater crab *Candiodipotamon rathbunae* (Decapoda, Brachyura, Potamidae) reflects five million years of the geological history of Taiwan. *Journal of Biogeography* **33**, 980–989.
- Strobel HM, Hays SJ, Moody KN, Blum MJ and Heins DC** (2019) Estimating effective population size for a cestode parasite infecting three-spined sticklebacks. *Parasitology* **146**, 883–896.
- Tajima F** (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.
- Tobias ZJ, Yadav AK, Schmidt-Rhaesa A and Poulin R** (2017) Intra- and interspecific genetic diversity of New Zealand hairworms (Nematomorpha). *Parasitology* **144**, 1026–1040.
- van Schaik J, Dekeukeleire D and Kerth G** (2015) Host and parasite life history interplay to yield divergent population genetic structures in two ectoparasites living on the same bat species. *Molecular Ecology* **24**, 2324–2335.
- Weller AM, Rödelberger C, Eberhardt G, Molnar RI and Sommer RJ** (2014) Opposing forces of A/T-biased mutations and G/C-biased gene conversions shape the genome of the nematode *Pristionchus pacificus*. *Genetics* **196**, 1145–1152.
- Yang TF, Lee T, Chen CH, Cheng SN, Knittel U, Punongbayan RS and Rasdars AR** (1996) A double island arc between Taiwan and Luzon: consequence of ridge subduction. *Tectonophysics* **258**, 85–101.