Spatial diversity of the endemic New Zealand mayfly lineage *Ichthybotus* (Ephemeroptera: Ichthybotidae)

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Science in Ecology

By

Charlotte Bridger

Massey University
Palmerston North
2022

Table of Contents:

Acknowledgments

My first and biggest thank you must go to my supervisors Steve Trewick and Mary Morgan-Richards. You two have been wonderful and I couldn't have asked for better support and guidance. You have taught me so much and made me feel so welcome as a Massey postgraduate. I really appreciate your constant encouragement and humour, from the start of this project until the end.

To Dr Steve Pohe, I appreciate you trusting me with your precious Ichthybotus specimens and allowing me to build on the amazing research you have done. Without your meticulous work with New Zealand mayflies, I would never have found my way to this project.

I was honoured to receive funding from the Graduate Women New Zealand Manawatu Postgraduate Scholarship. This helped me to fund my research and allowed me to spend less time working retail, more time chopping off mayfly legs.

Thanks so much to the Phoenix Group. Your passion and hearing about all your exciting research (both successes and setbacks) has really inspired me and I wish you all the best with the rest of your work. Knowing that you guys would always be there to help if I needed took a huge weight off my shoulders and made me feel less lost in the wild world of academia. I want to thank David and Leo for their patience and willingness to sit with me while I tried to get my head around all my R code errors!

To my friends and office mates within the ecology group: Ben, Fern, Ruby, and Caitlin. You guys were a highlight of this whole experience. Thank you so much to Libby and Polly. Not sure how to thank you for 15 years of friendship and emotional support. Thanks for giving me a life outside of this thesis, and always being there for me when I need you. I love you guys.

I'm so lucky to have been surrounded by such a supportive family. Special thank you to my Aunty Ally for letting me take over your house for a year. Thanks, Mum, for being a text away, and for allowing me a much-needed break.

Finally, this thesis is dedicated to Dad. I can't even begin to put into words how thankful I am for all your love, support, and sacrifice. You are the reason for my love for learning and research, spending hours and hours helping with collecting woodlice in high school, to mayflies for this thesis. You are an absolute wonder to me, and I love you so much.

Figures and Tables Index

Figures

Chapter 1:

Chapter 2:

Chapter 3:

- **Figure 3.5.** Haplotype networks of *I. hudsoni* (North Island) and *I. bicolor* (South Island) (617 bp of COI mtDNA).

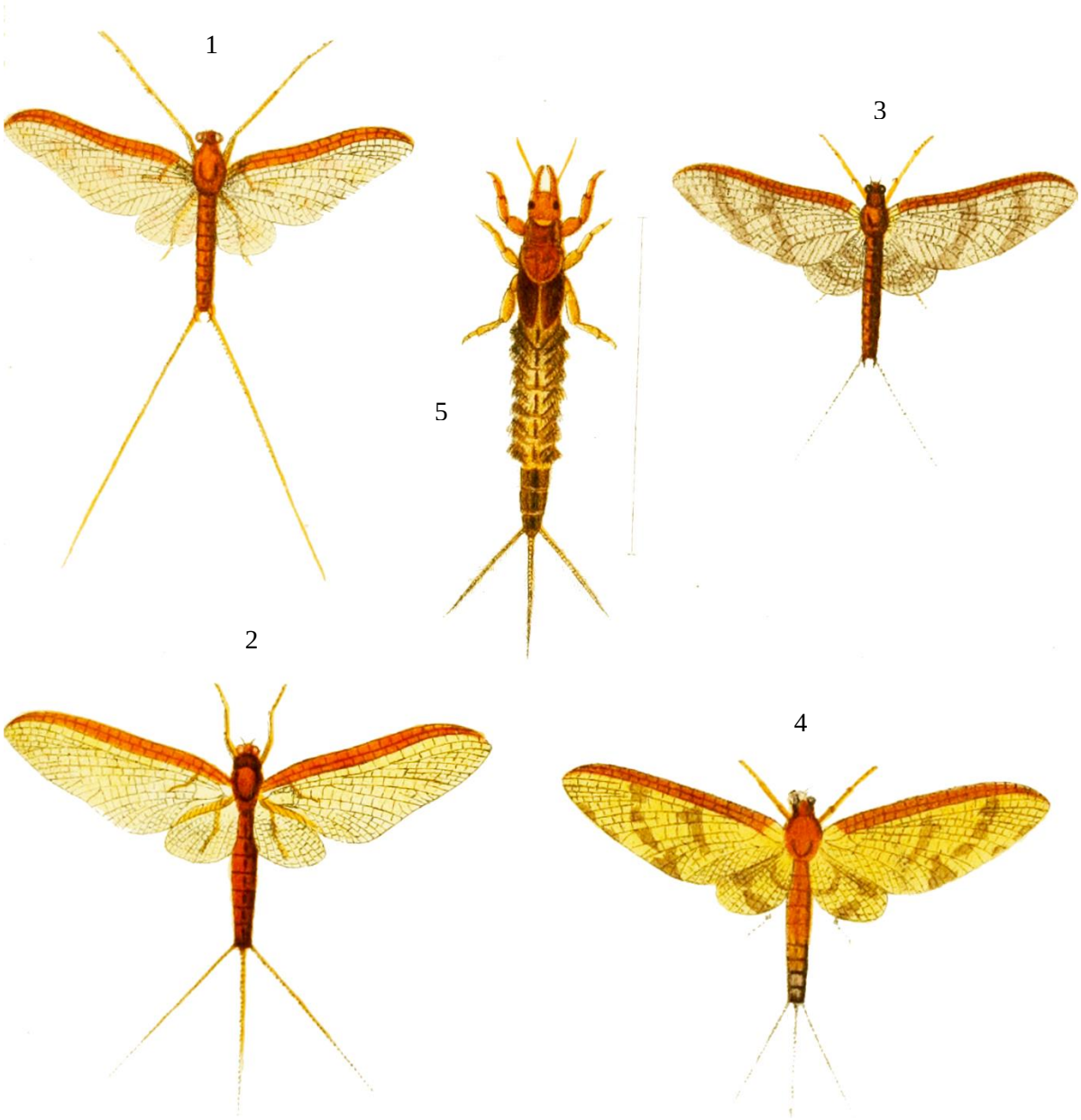
Tables

Chapter 2

- **Table 2.1.** Variance inflation factor for the bioclimatic variables retained in ecological niche models of two species of New Zealand burrowing mayfly *Ichthybotus.*
- **Table 2.2.** Ecological Niche model fit statistics for the New Zealand mayflies, *Ichthybotus hudsoni* and *I. bicolor*. Upper: True Skill Statistic (TSS) for each mayfly species and model (GLM, GBM, GAM, CTA, ANN, RF) included in ensemble model, across 5 runs with the mean of the 5 runs. Lower: Relative Operating Characteristic (ROC) for each mayfly species and model (GLM, GBM, GAM, CTA, ANN, RF) included in the ensemble model, across 5 runs with the mean of the 5 runs.
- **Table 2.3.** Ensemble model evaluation statistics: the score, the threshold (for transforming continuous likelihoods into binary data with an optimised score), sensitivity (% of presences correctly predicted) and specificity (% absences correctly predicted).
- **Supplementary Table 2.1.** Shows all site locations (164) and their co-ordinates, sorted form northern to southernmost. Includes presence (1) and absence (0) of species and location and the source of the data.
- **Supplementary Table 2.2.** Codes for the 19 bioclimatic variables from WorldClim.

Chapter 3

- **Table 3.1.** Summary of variation among 617bp mtDNA COI for population samples of *I. hudsoni and I. bicolor*, organised by island and region. Shows locations (ordered from north to south) and genetic diversity indices for each site and species, and total data estimates (π, nucleotide diversity; h, number of haplotypes; k, average number of nucleotide differences; Hd, haplotype diversity).
- **Table 3.2.** Table of pairwise population data used for *I. hudsoni* mantel test.
- **Supplementary Table 3.1.** North Island *I. hudsoni* CO1 haplotypes (73) and their locations.
- **Supplementary Table 3.2.** South Island *I. bicolor* CO1 haplotypes (22) and their locations.



Ichthybotus hudsoni after Hudson (1904) New Zealand Neuroptera: 1 male imago, 2 female imago, 3 male sub-imago, 4 female sub-imago, 5 nymph.

Chapter 1: General Introduction

Mayflies (Ephemeroptera) are insects with a global distribution, only absent from Antarctica. As nymphs they live in fresh water, with a brief, non-feeding, phase as winged adults. The Order Ephemeroptera is a lineage that goes back over 300 million years with 42 extant families and over 3,000 living species (Barber-James et al., 2008). Mayflies provide many ecosystem services as they clean freshwater, and support many processes such as decomposition, nutrient cycling. Their presence and abundance in waterways provide useful indicators of water quality, pollution, and ecosystem health (Ogden and Whiting, 2005; Jacobus et al., 2019). Culturally they provide artistic inspiration and have festivals dedicated to them reflecting the cyclic and ephemeral nature of their appearance. Mayflies also form a valuable human food source in India, Mexico, Malawi, Japan, China, Papua New Guinea, Kenya, Tanzania, and Uganda, having one of the highest protein contents of any edible insect (Jocobus et al. 2019; Zhao et al. 2021).

New Zealand has 57 endemic species of mayfly including three endemic families (Pohe 2018; Trewick et al. 2022). There are high levels of endemism in the whole of the freshwater invertebrate community in New Zealand where rivers are often short, steep, and fast flowing due to the country's current mountainous terrain. Disturbances are frequent in New Zealand waterways, reflected in dynamic invertebrate community structures (Thompson & Townsend, 2000). Stream invertebrate composition is considered a key indicator of anthropogenic land use because of sensitivity to changes in environmental attributes including nutrient and sediment loading and temperature (Death and Winterbourn, 1995).

1.1 Study Species

The endemic Ichthybotidae are burrowing mayflies, (Hitchings, 2008) with just two endemic New Zealand species that are amongst the largest and most colourful mayflies in the country (Hitchings, 2008; Pohe, 2019). There are two species within the family; *Ichthybotus hudsoni*, which appears to be restricted to the North Island; and *Ichthybotus bicolor*, restricted to low elevations in the South Island, particularly the northwest (Hitchings, 2008; Pohe, 2019). They are semi-aquatic insects, differentiated by the subtle colour difference of the wings and geographic separation by Cook Strait (Figure 1.1) (Chrisholm, 1984; Pohe, 2019).

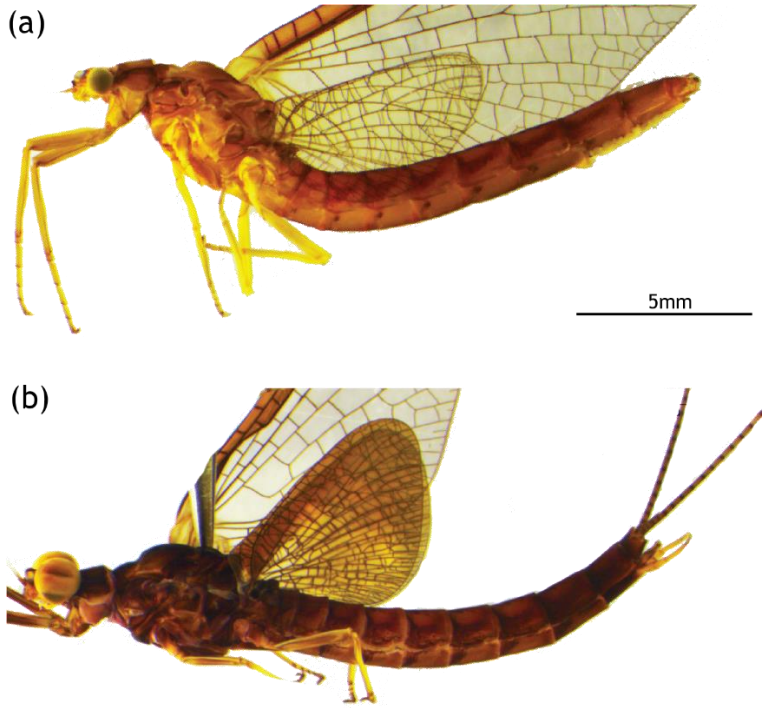


Figure 1.1. Side-by-side of wing and body colour difference between (a) female imago *I. hudsoni*, (b) male imago *I. bicolor*.

1.1.1. Taxonomy

The most recent description of the taxonomy of the *Ichthybotus* species' is as follows (Pohe, 2018):

Class: Insecta
Order: Ephemeroptera
Suborder: Furcatergalia
Family: Ichthybotidae (Demoulin, 1957)
Genus: *Ichthybotus*
Species: *Ichthybotus hudsoni* and *Ichthybotus bicolor*

The species *Ichthybotus hudsoni* was described by McLachlan in 1894 based on *I. hudsoni* specimens in the Wellington area (McLachlan, 1894). It was initially placed in the northern hemisphere genus with other similar burrowing mayflies, but Eaton (1899) established the endemic New Zealand genus *Ichthybotus* for it. In 1923 Tillyard described the second species, *I. bicolor*, in the Nelson area. The treatment of *Ichthybotus* as two species rests upon a difference in adult wing colouration (Tillyard, 1923; Chrisholm 1984), specifically the darker hindwings of adult *I. bicolor* although *I. bicolor* also has noticeably darker colouring overall (Figure 1.1). Some studies suggest that there are differences in nymphal habitat, morphology, and size of the two species (Phillips, 1930; Tillyard, 1923), however these distinctions remain ambiguous (Chrisholm, 1984). *Ichthybotus* is currently considered restricted to New Zealand (McCafferty, 1973) and though initially assigned to the Ephemeridae family, the

morphology of *Ichthybotus* wings and tusks was considered sufficient to differentiate it from other genera and the two species in a family of their own, Ichthybotidae (Demoulin 1975).

1.1.2. Feeding

The nymphs of most burrowing mayflies are sediment/deposit feeders with a few being filter feeders. *Ichthybotus hudsoni* was classified by Chrisholm (1984) using Cummins 1973 classification as a fine particle detrivore or a collector rather than a true filter feeder. Their diets consist of mostly organic detritus (possibly around 85-95%). Their two large mandibular tusks may not have a role in feeding, rather burrowing. The gills may also assist in feeding by creating an admittedly weak feeding current that directs detritus. Like all mayflies, adult *Ichthybotus* do not have functional mouthparts. They rely on food stored during previous life stages (Sartori & Brittain, 2015).

1.1.3. Habitat

Both *Ichthybotus* species prefer slow moving water and relatively stable substrate for burrowing. They also tend to live in undisturbed native forest regions. Ideal habitat is relatively uncommon in New Zealand due to relatively high proportions of mountainous terrain (increasing water movement) and lowland areas being modified for agriculture, altering substrate, and run off among other factors. *Ichthybotus* can burrow into mud, sand, and gravel substrates however they prefer coarse substrate (Chrisholm, 1984). Coarse substrate allows for the flow of oxygen through burrows. During Chisholm's study at Gollans Stream nymphs were only found in coarse substrate. *Ichthybotus* sensitivity makes them good indicators of water quality.

1.1.4. Life Cycle

Ichthybotus have a two-to-three-year life cycle (Phillips, 1930; Hopkins, 1976; Chrisholm, 1984). Their life cycles appear to be flexible; allowing for faster development where environmental conditions and food resources are ideal, and lengthened development where conditions are not ideal (Sartori & Brittain, 2015). Their lifecycle has three stages: nymph, subimago and adult, with subimago being unique to mayflies. The aquatic nymphal life stage is dominant i.e., they spend most of their lives as eggs or nymphs (Sartori & Brittain, 2015). They undergo various moults then emerge, transitioning from aquatic nymph to terrestrial winged subimago (Chrisholm, 1984). The adult stage lasts anywhere between a few hours to weeks, in which mating, and oviposition occurs. Only the subimago and adult stage allow for dispersal between water catchments.

1.2. Climate Change and the history of the Cook Strait

Spanning from 34°25'S - 47°20'S latitude, the country of New Zealand has varied climatic conditions (Wallis & Trewick, 2009). Most invertebrate species are not widespread across New Zealand and there is high regional endemicity (Taylor-Smith et al., 2019). Climate is one important factor in determining

current species distributions. The effects of past climate change will offer insight into how species might respond to ongoing climate change (Dawson et al, 2011). Many species experience range shifts because of changing climate, which is expected to increase with continuing climate change (Chen et al., 2011; Garcia et al 2014). Glacial cycles have significant effects on the distribution of species, with the Last Glacial Maximum (LGM) occurring around 20,000 – years ago during the last phase of the Pleistocene epoch (Suggate & Almond, 2005; Trewick et al., 2011). The LGM caused glaciation in the western South Island, covering 30% of the island and lowering sea levels (Suggate & Almond, 2005; Wallis and Trewick, 2009). The glaciers in the South Island had nunataks and valleys that remained unglaciated, which could have served as fragmented refugia (Wallis & Trewick, 2009). During the glacial cycles forest cover is likely to have been reduced to coastal and warmer northern part of New Zealand (Newnham et al. 2013). Meanwhile, Cook Strait was bridged from south Taranaki to northwest Nelson (Hitchings, 2008; Trewick & Bland, 2012). The strait was reformed as glaciation ended (~21,000) and the sea level rose, becoming a marine-barrier to dispersal of terrestrial population. This led to the separation of populations of some species, i.e., the separation of northern and southern weka lineages (Trewick et al., 2017), while other taxa show little or no divergence concordant with the strait (Dumbleton, 1970, Pohe 2019). Chrisholm (1984) suggests that the land bridges that formed between the islands during the Pleistocene ice ages may not have been suitable to the *Ichthybotus* species. i.e., streams on the bridge did not provide the gravel bottom the nymphs need or forest was too open. Therefore, the two *Ichthybotus* species may have been separated since before the Pleistocene and never reconnected during the land bridges, or they were connected but separated by the most recent separation.

1.3. Significance

The lack of distribution and ecological knowledge are some of the biggest gaps to freshwater invertebrate conservation and management (Drinan et al., 2020). The response of aquatic invertebrates to past and ongoing climate change is poorly understood (Drinan et al., 2020). Current species distributions will reflect past geological and climatic history, and the present-day habitat (Pohe, 2019). Being able to infer and model how future climate change will impact freshwater invertebrates could help us mitigate biodiversity loss.

1.4. Outline

I examine the spatial structure of these two endemic species of burrowing mayfly, *Ichthybotus hudsoni* and *Ichthybotus bicolor*, to explore the influence of climate and geography on their current and past distributions in New Zealand. In Chapter 2, I use an ecological niche modelling approach to determine whether climatic variables can be used to explain the current distribution of these two species. Ecological niche models will be projected onto the inferred climatic conditions of the last glacial maximum of the Pleistocene to estimate the past distribution of the two species. This was a time when

the two main islands of New Zealand were connected the two mayfly species may have been in contact with one another. Then in Chapter 3 I use mtDNA variation phylogeographic analysis to determine whether the two species are genetically distinct lineages. Patterns of variation within widespread species can provide evidence of population history that might reflect changes in potential habitat such as forest cover since the last glacial maximum. If I detect higher genetic diversity at lower latitudes of this mayfly's range compared to most of its distribution this would suggest southward range expansion during the current post glacial period. Finally, Chapter 4 is a general discussion of the results of both chapters 2 and 3.

Chapter 2: Environmental envelope modelling of two burrowing mayflies during current and LGM climates

2.1. Introduction

2.1.1. Niche Concepts

Niche can be defined as the sum of the environmental characters that determine the position of an organism in its ecosystem and is usually considered a characteristic of a species. Niche includes all the conditions and factors required for population growth rate to be positive. The idea of an ecological niche has developed since its origins (Grinnell, 1917a; Grinnell, 1917b) which focused on abiotic factors of population growth, and eventually begun to include biotic factors, before being defined as a geographical combination of the two: an n-dimensional hypervolume defined by environmental and resource requirements (Hutchinson 1957, Silvertown 2004). The niche exists within the intersection of the realised environmental space and the species fundamental niche (Jackson & Overpeck 2000). The fundamental niche is the environmental conditions that ensure a positive population growth, excluding biotic factors (Pearman et al. 2007). This intersection is the potential niche space, which is restricted once biotic factors are considered. This is called the realised niche; the area within the fundamental niche that ensures a positive population growth rate, given the impact of biotic interactions (e.g., competition) (Jackson & Overpeck 2000; Silvertown 2004, Pearman et al. 2007). A recent concept is how species can also modify their environment to improve their niche, coined niche construction (Olding-Smee et al. 2003; Kylafis & Loreau, 2011). Changes to climate variables changes the conditions of the realised environmental space. Niche also considers the role that a species plays in its community by occupying a space and utilizing resources. The Grinnellian and Soberon and Peterson concepts of niche align best with the aims of this chapter. Both focus on the impact of large spatial scale environmental and climatic factors (Sales et al. 2021, Stevens et al. 2021). Soberon and Peterson explore distribution geographically across grid cells each with ‘n’ environmental features (Pulliam 2000, Sales et al. 2021).

2.1.2. Ecological Niche Models

Ecological Niche Models (ENMs), or species distribution models, are statistical tools for predicting the potential niche space that species can occupy across space and time (Pearson & Dawson 2003; Pearman et al. 2007; Phillips & Dudik 2008; Thuiller 2003; Elith & Leathwick. 2009; Valencia-Rodriguez at al 2021). With the increase in modelling software and accessible climate variables ENMs have become more popular in distributional ecology literature. They are correlative models which can use the known distribution of a species occurrences and environmental variables to determine which variables are potentially limiting their distribution. These models can also provide insight into range shifts or species

distribution overlap. Individual methods of ENMs can be combined to run ensemble models. A recent development, these models are more accurate and have high predictive ability (Araujo & New 2006a).

Ecological Niche Models have been used in many fields to predict occurrence, range shifts, genetic structure and understand biotic interactions. For example, prioritising and informing conservation efforts (Taubmann et al. 2010). Comparing realised ranges to potential ranges can show potential areas of expansion, or competition (Bulgarella et al. 2013). The accessibility of environmental databases such as WorldClim (Global climate data) (Hijmans et al. 2005) has also contributed to models that examine the effects of past and future climate projections. Models of the present can be applied to these time periods to infer how distributions have changed given the same niche requirements (Taubmann et al. 2010). They have been useful in many different fields, including terrestrial, aquatic, and marine studies. ENM approach differs depending on the field largely due to differences in dispersal ability (Elith & Leathwick, 2009).

2.1.4. Modelling Aquatic Taxa

ENMs have been utilised less for aquatic habitats, compared to terrestrial environments but their application has increased since the early 2000s (Elith & Leathwick, 2009). This is likely due to the challenges associated with collecting occurrence data for species of interest. Local properties of aquatic systems such as depth, flow rate, temperature, substrate (hydrological variables) are important aspects of niche, just as slope, aspect, soil type for terrestrial organisms (Carmelet-Rescan et al, 2021), however, broad climatic data provide valuable proxies for regional conditions. As such climate-based models can reflect local conditions that are not directly measured and provide a window on the impact of large-scale shifts in conditions.

For example, Taubmann et al. (2011) applied MAXENT ENM as part of their study of the European range of the mayfly *Ameletus opinatus*. They aimed to identify areas of suitable long-term habitat and genetic diversity. Their models scored very highly in terms of predictive accuracy and were used to demonstrate decreasing suitable habitat extent. Domisch et al. (2013) used a Biomod ensemble model to predict the distribution of 191 stream macroinvertebrates under two climate change scenarios. They found that freshwater organisms are sensitive to climate change effects, particularly if they are cold adapted. Analysis of the Pleistocene history of a semi-aquatic invertebrate in China showed an expansion of the cold adapted species during the LGM (Ye et al. 2014). This conclusion was supported by phylogeographic analysis. The effectiveness of ENMs for freshwater species was by Valencia-Rodriguez et al. (2021) by modelling the distribution of the freshwater fish *Bycon henni*, which is endemic to the Colombian Andes. They included climate and hydrological variables and concluded that ENMs can be useful for assessing freshwater distributions. (Valencia-Rodriquez et al., 2021).

2.1.5. Chapter Aims

1. Determine whether the distributions of freshwater insects in New Zealand can be modelled using basic climatic variables such as temperature and rainfall.
2. If ENM are plausible, do climate variables explain the lack of overlap between the two species of burrowing mayfly in New Zealand?
3. Project ecological niche models based on current distribution to estimate potential distribution during Last Glacial Maximum. This will help determine how climate has impacted their distribution (i.e., ranges shifts, contractions, expansion) and identify any potential refuges during the LGM.

2.2. Niche Modelling Methods

2.2.1. Collection of *Ichthybotus* location records (Response Variable)

Data on the current distribution (presence and absence) of two *Ichthybotus* species was collected from their full range by taxonomist Dr. Stephen Pohe, and me (Pohe, 2018; Pohe 2019). Sites were included from iNaturalist sightings after identifications were verified; crucial for "gold standard" modelling (Zurell et al. 2020). A nationwide survey was carried out across New Zealand's three largest islands. The initial survey (carried out by Pohe) intended to sample all New Zealand's mayfly species, therefore includes sites where *Ichthybotus* was absent, allowing true absences to be used in the niche modelling. Minimum distance between sites was 1km. Sampling bias was reduced by using methods and locations to ensure a wide range of New Zealand mayflies were caught rather than focusing on assumed *Ichtybotus* spp habitat. The use of seven latitudinal zones across the country ensured Pohe's 83 sampling locations were distributed evenly over New Zealand and not concentrated around easy access points (Figure 2.1 and 2.2). Ensuring that sampling covers the full species range and is not biased is important for modelling realized niche (Zurell et al. 2020). Zones contained locations, which each contained three sites i.e., nested stratified sampling (Zone/Region > Location > Site).

Ichthybotus hudsoni was found in 12 North Island locations (n = 67), *Ichthybotus bicolor* in 10 South Island locations (n = 18) (Supplementary Table 2.1). Specimens were collected between 2013 and 2021 (Pohe) and 2021-22 (me). GPS co-ordinates were recorded at each location. The final number of locations was 114, with 164 sites (Supplementary Table 2.1). Nymphs were sampled by kick sampling in streams, disturbing the sediment under displaced rocks, and catching specimens in a net held downstream. Adults were sampled with a light trap at night (Pohe, 2017). *Ichthybotus* specimens were identified on a white tray using unique physical traits (jaws of nymphs and wings of adults) and were transferred to 99% ethanol for later DNA extraction and sequencing (Chapter 3). Correct identification, and identification by multiple means, are crucial to achieve the gold standard methodology for SDM's (Araujo et al. 2019). Genetic sequencing also helped to confirm identification and reduce the risk of

misidentification (Zurell et al. 2020). Thirty-two presence locations were included from iNaturalist records (2013-2022) (Supplementary Table 2.1).

Presence/absence tables were constructed for each species. Real absence data were stream sites Pohe surveyed and collected mayflies but did not find *Ichthybotus* specimens (*hudsoni* n = 97, *bicolor* n = 146) therefore I did not need to use pseudo-absences, thus eliminating unreasonable occurrence records (Araujo et al. 2019). Site coordinates were cut off after the 2nd decimal point to match the scale of the climate data, and all replicates were removed (Sillero & Barbosa, 2021). Data was imported into R as a table of corresponding site latitudes and longitudes with 1s and 0s representing the presence or absence of either species. To understand how the current distribution of these two *Ichthybotus* species relates to climate the two species were modelled separately. Understanding the potential of climate to be limiting their current distribution within NZ is essential for projecting ecological niche on to past climate models. Coordinates were changed into spatial points data frame class to allow the extraction of predictive data for each site. Stream sites with *Ichthybotus bicolor* were used as "absence" locations for *Ichthybotus hudsoni* and vice versa.

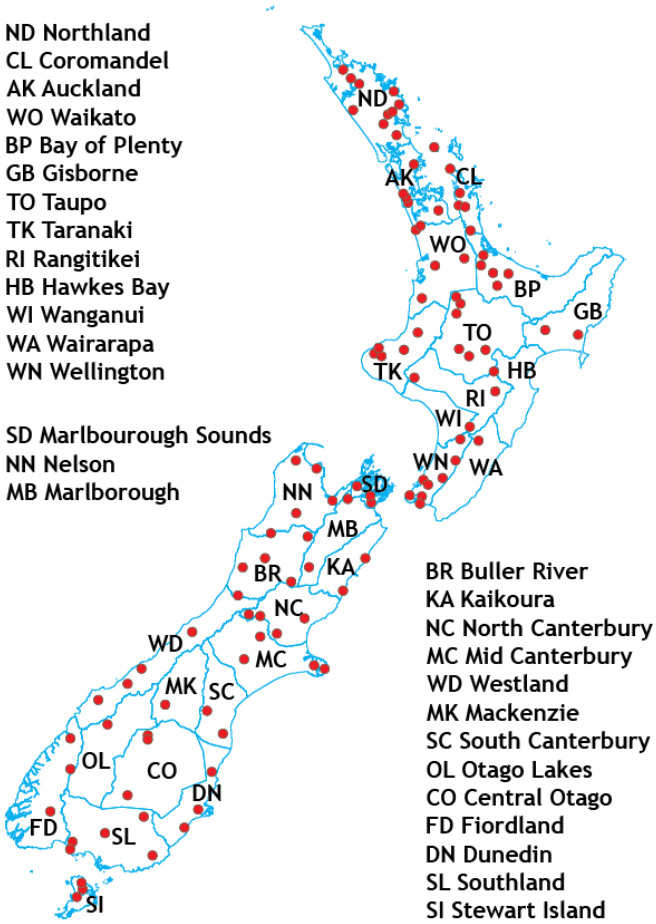


Figure 2.1. Sampling locations across 'Crosby' entomological regions of Aotearoa New Zealand (Crosby et al 1998) used for ecological niche modelling of two endemic species of burrowing mayfly (*Ichthybotus*)

This study has a nested purpose i.e., explaining the current distribution and determining how it relates to climate, mapping the potential distribution within NZ at the current time, and transferring/predicting past distributions. Accounting for the fact that transfer should not be attempted without thorough understanding of the model (Araujo et al. 2019; Zurell et al. 2020). (i.e., understanding of the past distribution necessitates understanding of how climate currently affects *Ichthybotus*).

2.2.2. Climate variable layers (Predictor Variables)

Model predictor variables consisted of 19 bioclimatic variables (Supplementary Table 2.2), obtained from the WorldClim website (https://www.worldclim.org/, accessed 18/03/21). These variables are derived from rainfall and temperature measurements to create variables applicable to biological modelling (Hijmans et al., 2005). I accessed 'current' data (data averaged over 1960-1990) and data modelled for the climate during the last glacial maximum (LGM) at a resolution of 2.5 arc minutes. The variables were all stacked, cropped to the extent of New Zealand, and formatted as a raster data set. I used the 'usdm' package variance inflation factor (VIF) analysis to reduce multicollinearity between the 'current' variables ('VIF' R package, VifCor function). This is a stepwise process which calculates the VIF and excludes highly collinear variables (in pairs of collinear variables one is removed). I excluded all variables with a VIF higher than 10 (Supplementary Table 2.2, Table 2.1). This increases parsimony and minimises over-fitting during the modelling process (i.e., increases simplicity and applicability beyond the data; Fletcher et al. 2016). Models have been shown to perform better with species-specific non-redundant predictor sets (Cengic et al. 2020). In this case the correlation threshold was set to 0.85, 1 being highly correlated. Bioclimatic variables are correlated due to trends in climate, so the elimination of all collinearity is not possible, and the goal is reduction only. After this process the 19 variables were reduced to a subset of eight bioclimatic for further analysis (Table 2.1). I considered variable importance to check for "biological plausibility" (Zurell et al. 2020).

Table 2.1. Variance inflation factor for the bioclimatic variables retained in ecological niche models of two species of New Zealand burrowing mayfly *Ichthybotus.*

Variable	Description	VIF
Bio 18	Precipitation of the warmest quarter	6.213425
Bio 19	Precipitation of the coldest quarter	5.647249
Bio 3	Isothermality (Bio2/Bio7)	1.558058
Bio 4	Temperature seasonality (standard deviation *100	2.086173
Bio 5	Max temperature of warmest month	4.464746
Bio 8	Mean temperature of wettest quarter	2.259656
Bio 9	Mean Temperature of Driest Quarter	8.200605
Bio 15	Precipitation seasonality (Coefficient of Variation)	1.340981

Model Assumptions: In applying ecological niche models using distribution data I am assuming there is an equilibrium between the species range and environmental variation such as the two mayflies fill their niche and do not occur elsewhere. I assume no observation bias issues and independence of species observations whereby each species record represents new information. Ecological niche models will only be biologically meaningful if important predictors of distribution are included in the model. I am assuming that key explanatory variables are included and climate variables are free of error.

2.2.3. Ecological niche model building

The R package 'biomod2' version 3.5.1 (Thuiller et al., 2009) was used to model the environmental niche of *I. bicolor* and *I. hudsoni* separately. Biomod uses an ensemble model, which generates consensus distributions based on individual model projections. This lessens the effects of different model discrepancies and has been shown to give more robust estimates than individual models (Araujo and New, 2006a; Beaumont et al., 2016; Cengic et al. 2020), i.e., better separation of signal and noise (Araujo and New 2006, Dormann et al. 2018, Hoa et al. 2018). Six modelling classes were used to analyse the presence/absence data against the nine remaining 'current' climate variables: Generalized Linear Model (GLM), Generalised Boosting Model/Boosted Regression Tree (GBM), Generalised Additive Model (GAM), Classification Tree Analysis (CTA), Artificial Neural Network (ANN), and Random Forest (RF). Descriptions of each model are explained in Diniz-Filho et al. (2009). Modelling parameters were kept at the default values. Eighty per cent of the input data were used to calibrate the models, with the remaining 20% used to test them. The 'Prevalence' parameter was set at 0.5, allowing absences and presences to be weighted equally and VarImport was set to 3, allowing three permutations to estimate variable importance. Calibration and evaluation of each statistical model was repeated five times. Model accuracy was evaluated using the True Skill Statistic (TSS) and Receiver Operating Characteristic (ROC), due to their suitability for species distribution modelling (Allouche et al. 2006)

2.3. Results

Location records were used as true presence and absence points for ecological niche modelling. The *Ichthybotus hudsoni* dataset has 64 presences (all in North Island) and 96 true absences. The *I. bicolor* dataset has 17 presences (all in South Island) and 143 true absences.

2.3.1. Model Evaluation

A high variance inflation factor (VIF) shows that a variable is highly collinear with other variables in the model. A VIF of 10 is highly correlated, hence all twelve variables with a VIF above 10 were excluded (Cheng et al 2022). Eight climate variables were used in this ensemble model (Table 2.1). There is collinearity between all variables which is to be expected when they are all bioclimatic variables.

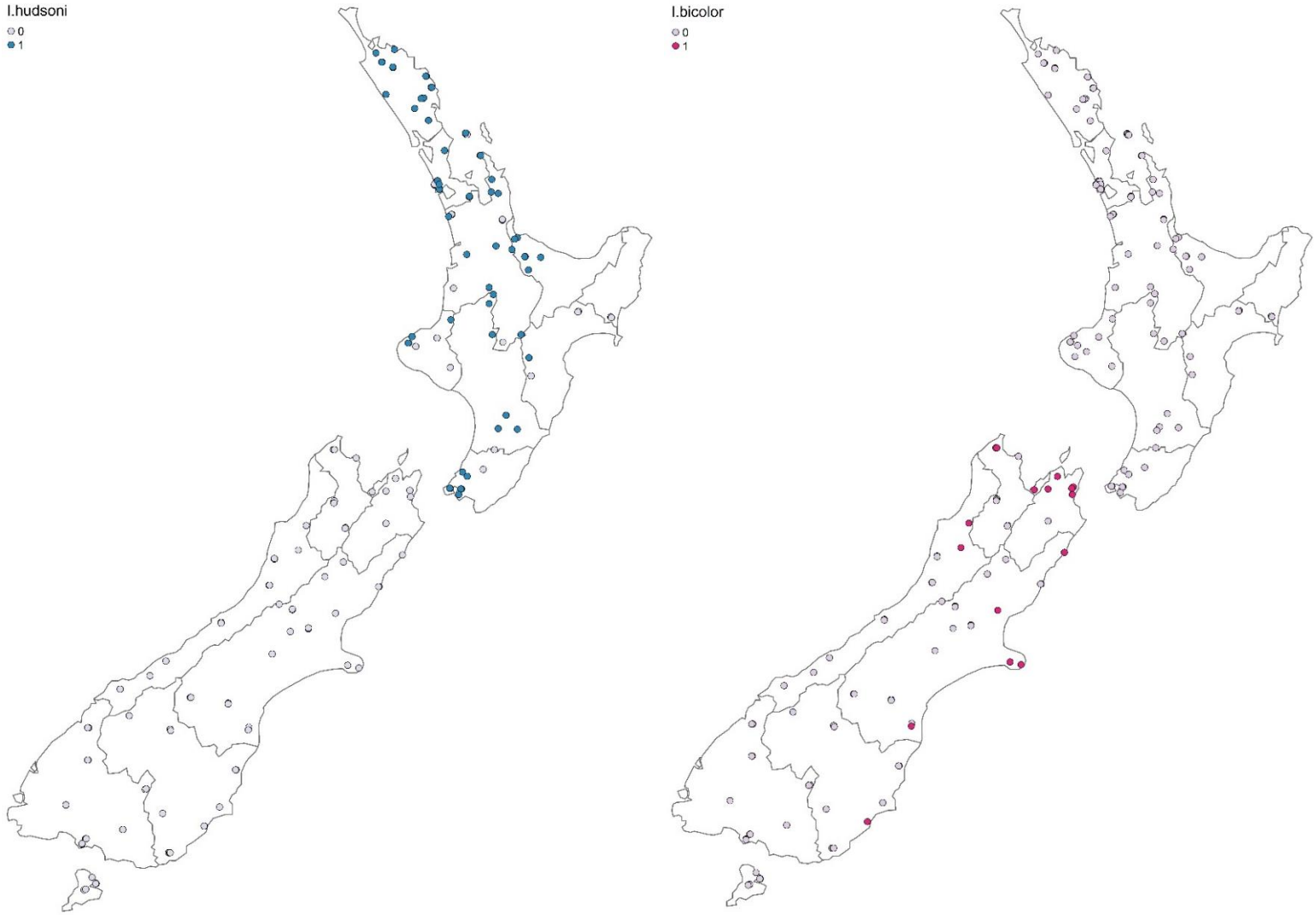


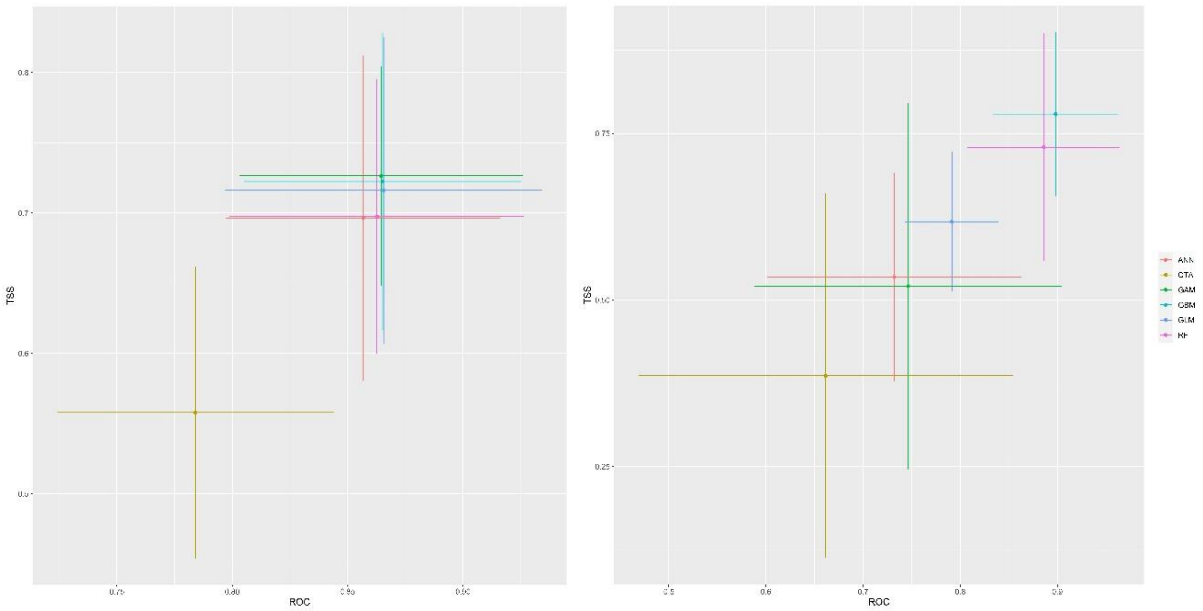
Figure 2.2. Presence/absence (left) *Ichthybotus hudsoni* and (right) *I. bicolor* at each of 114 locations surveyed for mayflies throughout New Zealand. Darker dots represent presence, lighter represent absence. Lines indicate recognised freshwater management regions.

Table 2.2. Ecological Niche model fit statistics for the New Zealand mayflies, *Ichthybotus hudsoni* and *I. bicolor*. Upper: True Skill Statistic (TSS) for each mayfly species and model (GLM, GBM, GAM, CTA, ANN, RF) included in ensemble model, across 5 runs with the mean of the 5 runs. Lower: Relative Operating Characteristic (ROC) for each mayfly species and model (GLM, GBM, GAM, CTA, ANN, RF) included in the ensemble model, across 5 runs with the mean of the 5 runs. Dark blue shows model runs that are excellent or with high accuracy (0.8–1), light blue is good accuracy (0.6–0.8), and unshaded is failed to model accurately (0–0.6) (Thuiller et al. 2009, ModOperating Manual for Biomod).

Model	*Ichthybotus bicolor*						*Ichthybotus hudsoni*					
	1	2	3	4	5	Mean	1	2	3	4	5	Mean
GLM	0.578	0.578	0.621	0.517	0.793	0.6174	0.895	0.684	0.632	0.737	0.632	0.716
GBM	0.931	0.655	0.862	0.655	0.793	0.7792	0.895	0.713	0.607	0.684	0.713	0.7224
GAM	0.655	0.362	0.862	0.147	0.578	0.5208	0.842	0.684	0.632	0.737	0.737	0.7264
CTA	0.276	0.612	0.681	0.362	0	0.3862	0.664	0.53	0.502	0.429	0.664	0.5578
ANN	0.69	0.655	0.517	0.293	0.517	0.5344	0.842	0.713	0.526	0.741	0.66	0.6964
RF	0.75	1	0.621	0.552	0.724	0.7294	0.818	0.713	0.555	0.741	0.66	0.6974
GLM	0.776	0.776	0.81	0.733	0.862	0.7914	0.96	0.866	0.777	0.895	0.83	0.8656
GBM	0.974	0.879	0.931	0.802	0.905	0.8982	0.935	0.879	0.769	0.879	0.864	0.8652
GAM	0.836	0.629	0.94	0.547	0.78	0.7464	0.939	0.868	0.773	0.895	0.848	0.8646
CTA	0.638	0.802	0.858	0.647	0.362	0.6614	0.783	0.816	0.755	0.704	0.862	0.784
ANN	0.802	0.828	0.75	0.504	0.776	0.732	0.911	0.885	0.787	0.903	0.798	0.8568
RF	0.922	1	0.862	0.793	0.853	0.886	0.919	0.891	0.757	0.895	0.852	0.8628

The true skill statistic and relative operating characteristics evaluation scores assess how well various models fit the data for *Ichthybotus*, and how accurately the models would predict a presence or absence correctly. The true skill statistic (TSS) which is sensitivity + specificity -1 (Allouche et al. 2006). It can rage from -1 to 1, with numbers closer to 1 means that the model is good at distinguishing presence and absence. It also indicates a good balance between sensitivity and specificity. Focusing on TSS is, recommended by Allouche et al. (2006). Shading shows predictive accuracy. Dark blue shows model runs that are excellent or with high accuracy (0.8-1), light blue is good accuracy (0.6-0.8), and unshaded is failed to model accurately (0-0.6) (Table 2.2) (Thuiller et al. 2009, ModOperating Manual for Biomod). Evaluation scores for individual models indicated good performance as indicated by relative operating characteristics (ROC) showing the relationship between sensitivity and specificity (Figure. 2.3). The common summary statistic for plot interpretation is AUC or the ROC area under the curve. Like TSS, the higher the ROC the better the model (Fawcett, 2006). Both species tend to score higher in ROC than TSS. *I. bicolor* models scored well in both TSS and ROC (most above 0.6 TS and 0.75 ROC), as shown by Figure 2.3. *I. hudsoni* also had high scores (above 0.65 for TSS and 0.85 for ROC) for all models except CTA. CTA scored the lowest for both metrics, however there was no outlying model scores. Aside from CTA, *I. hudsoni* scores were more clustered, therefore consistent across models.

Figure 2.3. Mean (dot) and standard deviations (lines) of TSS vs ROC scores of models included in the *Ichthybotus hudsoni* (left) and *I. bicolor* (right) ensemble models, Biomod2. Legend reads (top down) ANN, CTA, GAM, GBM, GLM, RF. Left: y axis (TSS) 0.5-0.8, x axis (ROC) 0.75-0.9. Right: y axis (TSS) 0.25-0.75, x axis (ROC) 0.5-0.9.



2.3.2. Predictor Variable Importance

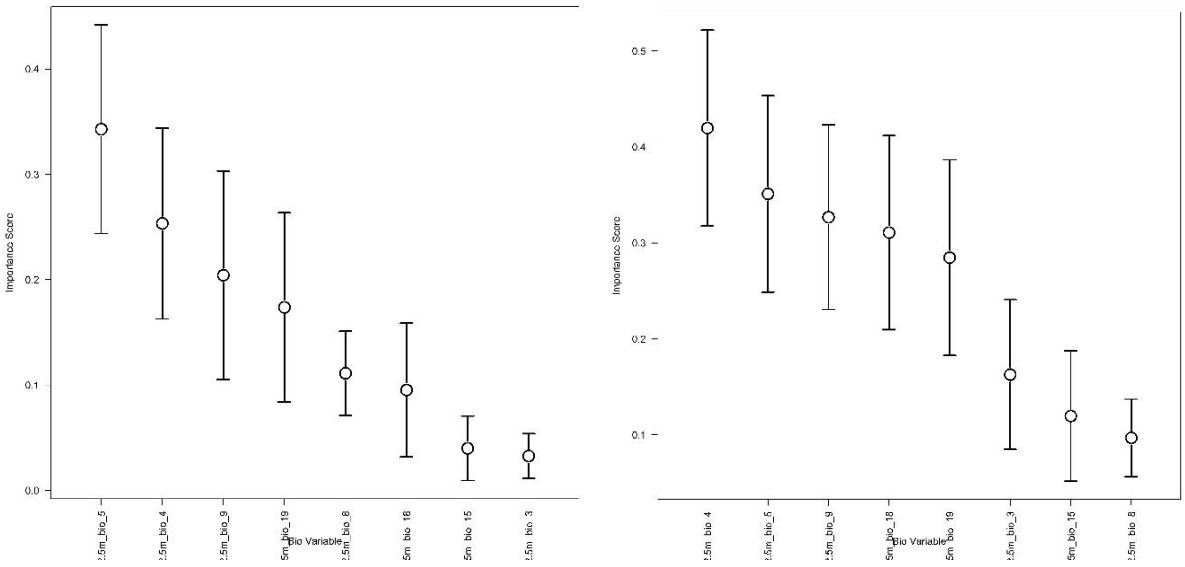


Figure 2.4. Ranked means (circles) with standard deviations of variable importance scores for *Ichthybotus hudsoni* (left) and *I. bicolor* (right) biomod2 ensemble models for the variables included in the model based on VIF scores. Left x axis labels read (left to right): bio_5, bio_4, bio_9, bio_19, bio_8, bio_18, bio_15, bio_3. Right x axis labels read (left to right): bio_4, bio_5, bio_9, bio_18, bio_19, bio_3, bio_15, bio_8.

There is a varied effect of a given predictor variable on the likelihood of presence (Figure 2.4). All *I. hudsoni* variables scored between 0 and 0.4 i.e., they were of low to moderate importance individually. There is no significant difference between the importance scores, though there are three stepped groups. The three most important appear to be bio5, bio9 and bio4, with bio3 and bio15 being the least important with low standard deviations. As the importance score increased, the standard deviation also seems to increase, meaning the average importance of these variables is less certain. The scores of *I. bicolor* were all between around 0.1 to 0.5, also all low to moderate importance individually. Both species had a similar order of importance. Bio4, bio5 and bio9 were the three most important (switched order), followed by bio18 and bio19 (switched), then bio8, bio15 and bio3 (switched) being the least important to both. All variables scored more important to distribution for *I. bicolor*.

The predictive ability of the final ensemble models was assessed using weighted means (EMmw) and comparing binary model predictions to the data's original species occurrences (Table 2.3). Both models showed high predictability, with both TSS above 0.84 and ROC above 0.97. Sensitivity, which is the percentage of presences correctly predicted, and specificity, the percentage of absences correctly predicted, were also high, above 91% and 93%. The cut off score is the value that maximises the sensitivity and specificity of the model. Areas on a map with values above the cut-off are classed presences in an optimised model.

Table 2.3. Ensemble model evaluation statistics: the testing score, the threshold (for transforming continuous likelihoods into binary data with an optimised score) i.e., cut-off score, sensitivity (% of presences correctly predicted) and specificity (% absences correctly predicted).

Species	Evaluation Method	Testing	Cut-off	Sensitivity	Specificity
I. bicolor	TSS	0.986	490.5	100	98.63
	ROC	0.998	490.5	100	98.63
I. hudsoni	TSS	0.849	515.5	91.045	93.814
	ROC	0.971	516.5	91.045	93.814

2.3.2. Current Ecological Models

The current predicted distribution (potential niche space) of *I. hudsoni* covers most of North Island and some regions in northern South Island (Figure 2.5). The highest probability of suitable habitat is in northland and coastal northern North Island, including islands around Auckland. Climate is less likely to be suitable inland likely due to higher elevation near Taupo, and expanding north and south along the Kaimanawa, Ruahine and Tararua ranges. There is some suitable climate in the northern tip of the South Island, however they have not been observed there in reality. Based on climatic variables, the potential niche space for *I. bicolor* is predicted to stretch most of the length of the east coast of South Island, and include some suitable habitat scattered across the North Island, predominantly in the southern North Island (Figure 2.5). There does not appear the be suitable habitat inland at higher elevations in the southern alps, or along the west coast of the South Island. Potential suitable habitat of *I. bicolor* inferred in northern South Island and in the Manawatu overlaps the predicted and realised distribution of *I. hudsoni*. Both species are predicted to have potential niche space on the coasts or lowland areas, with mountainous/steep terrain further inland unsuitable.

2.3.3. Past Ecological Model

Projecting ecological niche models, based on current distributions of these two mayfly species, onto models of past climate, suggest that areas of suitable climate for both species would have been widespread in New Zealand during the last LGM. *Ichthybotus hudsoni* may have had suitable habitat that was less fragmented and more widespread during the LGM than today. Given the cut-off of 515.5/516.5 for *I. hudsoni* the model suggests that potential niche existed country wide, particularly around the northwest coastal regions, and the southern tip of New Zealand. Suitable conditions for *I. bicolor* may not have existed during the LGM according to the ensemble model with maximised sensitivity and specificity (with a cut-off of 490.5). If *I. bicolor* was present during the LGM the most suitable climate would be along the east coast of the South Island and in the Whanganui basin. However, this is not predicted to be likely according to the model.

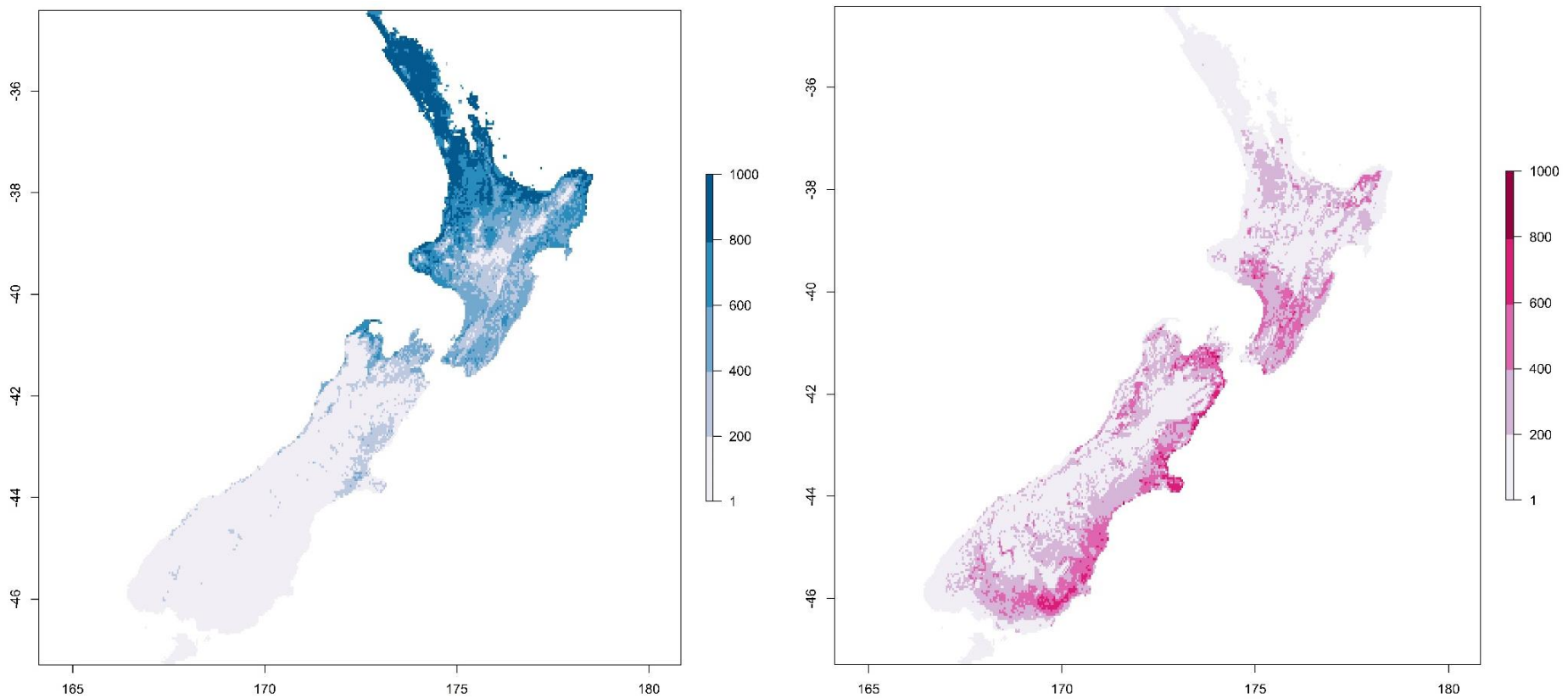


Figure 2.5. New Zealand, showing the predicted current distribution of *Ichthybotus hudsoni* (left) and *I. bicolor* (right), using biomod2 ensemble models. The axes show latitude (horizontal) and longitude (vertical). The colour gradient shows the model score as a representation of the probability of presence, ranging from lower (light) to higher (dark) likelihood. The cut-off score for I. hudsoni is 515.5/516.5, and the cut-off score for I. bicolor is 490.5. Both scores sit near the centre of the scale, therefore mid to dark tones indicate predicted presences.

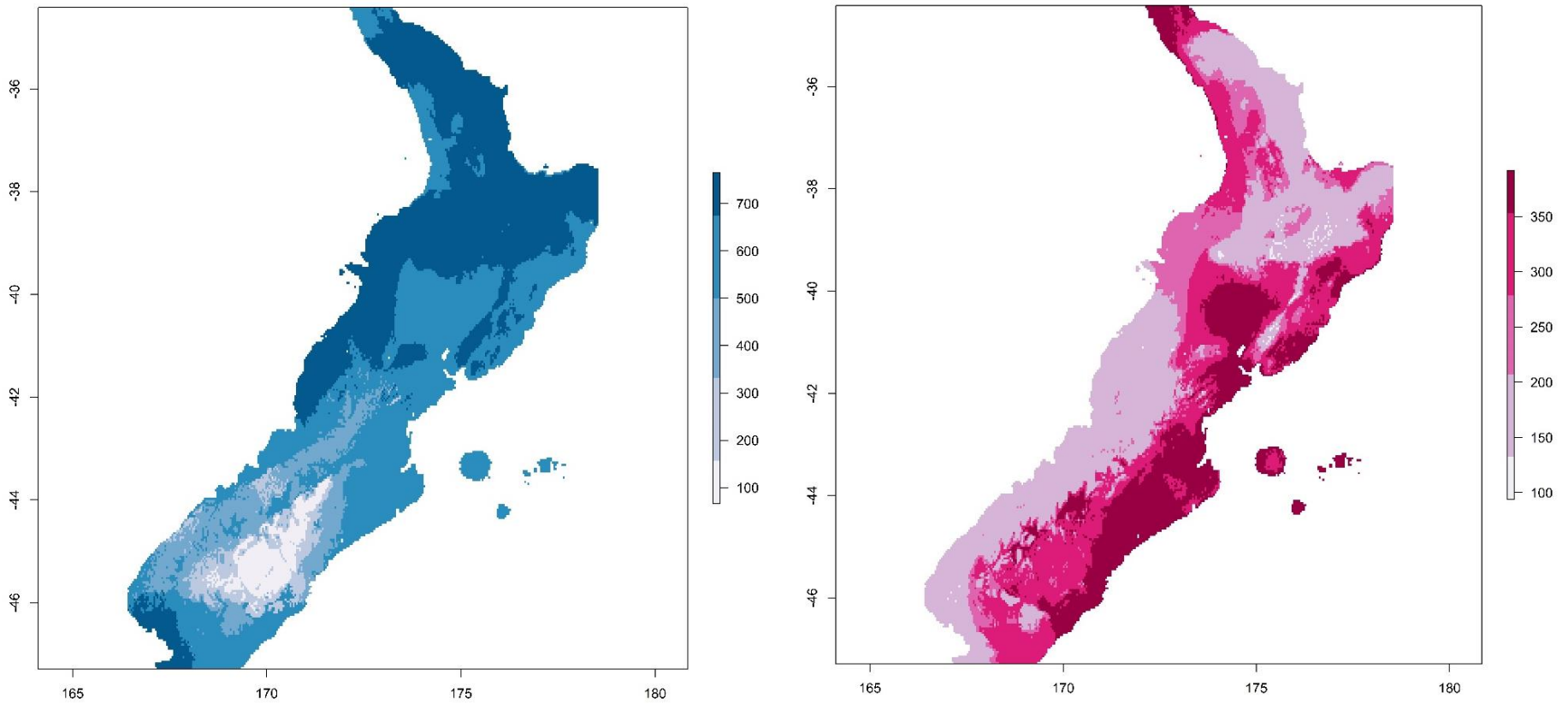


Figure 2.6. New Zealand, showing the predicted LGM distribution of *Ichthybotus hudsoni* (left) and *I. bicolor* (right), using biomod2 ensemble models. The axes show latitude (horizontal) and longitude (vertical). The colour gradient shows the model score as a representation of the probability of presence, ranging from lower (light) to higher (dark) likelihood. The cut-off score for *I. hudsoni* is 515.5/516.5, which sits close to the centre of the scale. Therefore, the two darker blue tones indicate predicted presences. The cut-off score for *I. bicolor* is 490.5, which is beyond the scale therefore all shades indicate predicted absences.

2.4. Discussion

2.4.1. Ecological Niche Modelling Predictive Accuracy

Individual models: Distribution data for both mayfly species included true absences which adds to the predictive accuracy of the model compared to the use of pseudo absences (Chefoaui & Lobo 2008). However, it is recommended that >30 presence locations (sample size; occurrences) be used, as model accuracy decreases with too few (Breiner et al. 2015, Hernandez et al. 2006, Stockwell & Peterson 2002, Wisz et al. 2008). The smaller dataset for *I. bicolor* probably explains the lower individual model evaluation scores compared to *I. hudsoni*. With only 17 locations where *I. bicolor* was found, each presence point is likely to have a significant effect on the model of this species' ecological niche model. These models, dependent on the quality and quantity of data, may be improved with additional sampling of *I. bicolor* specimens. Ensemble models scored better than individual models, as expected (Araujo & New 2006a, Breiner et al. 2015, Marmion et al. 2009). Surprisingly, the *I. bicolor* ensemble model had higher TSS and ROC scores than *I. hudsoni*, in contrast to individual models. This has occurred in previous studies; smaller range size or less prevalent species data yields higher scoring and more robust models (Allouche et al., 2006, Segurado & Araujo, 2004, Bouska et al. 2015, Koot et al. 2022). Lower scores for widespread species may reflect high environmental variance among locations (Koot et al. 2022). In other studies, sample size had no, or inconsistent impacts, on model receiver operating characteristic (ROC), often called AUC (area under the curve) scores (a statistic used widely to estimate predictive accuracy of presence-absence SDM's) (McPherson et al., 2004; Elith et al. 2006, Franklin, 2010, Bouska et al. 2015). However, AUC scores are independent of the prevalence of presence/absence points. Rare and specialist species can have a more predicable distribution leading to more robust models (Elith et al. 2006, Bouska et al. 2015). With *I. bicolor* as it is uncertain if the smaller sample size is due to specialised habitat or if it was found at fewer sites because of low abundance. Fewer records from iNaturalist could be due to both fewer humans in South Island and fewer streams with suitable substrate (Elith et al. 2006).

Sensitivity and specificity: Models for both species were able to distinguish between presence and absence points. Both had better sensitivity than specificity, i.e., were better at predicting presences correctly as opposed to absences (sensitivity being the chance true presences are included, specificity the chance that true absences are excluded from the predicted distribution). Both metrics were higher for *I. bicolor* perhaps due to that fact that there were fewer presence points to include. Models for *Ichthybotus hudsoni* had lower sensitivity reflecting previous studies that demonstrate the more common species generate models with more false positives (Syphard & Franklin, 2010). However, the study by Syphard & Franklin (2010) also resulted in uncommon species with more false negatives, which was not the case for this study of *Ichthybotus*. A small dataset can yield useful and interesting

results (Pearson et al., 2006) by predicting the environmental conditions that the species can occur, i.e., this would be useful for targeting areas for future sampling (Pearson et al. 2006).

2.4.2. Current Climate Projections

The current distribution of the two species of *Ichthybotus* mayfly matched the ecological niche space predicted by the ENMs relatively well. Both species appear to have available habitat in southern North Island and northern South Island, however we know from collection data they do not occur there i.e., each species is restricted to one island with no cross over. There seem to be gaps around areas of higher elevation, likely due to harsher climates. The mismatch between the current distribution of these mayfly species and their predicted niche under current conditions indicates that variables not included in these models are also limiting their distribution. This could involve abiotic variables not included in the analysis (e.g., stream speed and substrate), or biological interactions such as competition, predation, disease and/or parasitism. For instance, there are lots of areas included in the predicted niche where we can be confident, they are not found such as alpine areas. This is due to the difference in environmental space and geographic space. The models are based on environmental predictors, meaning they build a picture of the environmental space that a species can inhabit. The geographical space will include the 2d map coordinates and elevation data, which will also impact distribution but are not included; "ignorant of geographic proximity even when predictions are mapped into geographic space" (Elith & Leathwick, 2009). *Ichthybotus hudsoni* may be more suited to the climate in the North Island which excludes it from much of the South Island save for some small areas around the northern South Island. This is where competition between the two species may occur if these were realised niches. The niche predictions also assume that the species can disperse there, but with the presence of Cook Strait means this is likely not the case, i.e., biotic error (Pearson et al., 2006).

2.4.3. Last Glacial Maximum Projections

During the LGM New Zealand was a single land mass rather than two islands. Ecological niche modelling projected on to past climate unexpectedly indicates that the climatic envelope for *I. hudsoni* was more extensive and contiguous during the LGM. *I. hudsoni* was predicted to have a country wide distribution, unlike its observed current distribution. The climatic envelope has become more restricted to mostly the North Island. North/South distributions are not uncommon amongst birds such as whio and weka (Trewick et al. 2017) despite the land connection only 20,000 years ago. The predicted past distribution for *I. bicolor* appears to be the opposite, with no habitat present during the LGM. The most likely areas *I. bicolor* could have inhabited were along the east coast of the South Island and Whanganui Basin, which does roughly line up with the current distribution. The LGM suggested distributions imply that no shared niche space existed due to the lack of *I.bicolor* habitat. It could be suggested that the *I. hudsoni* distribution may represent the country wide distribution of a single species before the two lineages/species diverged. It is interesting to note that glaciation would have affected distribution

(Carmelet-Rescan 2021), which is not explicitly account for and does not appear to be completely excluded from potential niche in the LGM (though the Otago area is predicted to be the least suitable niche for I hudsoni in the LGM).

2.4.4. Climate variables as predictors and 'biotic errors' (Pearson et al., 2006)

It is important to acknowledge that modelling provides us with only a potential niche. Results may differ with sample size, and models are inherently inaccurate but useful nonetheless (Jimenez-Valvaerde et al. 2008). The distribution of the two *Ichthybotus* species studied here appeared to be affected by different climate variables, suggesting that they may have different ideal climates (different sensitivities). This could partially explain the distribution of one species confined to South Island and one species confined to North Island New Zealand. Alternatively, it is possible that their difference distributions have resulted in models that infer distinct environmental envelopes for each species. Their distinctive distributions could be due to processes that are not driven by selection from the climate, such as allopatric speciation, and/or competitive exclusion. For *I. hudsoni* the one variable that was significantly more important than the others was maximum summer temperature suggesting that this mayfly might be limited by extreme temperatures. As seen for alpine grasshopper species in New Zealand (Koot et al. 2022), *I. bicolor* had multiple variables that were of similar importance.

Species distribution models work on the assumption that the system is in equilibrium, which is often not the case, particularly in freshwater systems which follow a dynamic equilibrium model with biodiversity constantly changing with waves of recolonisation (Death 2002, Bouska et al. 2015). In freshwater species disturbances, such as floods and drought, are important determinants of community structure (Death 2002, Death 2003, Dodds et al. 2004, Resh et al. 1988, Lepori & Hjerdt, 2006). Freshwater ecosystems are often recovering from disturbances which are not considered in ecological niche models. (Bouska et al. 2015). Species distribution models typically require the assumption that realized distribution encompasses most of the potential habitat, but more complex models can be built to capture occurrence dynamics, including spatial dependence, therefore relaxing the species–environment equilibrium assumption (Zurell et al. 2020). Other factors affecting freshwater biodiversity include habitat structure (i.e., substrate), water quality and land use (Ormerod et al. 2010, Daufresne et al. 2007). Biotic interactions (i.e., food resources, predation, competition, disease, and parasitism), dispersal, phenotypic plasticity and evolution are assumed to have no effect on distribution, however we know they are limiting to distribution (Heikkinen et al. 2006, Lavergne et al. 2010). For example, predicting the distribution of two New Zealand wētā species relies heavily on understanding how they compete and displace each other (Bulgarella et al. 2013). Inclusion of biotic factors would improve model fit and accuracy, however they are difficult to account for (Pearson et al. 2006, Moilanen et al. 2008, Morin & Thuiller 2009, Boulangeat et al. 2012). Climate variables may be too coarse to account for floods, droughts or other stochastic events that will drastically affect community structure

(Daufresne et al. 2007). Can be tricky to account for interactions between climate variables and how their effects could change (Guisan et al. 2006). Climate variables such as precipitation and air temperature are important to freshwater invertebrates, and act as useful proxies for stream conditions, i.e., stream and air temperature are highly correlated (Durance & Ormerod, 2009). Freshwater systems are also affected by upstream conditions, slope, and hydrological conditions, which Jahnig et al. (2012) was able to successfully account for in local modelling, however, this was not achieved for large scale modelling. Climate variables are a great way of predicting distribution over a large scale, while habitat characteristics are more useful for a smaller/local extent study (Araujo et al. 2019). A model that could include both would be beneficial (Araujo et al. 2019). Difficult to include in a country wide model, especially when past conditions are unknown.

Predictions are more dubious outside of the temporal extent of the data (Thuiller 2003, Crimmins et al. 2013, Bouska et al. 2015, Elith et al. 2006). There may be novel conditions or environmental data from the LGM that are difficult to predict or unavailable. It would have been very useful to include factors like substrate, slope, and land use however this is not available for New Zealand during the LGM. Specially across large scales (Bouska et al. 2015). The relationship between climate variables and distribution of species may also change overtime. Bouska et al. (2015) explained this using slope and fish distribution. While slope may be a suitable proxy for discharge in the current climate, the slope may be the same in future climates however discharge patterns may change with other factors or anthropogenic impacts. Despite these omissions the models still score well, which suggests that these factors may not be necessary for a useful model. Depending on the research question, models that only include abiotic factors can still be useful, especially given the difficulties in obtaining biotic variable data at large scales.

2.2.4. Potential Biases and Errors (Guillera-Arroita 2017; Zurell et al. 2020)

One potential bias in the data is that Pohe's location criteria included being within a certain distance of native forest. This may be useful for this model, reducing the effect of stream pollution, disturbance, and other anthropogenic effects on mayfly presence. Data was collected initially for the purpose of looking at all NZ mayflies, so there is unlikely to be a bias toward Ichthybotus specific locations. Finally, it should be acknowledged that the data (presence/absence locations) was collected over 8 years, or were sourced from museum collections, so there is a potential for change in presence absence over time. Predictors and response variables are also from different time periods.

2.4.5. Implications for Population Genetics

If the projections of ecological models are realistic estimates of past distributions for these two *Ichthybotus* species, then we can predict that *Ichthybotus hudsoni* would have retained high genetic diversity throughout its range with potential for greater diversity in the northwest of North Island. There is often a link between population and distribution size, and genetic diversity (Frankham, 1996), with a

large distribution implying a larger population. *Ichthybotus bicolour* had no projected niche space during the LGM therefore we can predict lower overall genetic diversity in this species.

Supplementary Table 2.1. Shows all site locations (164) and their co-ordinates, sorted form northern to southernmost. Includes presence (1) and absence (0) of species and location and the source of the data (Pohe = (Pohe, 2018), thesis = gathered for the purpose of this thesis).

Site	Latitude NZGD2000	Longitude NZGD2000	*I. hudsoni*	*I. bicolor*	Source
Whangaroa	-35.01	173.71	1	0	Pohe
Kaitaia	-35.06	173.36	1	0	iNaturalist
Mangamuka	-35.19	173.48	1	0	Pohe
Mangamuka	-35.19	173.47	1	0	Pohe
Puketi	-35.26	173.68	1	0	Pohe
Puketi	-35.27	173.68	1	0	Pohe
Russell	-35.39	174.31	0	0	Pohe
Russell	-35.39	174.3	1	0	Pohe
Whananaki	-35.55	174.4	1	0	Pohe
Whananaki	-35.55	174.41	1	0	Pohe
Waipoua	-35.65	173.55	1	0	Pohe
Mangere	-35.7	174.24	1	0	Pohe
Pukenui	-35.7	174.26	1	0	Pohe
Mangere	-35.71	174.24	1	0	Pohe
Maungatapere	-35.71	174.21	1	0	iNaturalist
Tangihua	-35.85	174.09	1	0	iNaturalist
Mareretu	-36.02	174.35	1	0	Pohe
Little Barrier	-36.2	175.05	1	0	Pohe
Little Barrier	-36.21	175.05	0	0	Pohe
Little Barrier	-36.22	175.06	1	0	Pohe
Little Barrier	-36.22	175.08	0	0	Pohe
Pohuehue	-36.45	174.65	1	0	Pohe
Fantail Bay	-36.52	175.32	1	0	Pohe
Fantail Bay	-36.52	175.33	1	0	Pohe
Mahakirau	-36.86	175.54	1	0	iNaturalist
Cascades	-36.88	174.51	1	0	Pohe
Cascades	-36.88	174.52	1	0	Pohe
Waitakere	-36.88	174.52	1	0	iNaturalist
Cascades	-36.89	174.53	0	0	Pohe
Upper Nihotupu	-36.93	174.55	1	0	Pohe
Piha	-36.93	174.46	0	0	Pohe
Waiatarua	-36.93	174.55	1	0	iNaturalist
Piha	-36.94	174.46	0	0	Pohe
Huia	-37	174.56	1	0	Pohe
Huia	-37	174.55	1	0	Pohe
Te Puru	-37.04	175.53	1	0	Pohe
Coromandel Forest Park	-37.06	175.66	1	0	iNaturalist
Hunua	-37.1	175.12	1	0	Pohe
Hunua	-37.11	175.12	1	0	Pohe
Port Waikato	-37.36	174.79	1	0	Pohe
Port Waikato	-37.36	174.78	0	0	Pohe
Port Waikato	-37.39	174.73	1	0	iNaturalist
Waitawheta	-37.43	175.74	0	0	Pohe
Waitawheta	-37.44	175.74	1	0	Pohe

Whakamarama	-37.69	176.02	1	0	iNaturalist
Whakamarama	-37.71	175.97	1	0	iNaturalist
Te Tapui	-37.81	175.62	1	0	Pohe
Kaimai Saddle	-37.86	175.92	1	0	iNaturalist
Pirongia	-37.93	175.07	1	0	Pohe
Rotorua	-37.96	176.17	0	0	Pohe
Rotorua	-37.96	176.16	1	0	Pohe
Okere Falls	-37.97	176.46	1	0	iNaturalist
Hillcrest Rotorua	-38.15	176.23	1	0	iNaturalist
Tokoroa	-38.4	175.49	1	0	Pohe
Whareorino	-38.41	174.82	0	0	Pohe
Pureora	-38.5	175.57	1	0	iNaturalist
Pureora	-38.63	175.49	1	0	Pohe
Waikaremoana	-38.74	177.17	0	0	Pohe
Waikaremoana	-38.75	177.16	0	0	Pohe
Gisborne	-38.82	177.78	0	0	Pohe
Gisborne	-38.83	177.79	0	0	Pohe
Kotare	-38.86	174.77	1	0	Pohe
Tongariro	-39.07	175.55	1	0	iNaturalist
Kaimanawa	-39.07	176.09	1	0	iNaturalist
Hurdon	-39.1	174.04	1	0	iNaturalist
Pouiatoa	-39.12	174.51	0	0	Pohe
Turangi	-39.18	175.75	0	0	Pohe
Pukeiti	-39.19	173.98	1	0	iNaturalist
Pukeiti	-39.19	173.97	1	0	iNaturalist
Egmont	-39.24	174.11	0	0	Pohe
Stratford	-39.33	174.28	1	0	iNaturalist
Otakeho	-39.4	174.06	1	0	Thesis
Timahanga Station	-39.4	176.24	1	0	Pohe
Rotokohu	-39.54	174.76	0	0	Pohe
Sentry Box	-39.66	176.28	0	0	Pohe
Makohine	-40.22	175.81	1	0	Pohe
Turitea	-40.41	175.66	1	0	Thesis
Waewaepa	-40.42	176.02	1	0	Pohe
Kahutarawa	-40.46	175.61	1	0	Thesis
Collingwood	-40.71	172.56	0	1	Pohe
Collingwood	-40.71	172.58	0	1	Pohe
Eketahuna	-40.71	175.59	0	0	Pohe
Totaranui	-40.83	172.99	0	0	Pohe
Carterton	-40.99	175.38	0	0	Pohe
Katarawa	-41.03	174.99	1	0	iNaturalist
Upper Hutt	-41.09	175.08	1	0	iNaturalist
Harvey Bay	-41.12	173.73	0	1	Pohe
Otari-Wilton	-41.26	174.76	1	0	Thesis
Otari-Wilton	-41.26	174.75	1	0	iNaturalist
Wainuiomata	-41.27	174.97	1	0	iNaturalist
Wainuiomata	-41.27	174.96	1	0	Thesis
Waikawa	-41.27	174.03	0	1	iNaturalist

Picton	-41.29	174	0	1	iNaturalist
Richmond	-41.3	173.55	0	1	Pohe
Nelson	-41.31	173.29	0	1	iNaturalist
Remutaka	-41.35	174.92	1	0	iNaturalist
Robertson	-41.38	174.01	0	1	Pohe
Motueka	-41.44	172.58	0	0	Pohe
Motueka	-41.45	172.57	0	0	Pohe
Motueka	-41.47	172.57	0	0	Pohe
Avon Valley	-41.76	173.55	0	0	Pohe
Lyell	-41.79	172.05	0	1	Pohe
Rotoiti	-41.82	172.79	0	0	Pohe
Rotoiti	-41.83	172.78	0	0	Pohe
Reefton	-42.14	171.9	0	1	Pohe
Mororimu	-42.21	173.86	0	1	Pohe
Blackball	-42.26	171.45	0	0	Pohe
Blackball	-42.27	171.45	0	0	Pohe
Molesworth	-42.31	172.75	0	0	Pohe
Lewis Pass	-42.52	172.4	0	0	Pohe
Brunner	-42.64	171.35	0	0	Pohe
Brunner	-42.64	171.36	0	0	Pohe
Hawkswood	-42.66	173.42	0	1	Pohe
Hawkswood	-42.66	173.41	0	0	Pohe
Arthurs Pass	-42.91	171.54	0	0	Pohe
Mt White Station	-42.98	171.79	0	0	Pohe
Mt White Station	-42.99	171.79	0	0	Pohe
Waipara	-43.04	172.6	0	1	iNaturalist
Harihari	-43.17	170.45	0	0	Pohe
Harihari	-43.18	170.45	0	0	Pohe
Oxford	-43.25	172.09	0	0	Pohe
Oxford	-43.26	172.09	0	0	Pohe
Porters Pass	-43.3	171.75	0	0	Pohe
Mt Somers	-43.62	171.41	0	0	Pohe
Paringa	-43.72	169.41	0	0	Pohe
Okuti	-43.78	172.83	0	1	Pohe
Hinewai	-43.82	173.04	0	1	Pohe
Haast	-43.93	169.11	0	0	Pohe
Jackson Bay	-44.12	168.55	0	0	Pohe
Twizel	-44.24	169.87	0	0	Pohe
Twizel	-44.24	169.88	0	0	Pohe
Tekapo	-44.32	170.58	0	0	Pohe
Tekapo	-44.33	170.58	0	0	Pohe
Rob Roy	-44.5	168.72	0	0	Pohe
Waimate	-44.66	170.97	0	0	Pohe
Milford Sound	-44.67	167.96	0	0	Pohe
Milford Sound	-44.67	167.94	0	0	Pohe
Lindis Pass	-44.69	169.49	0	0	Pohe
Waimate	-44.7	170.96	0	1	iNaturalist
Lindis Pass	-44.71	169.5	0	0	Pohe

Te Anau	-45.13	167.95	0	1	Pohe
Te Anau	-45.13	167.94	0	0	Pohe
Oamaru	-45.27	170.73	0	0	Pohe
Oamaru	-45.27	170.72	0	0	Pohe
Waikaia	-45.54	169.04	0	0	Pohe
Waikaia	-45.55	169.02	0	0	Pohe
Borland	-45.77	167.53	0	0	Pohe
Dunedin	-45.8	170.42	0	0	Pohe
Black Gully	-45.89	169.35	0	0	Pohe
Akatore	-46.07	170.13	0	1	iNaturalist
Gore	-46.12	168.6	0	0	Pohe
Otautau	-46.25	167.9	0	0	Pohe
Otautau	-46.25	167.91	0	0	Pohe
Invercargill	-46.32	167.83	0	0	Pohe
Invercargill	-46.33	167.83	0	0	Pohe
Catlins	-46.45	169.47	0	0	Pohe
Catlins	-46.45	169.46	0	0	Pohe
Catlins	-46.45	169.48	0	0	Pohe
Catlins	-46.45	169.49	0	0	Pohe
Bungaree	-46.8	168.02	0	0	Pohe
Oban	-46.89	168.1	0	0	Pohe
Oban	-46.89	168.09	0	0	Pohe
Rakeahua	-46.97	167.9	0	0	Pohe
Rakeahua	-46.98	167.87	0	0	Pohe

Supplementary Table 2.2. Codes for the 19 bioclimatic variables from WorldClim.

Variable	Description
BIO1	Annual Mean Temperature
BIO2	Mean Diurnal Range (Mean of monthly (max temp – min temp))
BIO3	Isothermality (BIO2/BIO7) (×100)
BIO4	Temperature Seasonality (standard deviation ×100)
BIO5	Max Temperature of Warmest Month
BIO6	Min Temperature of Coldest Month
BIO7	Temperature Annual Range (BIO5-BIO6)
BIO8	Mean Temperature of Wettest Quarter
BIO9	Mean Temperature of Driest Quarter
BIO10	Mean Temperature of Warmest Quarter
BIO11	Mean Temperature of Coldest Quarter
BIO12	Annual Precipitation
BIO13	Precipitation of Wettest Month
BIO14	Precipitation of Driest Month
BIO15	Precipitation Seasonality (Coefficient of Variation)
BIO16	Precipitation of Wettest Quarter
BIO17	Precipitation of Driest Quarter
BIO18	Precipitation of Warmest Quarter
BIO19	Precipitation of Coldest Quarter

Chapter 3: Phylogeography of Ichthybotus

3.1 Introduction

3.1.1. Phylogeography Concepts

The concept of phylogeography was coined in 1987 and has grown it its first 30 years (Avise et al 2001, Avise, 2009). It involves using genetic, spatial, and temporal data, and coalescent theory to infer past and present relationships between populations (Trewick et al 2011, Beheragaray 2008, Trewick, et al 2022). Phylogeography often utilises mitochondrial DNA (mtDNA) due to being compact and relatively quick to accumulate nucleotide substitutions i.e., evolution is rapid and divergence in isolated populations will accumulate relatively quickly (Avise, 2009, Morgan-Richards et al, 2017). This leads to mtDNA having high sequence variation within species, which is crucial for phylogeographic studies at species level (Avise, 2009). Mitochondrial DNA is also passed down maternally and does not undergo recombination as part of sexual reproduction. This means the only changes that will occur from generation to generation in these molecules is through mutations. Different mtDNA sequences found in a population are referred to as haplotypes. Due to the rapid evolution of these molecules, there are often a relatively large number of haplotypes present in a population or species allowing for lineages and patterns to be inferred within and between these groups. Another important concept in phylogeography is the theory of coalescence. This theory states that although there are distinct genetic lineages in populations, they all coalesce to a common ancestor. These theories combined allow the examination of population diversity across a landscape, in this case the whole of New Zealand, and into the past at what may have caused the genetic structure within a species (Trewick et al. 2022).

3.1.2 Potential patterns inferred and phylogeography in New Zealand

The goal of phylogeography is to infer patterns in space and time, with relation to genetic evolution. There have been various studies carried out on New Zealand invertebrates. These have used current distributions and phylogeography to reconstruct past refugia (Buckley et al., 2009; Buckley et al., 2010; Marske et al., 2011), explore taxonomic relationships and distinguish between species (Atkinson, 2006, Chinn & Gemmell, 2004). There are various distribution patterns that can be inferred from this field of study. Range expansion from a single refugia leads to high to low genetic diversity moving into peripheral or more recently colonised localities (Excoffier et al. 2009). In New Zealand many species are inferred to have had refuge in northern North Island during the LGM, before expanding south. This was the case in the mayfly *Acanthophlebia* which is now widespread across the North Island (Trewick et al. 2022), and in New Zealand stick insects *Clitarchus hookeri* and *Argosarachus horridus*, which both found refuge in the northern North Island and eastern South Island during the LGM (Buckley et al 2009; Buckley et al., 2010). Moving out of a refugia often has a bottleneck effect on founded populations, this is common when populations move into previously glaciated areas, such as in the

South Island during the LGM (Excoffier et al 2009, Buckley et al 2010, Marshall et al 2009). When a population expands from multiple glacial refugia they each produce local regions with high genetic diversity and the potential for hybrid zones to form between localities. This scenario would not necessarily result in decreasing genetic diversity moving away from the refuge (Excoffier et al 2009). This was found to be the case for cicada in the genus *Kikihia*. Like the stick insects (Buckley et al 2009, Buckley et al 2010) this genus was likely restricted to the North Island and northern South Island during the LGM. It shows split lineages across the Cook Strait and during the Pleistocene in the northern South Island, these lineages diverged down either side of the Southern Alps i.e., this genus shows allopatric divergence around Cook Strait and the Southern Alps (Marshall et al., 2008; Marshall et al., 2009; Marshall et al., 2011). *Kikihia muta* hybridized relatively recently due to post glacial secondary contact in certain locations. Finally, if populations survive in situ then large stable populations result in widespread high genetic diversity. This is seen in both tree and ground wētā, *Hemideina crassidens* and *Hemiandrus pallitarsis,* which have maintained high genetic diversity due to maintaining one stable and large population increasing the rate of mutation in the population over a long period of time (Chappell et al, 2012, Morgan-Richards et al. 2017). New Zealand alpine grasshoppers have high species diversity owing to large habitats and population size during the Pleistocene and maintaining this diversity despite a smaller population (Carmelet-Rescan et al, 2021). This is despite small populations been considered to have limited genetic diversity (Frankham, 1996). These trends can all help piece together the history of a species, and given their New Zealand context, may be relevant to mayfly distribution.

3.1.3. Phylogeography in Mayflies

Mayflies generally have low dispersal ability with mountains and oceans being major barriers (Barber James 2008). They can drift downstream as nymphs or colonize habitats over short distances during their winged stage, moving around 59m maximum from the stream for males and 44m for females (Petersen et al. 2004, Sartori & Brittain 2015). Their dispersal distance is considered restricted due to their short lifespan and fragility as adults. Some studies have found evidence of transoceanic dispersal of some mayfly species between Madagascar and continental Africa (Monaghan et al, 2005, Gattolliat, 2004). So long distance dispersal might be possible for some species. However, Kaneko et als (2021) study concluded that mayflies on the Japanese archipelago island Kyushu were genetically distinct, and *Acanthophelbia* in New Zealand can only be found on the North Island (Trewick et al. 2022) suggesting that transoceanic dispersal does not often occur. Species with low dispersal ability typically have geographically localised haplotypes that are much more similar than those at a distance, as intermixing over distance is less common than in species with high dispersal ability (Barber James 2008). There are often deeper divisions between populations, tracking back to a separation event or refugia (Avise, 2009). Three distinct lineages of baetid mayflies were identified across three Mediterranean islands

with no admixture in shared locations (Bisconti et al., 2016) in an example of applying phylogeography to endemic island mayflies.

3.1.4. Chapter Aims:

1. Test to see whether the current species level treatment of *Icthybotus* comprises the expected two reciprocally monophyletic genetic clusters based on wing colouration, and spatial distribution (islands). The alternative been one taxon being nested with the other.
2. To explore the genetic diversity of both species across New Zealand, with particular focus on the intraspecific diversity of *I. hudsoni*.
3. To determine if phylogeographic data corroborates the niche modelling trends inferred in Chapter 2 which suggested wider distribution in the past than currently observed for *Ichthybotus hudsoni*, which would have retained high genetic diversity throughout its range. *Ichthybotus bicolour* had no suitable projected niche space during the LGM, but is currently more suited to the east coast, therefore I predict relatively low genetic diversity throughout its range with possible higher diversity along the east coast.

3.2 Phylogeography Methods

3.2.1. Collection of Mayfly Specimens

Ichthybotus specimens for genetic analyses were collected from the full range of their distribution (Pohe, 2018, Pohe, 2019). Putative *Ichthybotus hudsoni* was collected from 12 North Island locations, and *Ichthybotus bicolor* from 10 locations in South Island (Table 3.1; Figure 3.1). Specimens were obtained either by light trapping of flying subimagos and imagos (Pohe et al, 2017) or as nymphs. Nymphs were captured by kick sampling in small, gravel-bottomed streams, disturbing the sediment under displaced rocks, and catching specimens in a net downstream. *Ichthybotus* specimens were sorted from other taxa on site using physical traits and were immediately preserved in 99% ethanol.

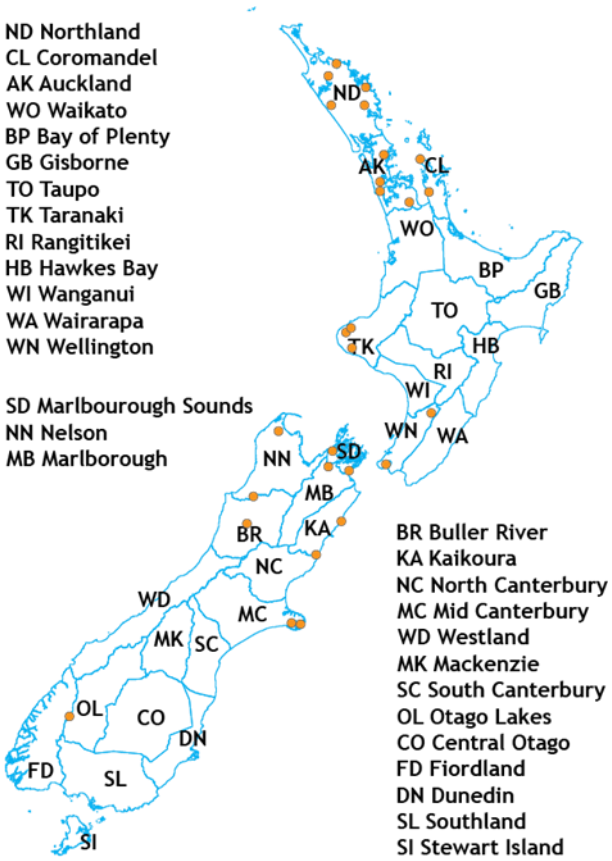


Figure 3.1. Sites where the New Zealand burrowing mayfly (*Ichthybotus hudsoni* and *I. bicolor*) specimens were collected for genetic analyses (Table 3.1).

3.2.2. DNA Extraction and Amplification

DNA was extracted from legs of nymph or adult specimens by crushing in 20uL of 10mg/mL proteinase K and 300uL digestion buffer was added (reheated and resuspended TNES: 2.5ml Tris (1M), 4.0ml NaCl (5M), 2.0ml EDTA (500mM), and 5ml SDS (sodium dodecyl sulphate), made up to 50ml with H2O), and incubated at 55ºC until fulling lysed. 85uL 5M NaCl solution was added, shaken well for at least 15 seconds before spinning at 14000rpm for 5 mins. 300uL of the resulting supernatant was pipetted into a new tube, with an equal amount of ice cold absolute EtOH and inverted to mix, to precipitate the DNA. This solution was spun for 10 mins at 14000rpm to pellet DNA. The EtOH was removed by pouring off, and rinsed with 400uL 70% EtOH, and spun for further 5 mins at 14000rpm. The EtOH was removed, and DNA allowed to dry in an incubator before re-suspending the DNA in 50uL of distilled water. DNA was run out on a gel to ensure DNA was of high molecular weight (not fragmented).

The cytochrome oxidase subunit-I (COI) gene region of mitochondrial DNA was amplified using the universal primer pair LCO - 1490: 5′-GGTCAACAAATCATAAAGATATTGG-3′, and HCO 2198: 5′-TAAACTTCAGGGTGACCAAAAAATCA-3′ (Folmer et al. 1994). This COI region is commonly

used in studies of invertebrate diversity for three reasons: (1) the same primers anneal to the DNA of a wide range of organisms resulting in successful amplification (2) mitochondria are haploid, so a single allele/haplotype is amplified per individual making sequencing uncomplicated (3) the gene accumulates synonymous substitutions rapidly providing information to infer phylogenies and to differentiate species. This region of the COI gene has proved widely useful in species level studies of insects and is applied in DNA barcoding studies (Hebert et al., 2003; Ward et al., 2009).

The DNA (1uL) was added to a PCR master mix: 1uL 10x buffer dream taq, 1uL dNTP's 2mu per nucleotide, 0.4uL of 1490 primer, 0.4uL of 2198 primer and 0.1uL Taq enzyme (added in that order). For a volume of 10uL I set the following protocol: (1) Activation of the polymerase at 95°C for 3 min. (2) Thermal cycling of 36 cycles of denaturation at 95°C for 20s, annealing at 48°C for 15s and elongation at 72°C for 1min. (3) Extension at 72°C for 5mins. I used the same lab, machinery, and equipment for all specimens. Amplified COI fragments were run on a gel to check PCR was successful and these products were then sequenced using Big- Dye chemistry (Perkin Elmer) following the manufacturer's protocols on an ABI3730 DNA analyser (Macrogen Inc).

3.2.3. Sequencing and analysis

DNA sequences were trimmed and translated to ensure there are no stop codons using Geneious Prime® (2021.2.2 Build 2021-07-19 12:20 Java Version 11.0.11+9 (64 bit)) (Kearse et al 2012). Trimmed sequences were aligned, and unique sequences were extracted to identify unique haplotypes. The Geneious tree builder was used to explore relationships between populations and species. Outgroup sequences were obtained from GenBank using a Blast search of available data, as no other representatives of the mayfly family Ichthybotidae exist. The most similar available COI sequences were for Ephemeridae *Ephemera serica* (OK018134) and *Hexagenia limbate* (HM137936) and Neoephemeridae *Potamanthellus edmundsi* (GU714180).

Alignments and trait matrices were imported into PopART (Leigh & Bryant, 2015) to generate networks (Bandelt et al. 1999) representing haplotype diversity within each of the putative *Ichthybotus* species. Haplotype diversity and nucleotide diversity were calculated using DNA Sequence Polymorphism (DnaSP, Version 6.12.03 2018-02-22 (64 bit)) (Rozas 2009, Rozas et al 2017). for all population samples with a sample size of 1 or above. Population samples smaller than 5 were excluded from the next stages of analysis, as their small size makes their summary statistics unreliable. Population genetic structure is expected to reflect historical processes such as species limits, expansion from refugia, bottleneck and connectivity. Population sample haplotype diversity (Hd) and nucleotide site diversity (π) were compared using R ANOVA models. (R Core Team, 2021, Girden, 1992). Within population genetic diversity was estimated using haplotype diversity (*Hd*) and nucleotide diversity (π). Haplotype diversity is defined as the probability that two randomly selected alleles/haplotypes are different.

Nucleotide diversity measures a population's degree of polymorphism. It is expressed as the average number of nucleotide differences per site in a pairwise comparison of DNA sequences.

The relationship between geographic distance and genetic diversity was examined in the better-sampled North Island species, *I. hudsoni* using population sample latitude and Hd values, and pairwise Φ_{ST}. The relationship between latitude and Hd within *I. hudsoni* was examined using R linear models. Genetic structure resulting from gene flow would leave a signature of isolation by distance (IBD). Pairwise genetic distances (Φst) were calculated using DNAsp, and geographic distances were calculated using the program Geographic Distance Matrix Generator Version 1.2.3. (Ersts, 2006) These data sets were used to test for isolation by distance by assessing the correlation between the pairwise genetic matrix and geographic distances (Table 3.2) using a Mantel test (Mantel, 1967, Diniz-Filho et al 2013) implemented in IBD Version 1.52 (Bohonak, 2002). Finally, using the suggested substitution rate of invertebrates and known divergence rates, the divergence time of the two species lineages was calculated (Morgan-Richards et al, 2019, Trewick et al 2022).

3.3. Results

3.3.1. Genetic diversity

An alignment (617bp) of 188 individual mtDNA sequences revealed 120 unique COI haplotypes (Figure 3.3 and 3.5). The North Island mayfly *Ichthybotus hudsoni* had 93 haplotypes (n = 136) (Supplementary Table 3.1) while I. *bicolor* had 27 (n = 52) (Supplementary Table 3.2). The two *Ichthybotus* species have distinct mtDNA haplotypes supporting the current taxonomy of the two species. Despite sampling from the northern South Island, and the Taranaki region, where the two islands were once connected by a land bridge there are no shared haplotypes as expected in interspecific gene flow is occurring (Figure 3.5). Most haplotypes sampled were restricted to few populations from within the same geographical region.

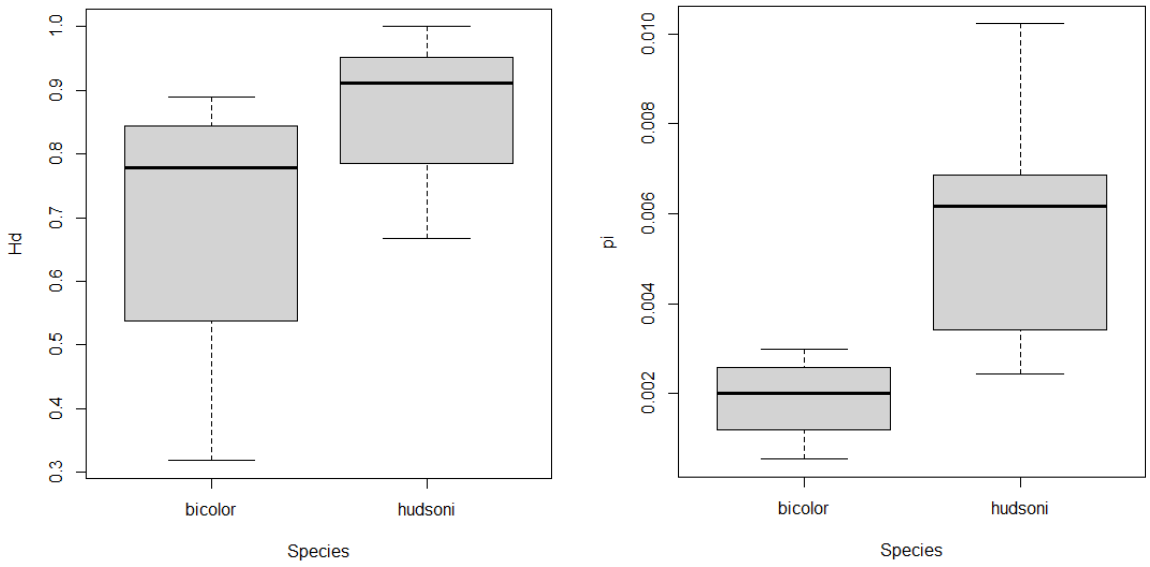


Figure 3.2. Boxes plots showing (left) the means and quartiles of Hd of *I. bicolor* and *I. hudsoni* and (right) the means and quartiles of π of *I. bicolor* and *I. hudsoni*.

Within population genetic diversity was estimated using haplotype diversity (H) and nucleotide diversity (π). Haplotype diversity is defined as the probability that two randomly selected alleles/haplotypes are different. Nucleotide diversity measures a population's degree of polymorphism. It is expressed as the average number of nucleotide differences per site in a pairwise comparison of DNA sequences. In both measures of diversity *I. hudsoni* scores highest. There was a significant difference between the total sampled haplotype diversity (Hd) and nucleotide diversity (π) values of the two species (Figure. 3.2). *I. hudsoni* had a higher mean Hd (0.87782) than *I. bicolor* (0.75209) (ANOVA; $F = 4.8509$, $p = 0.04369 < 0.05$). The results of an ANOVA comparing π was the same with *I. hudsoni* having a significantly higher nucleotide diversity ($F = 7.4421$, $p = 0.01556 < 0.05$) (Figure 3.2.).

The two Ichthybotus species are more closely related to each other than they are to the outgroups. There are seven main clades in the Ichthybotus species, four in *I. hudsoni* and three in I. bicolor (Figure 3.3). The first and second *I. hudsoni* clades are found exclusively in ND and AK. The third is found in AK, CL, TK, WI and WN. The majority of clade three was found in AK and the CL, but this could be due to a larger sample size in these areas. The fourth clade is found in one location in AK, CL, TK, WI and WN. These clades are also shown in the network (Figure 3.5). It can be inferred from the phylogenetic tree that the lineage split into the two species. The *I. hudsoni* may have split into three lineages 1H, and 2+3+4H, before splitting into 2H and 3+4H. The first two clades are exclusively in ND and AK. The last clade then possibly split into 3H and 4H, with 3H being more likely in Ak and CL, and 4H being more likely in TK, WI and WN (Figure 3.3). The first *I. bicolor* clade is found in SD, BR, KA, MC, and OL. The second was found in NN, SD and MB. The third was found exclusively in NN (CW

specifically). The phylogenetic tree suggests that the *I. bicolor* may have split into lineages 1B (found exclusively in CW in NN) and 2+3B (found in the rest of the South Island). The other clade may have then split into 2B, present in NN, SD and MB (i.e., The northern South Island, and 3B present in SD, KA, MC, and OL (i.e., northern, mid and southern South Island).

The average and maximum interspecific genetic distances were 0.041 (4.1%) and 0.051 (5.1%) respectively. Within *I. hudsoni* sampled haplotypes differed by a maximum of 2.7% and within *I. bicolor* a maximum of 1.2%. Using the insect intraspecific divergence rate of 0.0792 substitutions per site, per million years (Morgan-Richards et al, 2019) and an insect interspecific substitution rate of 0.115 per site per My (Popadopoulou et al, 2010) the divergence of the two species lineages was calculated. This resulted in an estimate of less than 1 Mya between the two species (5.7%) and ~300,000 ya within each species (2.7% - 1.2%).

3.3.2. North Island *I. hudsoni*

The North Island species of burrowing mayfly (*I. hudsoni*) was collected from 17 locations with mean sample size of 8 (Table 3.1). All populations, excluding OT and PK (both n=1), had unique haplotypes (Supplementary Table 3.1; 15 populations out of 17). Most populations (13) had more unique haplotypes (haplotypes only found in that population) than shared haplotypes. Sites PS (n= 12), CA (n=1) and WA (n=5) consisted of entirely unique haplotypes. Out of 73 *I. hudsoni* haplotypes 54 were only found in one location. The most widely encountered haplotype, Ih_11, was found in 6 population samples in Northland (ND) and Auckland regions AK). Two haplotypes (Ih_33 and Ih_44) were found in 4 population samples and 2 (Ih_56 and Ih_69) in three. Within the *I. hudsoni* sample, most haplotypes differed from one another by one or few mutations with the maximum pairwise difference was 16 between Ih_19(RF) - Ih_64(WA) (Figure 3.3). The average genetic distance across all pairwise haplotype comparisons for *I. hudsoni* was 0.012 (1.2%), with the maximum 0.027 (2.7%).

The evolutionary relationships inferred for the haplotypes (see network, Figure 3.5) reveal four haplotype clusters denoted 1H – 4H. Clusters 1H and 2H comprise specimens from Northland and Auckland only, with no obvious separation between these two regions. Haplotype Ih_11 is the most common in these specimens, consisting mostly of ND specimens. Cluster 4H is made up of specimens from all over the NI, excluding Northland. Auckland and CL samples have haplotypes that form cluster 3H, with the most common haplotype being Ih_45. Southern NI regions (TK, WI and WN) only have haplotypes from a single cluster (4H). Ih_69 and Ih_56 are the most common haplotypes in Southern NI specimens. Upon comparison with the phylogenetic dree diagram Ib_52, Ib_53 and Ib_54 are likely to be a part of 3H.

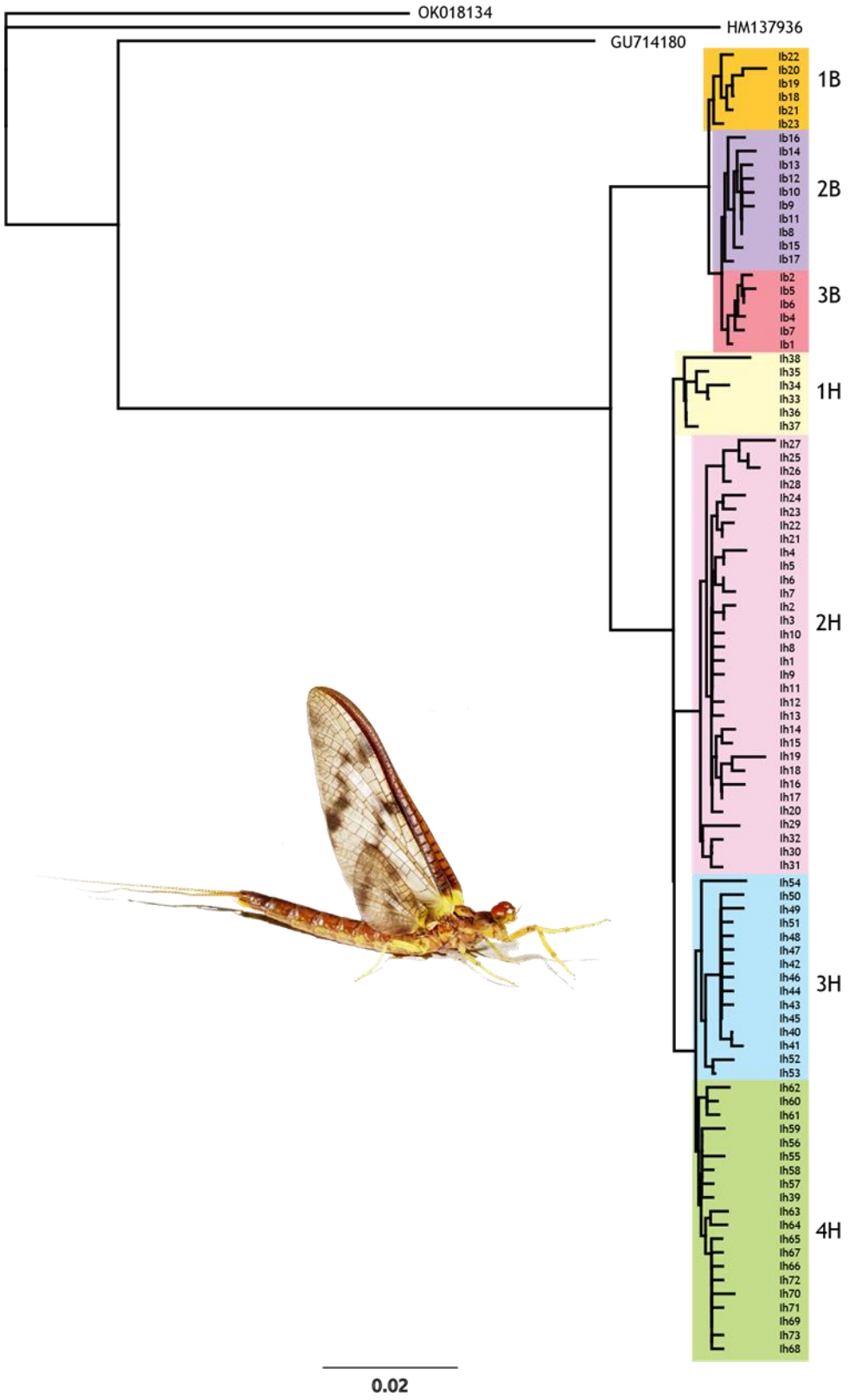


Figure 3.3: Phylogenetic tree of all mtDNA COI haplotypes (617 bp) from two New Zealand mayfly species (*Ichthybotus* sp), and three outgroup specimens Ephemeridae *Ephemera serica* (OK018134) and *Hexagenia limbate* (HM137936) and Neoephemeridae *Potamanthellus edmundsi* (GU714180). Tree inferred using Maximum Likelihood. Ih represents *I. hudsoni* haplotypes, while Ib is *I. bicolor* haplotypes. Colours represent different clades within species from top to bottom: 1B, 2B, 3B, 1H, 2H, 3H, and 4H.

Table 3.1. Summary of variation among 617bp mtDNA COI for population samples of *I. hudsoni and I. bicolor*, organised by island and region. Shows locations (ordered from north to south) and genetic diversity indices for each site and species, and total data estimates (π, nucleotide diversity; h, number of haplotypes; k, average number of nucleotide differences; Hd, haplotype diversity). Species and total data estimates are calculated from all populations n>1.

Region	Region Code	Site Name	Site Code	Lat	Long	Species	n	h	k	Hd	π
Northland	ND	Whangaroa	WH	-35.01	173.71	*hudsoni*	7	6	2.0952	0.9523	0.0034
		Mangamuka Gorge (Tapapa Stream)	MG	-35.19	173.47	*hudsoni*	12	7	3.7878	0.8333	0.0061
		Russell Forest (Maruku Stream)	RF	-35.39	174.31	*hudsoni*	14	11	4.2197	0.9560	0.0068
		Waipoua River	WR	-35.65	173.55	*hudsoni*	10	7	4.0444	0.9111	0.0065
		Mangere Stream	MS	-35.7	174.24	*hudsoni*	8	8	6.9333	1.0000	0.0094
Auckland	AK	Pohuehue Stream	PS	-36.45	174.65	*hudsoni*	12	7	5.8333	0.8939	0.0094
		Casades	CA	-36.88	174.52	*hudsoni*	1	1			
		Huia	HU	-37	174.55	*hudsoni*	8	7	6.2857	0.9642	0.0102
		Hunua Ranges	HR	-37.11	175.12	*hudsoni*	7	6	3.0476	0.9523	0.0049
Coromandel	CL	Fantail Bay	FB	-36.52	175.33	*hudsoni*	12	6	2.1969	0.7575	0.0035
		Te Puru	TP	-37.04	175.53	*hudsoni*	7	6	3.9047	0.9523	0.0063
Taranaki	TK	Otakeho Stream	OT	-39.4	174.06	*hudsoni*	1	1			
		Ratapihipihi reserve (Pukeiti)	PK	-39.19	173.97	*hudsoni*	1	1			
		Ratapihipihi reserve (Hurdon)	HD	-39.1	174.04	*hudsoni*	11	5	1.7090	0.7636	0.0027
Whanganui	WI	Kahutarawa	KA	-40.46	175.61	*hudsoni*	12	5	1.6818	0.6666	0.0027
Wellington	WN	Otari-Wiltons	OW	-41.26	174.75	*hudsoni*	8	5	1.5000	0.7857	0.0024
		Wainuiomata	WA	-41.27	174.76	*hudsoni*	5	4	2.6000	0.9000	0.0042

Nelson	NN	Collingwood (Kaituna River)	CW	-40.71	172.56	*bicolor*	9	6	1.8333	0.8888	0.0029
		Lyell	LY	-41.79	172.05	*bicolor*	1	1			
Marlborough Sounds	SD	Harvey Bay Stream	HB	-41.12	173.73	*bicolor*	12	3	1.1363	0.7575	0.0018
		Robertson Ranges (Pukaka Stream)	RR	-41.38	174.01	*bicolor*	12	3	0.3333	0.3181	0.0005
Marlborough	MB	Richmond Ranges (Elvy Stream)	RM	-41.3	173.55	*bicolor*	6	4	1.3333	0.8000	0.0021
Buller	BR	Reefton	RE	-42.14	171.9	*bicolor*	1	1			
Kaikoura	KA	Mororimu	MO	-42.21	173.86	*bicolor*	1	1			
		Hawkswood	HW	-42.66	173.41	*bicolor*	3	3	1.3333	1.0000	0.0021
Mid Canterbury	MC	Okuti	OK	-43.78	172.83	*bicolor*	1	1			
		Hinewai	HN	-43.82	173.04	*bicolor*	2	2	2.0000	1.0000	0.0032
Otago Lakes	OL	Te Anau	TA	-45.13	167.94	*bicolor*	4	2	0.5000	0.5000	0.0008
I hudsoni Total Estimates							131	71	6.3262	0.9690	0.01029
I bicolor Total Estimates							48	21	2.2109	0.8359	0.00360
Total Data Estimates							179	92	12.9331	0.9718	0.02103

3.3.3 South Island *I. bicolor*

The South Island species of burrowing mayfly (*I. bicolor*) was collected from 11 locations with mean sample size of 4.7 (Table 3.1). Seven populations out of 11 had unique haplotypes (Supplementary Table 3.2). Around half of the populations (6, including HN in which ½ of the haplotypes were unique) had more unique haplotypes than shared haplotypes found in other populations. There were 4 populations in *I. bicolor* that had no unique haplotypes: LY, MO, OK and TA (all with n ≤2). Sites CW and RE consisted of entirely unique haplotypes. Out of 22 *I. bicolor* haplotypes 19 were only found in one location. The maximum number of locations that any haplotype was found in was Ib_6 and Ib_8, in 4 locations. There was one haplotype with 3 locations (Ib_2). The network (Figure 3.5) shows that these haplotypes differed from one another by between 1 and 7 mutations with the maximum pairwise difference between; Ib_4(HW) and Ib_20(CW), Ib_4(HW) and Ib_23(CW), Ib_5(HN) and Ib_20(CW), Ib_5(HN) and Ib_23(CW), Ib_14(RM) and Ib_22(CW), and Ib_16(HB) and Ib_22(CW). The average genetic distance across all pairwise haplotype comparisons for *I. bicolor* was 0.006 (0.6%), with the maximum 0.012 (1.2%). Less haplotype diversity was observed in *I. bicolor* compared to the northern species, *I. hudsoni*.

The evolutionary relationships inferred for the haplotypes (see network, Figure 3.5) reveal three main clusters: Cluster 1B is made up of low frequency NN haplotypes in the furthest northwest SI. Cluster 2B is a mix of SD, NN, MB and BR. This cluster includes the most common SI haplotype, Ib_8, with 20 specimens from NN, SD and MB, i.e., from the northwest SI. However, this is where most of the specimens are from so a strong bias toward this area would explain this trend. This cluster is widespread geographically, encompassing the northwest and central SI. This cluster connects to 3B via haplotypes from MC, MB, SD, and BR. Cluster 3B contains haplotypes form KA, MC, and OL, representing all the southern South Island haplotypes and the eastern South Island. Comparing the phylogenetic tree to the haplotype network, Ib_1 and Ib_2 are likely to be a part of 3B.

Within I. bicolor, nucleotide diversity was highest in the north of its range (Collingwood; $\pi = 0.00298$) but this level of diversity is like the lower estimates for I. hudsoni population samples in the south (Table 3.1).

3.3.4. *I. hudsoni* Results: Isolation by Distance

The following analyses were carried out on *I. hudsoni* data and included sites with at least six specimens. There was a linear relationship found between latitude and nucleotide diversity (π) ($t = 2.135$, $p = 0.0561$) (Figure 3.4). It shows that the lower the latitude (or the closer to the equator) the higher π is expected to be, i.e., there is higher nucleotide diversity in *I. hudsoni* populations in the north of the North Island, compared to the south. A mantel test carried out on *I. hudsoni* pairwise population data (Table 3.2) suggests that there is a correlation between the geographic distance between samples and

the genetic distance (Z = 10380.7834, r = 0.6622, $p \leq 0.001$) (Figure 3.4). Therefore, there seems to be increased genetic diversity across longer distances and more similar haplotypes are generally closer together.

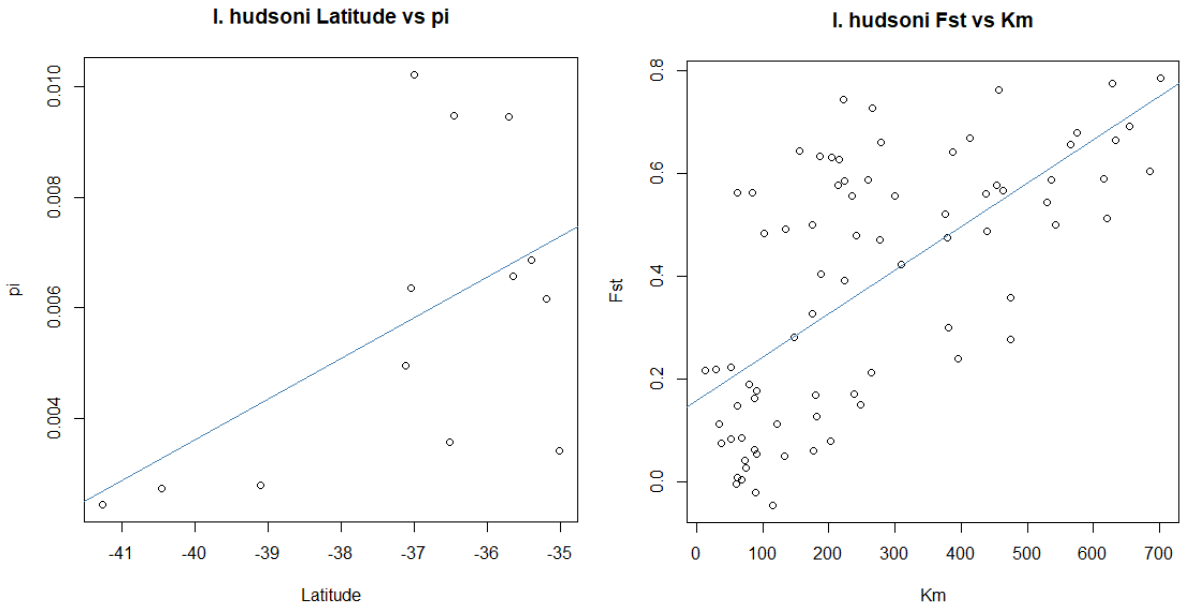


Figure 3.4. Scatterplot showing the relationship between *I. hudsoni* site latitude and nucleotide diversity (left). Each black dot represents one site. There is a positive relationship shown by the blue trendline. Scatterplot showing the relationship between pairwise geographic distance (Kms) and genetic distance (Φ_{ST}) for all *I. hudsoni* sample populations with at least 6 individuals (right). Each pairwise population comparison (Table 3.2) is represented by a dot, and the linear trendline in blue shows a positive relationship.

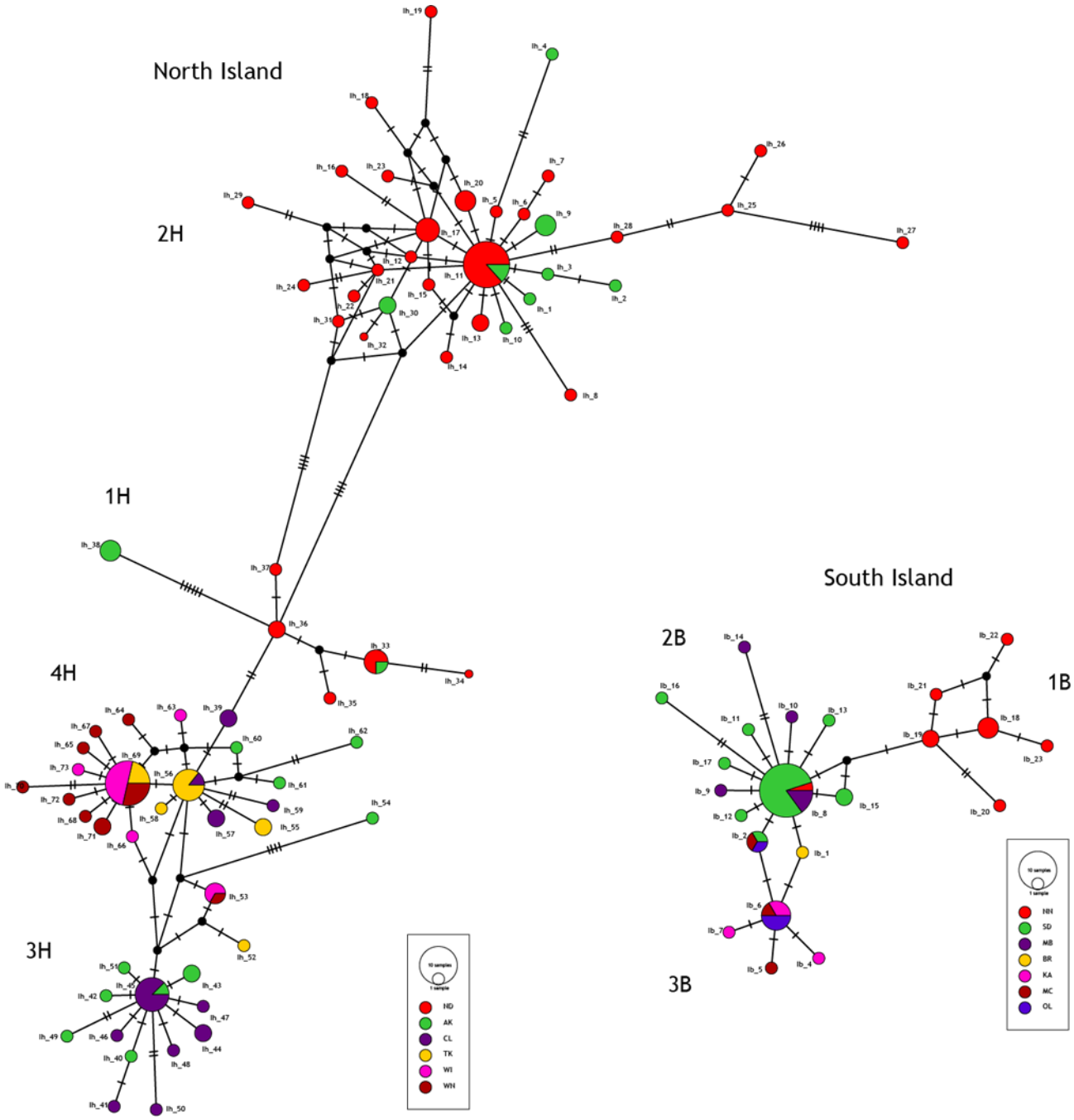


Figure 3.5. Haplotype networks of *I. hudsoni* (North Island) and *I. bicolor* (South Island) (617 bp of COI mtDNA). The nodes represent the haplotypes present in each species, all labelled with the assigned haplotype number. Node colour indicates the Crosby region, node size shows the number of samples with that haplotype. Small black dots at a node represent a haplotype that was not sampled. Clusters are labelled to match the clades in the phylogenetic tree, 1H etc. for *I. hudsoni* and 1B for *I. bicolor*.

Table 3.2. Table of pairwise population data used for *I. hudsoni* mantel test. P1 = population 1, P2= population 2 i.e., population being compared, Φ_{ST} = measure of genetic differentiation among population samples, and Km = geographic distance between populations in kilometres.

P1	P2	Km	Φ_{ST}
WH	MG	29.65	0.21807
WH	RF	69.05	0.08451
WH	WR	72.71	0.04067
WH	MS	90.64	0.17565
WH	PS	181.41	0.12598
WH	HU	223.68	0.39206
WH	HR	265.98	0.72727
WH	FB	222.85	0.74429
WH	TP	279.12	0.66129
WH	HD	456.24	0.76376
WH	KA	629.28	0.77498
WH	OW	701.66	0.78582
MG	RF	79.51	0.18959
MG	WR	51.72	0.08215
MG	MS	90	-0.02107
MG	PS	176.12	0.05879
MG	HU	180.52	0.16919
MG	HR	260.14	0.5887
MG	FB	223.78	0.58487
MG	TP	276.99	0.47034
MG	HD	438.18	0.55971
MG	KA	616.03	0.58892
MG	OW	684.89	0.60464
RF	WR	74.69	0.0261
RF	MS	35.09	0.11282
RF	PS	121.91	0.11109
RF	HU	175.01	0.32688
RF	HR	204.81	0.63213
RF	FB	155.79	0.64445
RF	TP	213.87	0.57739
RF	HD	413.69	0.66966
RF	KA	575.8	0.68017
RF	OW	654.57	0.69142
WR	MS	62.64	0.00803
WR	PS	133.16	0.05021
WR	HU	147.36	0.28013
WR	HR	214.97	0.62841
WR	FB	187.14	0.63429
WR	TP	235.49	0.55556
WR	HD	386.49	0.64284
WR	KA	565.02	0.65574
WR	OW	633.18	0.666
MS	PS	91.28	0.05432
MS	HU	61.87	0.14716
MS	HR	175.65	0.50095
MS	FB	133.95	0.49119
MS	TP	188.73	0.40419
MS	HD	378.9	0.47574
MS	KA	543.29	0.49977
MS	OW	620.53	0.51346
PS	HU	13.62	0.21602
PS	HR	84.58	0.56374
PS	FB	61.36	0.56265
PS	TP	102.35	0.48424
PS	HD	299.84	0.55593
PS	KA	454.16	0.57661
PS	OW	535.52	0.58782
HU	HR	52.1	0.22222
HU	FB	87.72	0.1622
HU	TP	87.22	0.0614
HU	HD	238.01	0.17035
HU	KA	396	0.23816
HU	OW	474.53	0.27575
HR	FB	68.29	0.00329
HR	TP	37.24	0.07428
HR	HD	240.87	0.47973
HR	KA	375.33	0.52136
HR	OW	463.08	0.56689
FB	TP	60.57	-0.00499
FB	HD	308.79	0.42327
FB	KA	439.28	0.48851
FB	OW	530.04	0.54499
TP	HD	263.88	0.21119
TP	KA	380.78	0.29959
TP	OW	474.56	0.35876
HD	KA	202.38	0.07901
HD	OW	247.91	0.1494
KA	OW	114.77	-0.04608

3.4. Discussion

3.4.1. Interspecific Genetic Diversity

The distribution of haplotypes observed in samples of *I. hudsoni* and *I. bicolor* supports the idea that these represent two distinct reciprocally monophyletic lineages (Tillyard, 1923, Chrisholm, 1984, Pohe 2018). The genetic divergence estimated from mtDNA sequences between the species is around 0.051 (5.1%), which is relatively low when compared to similar studies of New Zealand insects (Trewick et al. 2011; Morgan-Richards et al. 2017). Estimated interspecific divergence of stream invertebrates using the same locus have reported between 18% (Ball et al 2005) and 24-32% (Beet, 2016). Given the morphological similarity of these two mayflies, there was a possibility that there would be no genetic divergence between the two species. However, my data reveals a genetic difference without evidence of gene flow between the two species, supporting separation of the species based on concordance of adult wing colour, mtDNA and geography (Tillyard, 1923, Phillips, 1930, Chrisholm 1984).

There were no haplotypes shared by the two species, with *I. hudsoni* constrained to the North Island and *I. bicolor* to the South Island. This implies there is no contact or gene flow between them. This supports the idea that Cook Strait acts as an effective barrier to dispersal for these mayflies. Before anthropogenic effects, aquatic insect species were most impacted by climate cycling and the changing landscape because of tectonic movement (Wallis & Trewick, 2009). The formation of straits is likely to be an important abiotic phylogeographic factor in this case (Trewick et al. 2011, Kumar & Kumar, 2018). The impact of a strait on dispersal depends on position, width and depth, and the ability of a species to disperse. During the last Glacial Maximum there was land connection between North and South Islands of New Zealand (Trewick & Bland, 2012). Therefore, it is unexpected that Cook Strait's influence on dispersal should be observed in the genetic structure of so many endemic species, including land snails (*Wainuia umula*) (Efford et al 2002), fungus beetles (*Epistranus lawsoni* and *Pristoderus bakewelli*) (Markse et al 2011), earthworms (Acanthodrilinae) (Buckley et al 2011), other mayflies *Acanthophlebia* (Trewick et al 2022) and both flighted and flightless birds (Trewick et al. 2017). The current marine strait does not explain why the two mayflies studied here did not connect and mix during the last glacial period when the two main islands would have been bridged (Trewick et al 2012). It is possible that land bridges during the last glacial maximum did not last long enough or were too dry to allow these species with low capacity for dispersal to expand across (Greaves et al 2007). Whether a strait acts as a barrier to dispersal depends on the width, depth and location of the strait (Kumar & Kumar 2018) i.e., a mayfly would have to travel 35km across the cook strait to disperse to another island, which appears sufficient to block dispersal. Competitive exclusion might be a better explanation of why the two mayfly species are each restricted to a single main island. My ecological niche modelling suggests it is not purely climatic differences that exclude the overlap of the two species.

The timeframe inferred from the interspecific divergence rate (somewhere between 300,000 ya and <1 Mya) is several glacial cycles within the latter part of Pleistocene when southern North Island was mostly emerged (Trewick & Bland, 2012). This might suggest that the niche model for LGM applies to previous glacial cycles i.e., ancestral *Ichthybotus* may have been able to spread across both islands, getting sundered by interglacial periods. If this occurred a few cycles back, subsequent connection and potential gene flow might not be evident if occupying population swamped any genetics coming in (rarely) from other island.

3.4.2. Intraspecific Phylogeography of *I. hudsoni*

In New Zealand stoneflies a species with "strong flight mediated dispersal", *Z. decorata*, had less genetic differentiation over its range when compared to the *fenestrata* species group with weaker flight dispersal (McColloch, 2010). With mayflies such as *Ichthybotus* with weak dispersal one expects to detect population genetic structure as their lack of long range dispersibility would contribute to isolation by distance. The average intraspecific distance for *I. hudsoni* was 1% which is the same as previously reported averages, around 1% in mayflies (Ball et al 2005) and 1.2 - 1.4% for caddisflies, stoneflies, and mayflies (Beet, 2016). The average intraspecific distance for *I. bicolor* was around 0.03% which is below the average for mayflies.

Within *I. hudsoni* I observed decreasing nucleotide diversity from the north to the south. This cline in genetic diversity suggests refuge in the north followed by a range expansion south as observed in the mayfly *Acanthophlebia cruentata* (Trewick et al. 2022). In two wētā species, *Hemideina crassidens* (southern North Island and South Island) and *H. thoracica* (North Island) both have recently expanded south, after seeming to be restricted to more northern habitats (Bulgarella et al 2013). This was attributed to contracting away from glaciation and then expanding back south during the current interglacial period. This is likely to be the case with *I. hudsoni*. Higher diversity in the north suggests that populations in this area have either persisted for longer or have been significantly bigger than southern populations. This is also seen in many other New Zealand species, having been called 'northern diversity, southern purity' by Ellis et al (2015). The *Argosarchus horridus* and *Clitarchus hookeri* stick insects are both widespread across the country. Both species display high diversity in the northern North Island and lower diversity in southern North Island and South Island (Buckley et al 2009, Buckley et al 2010; Morgan-Richards et al. 2010). Cicadas (*Kikihia sabalpina*) expanded from glacial refugia in the north of the South Island south, indicated by common haplotypes either side of the southern alps (Marshall et al 2009, Bulgarella et al 2013). Tree wētā and giraffe weevils also show nucleotide diversity decreases southward, suggesting northern refugia. However, the isolation by distance analysis makes this even more interesting. There is a clear signal of isolation by distance (IBD) within the *I. hudsoni* population samples, i.e., populations are genetically isolated having ample time to differentiate since coalescence. IBD considers the direct or straight-line geographic distance between populations,

with no accounting for landscape/geographic barriers to dispersal which would have a significant effect on the ease of dispersal (van Strien et al 2015). When range expansion has been very recent (e.g., in the mayfly *Acanthophlebia cruentata*; Trewick et al. 2022) then population divergence has not accumulated, and gene flow cannot be detected as IBD. Thus, detection of isolation by distance suggests stable populations have maintained connections via gene flow over time. From which one might infer that southern populations have existed in their current positions long enough to accumulate genetic differences, but not long enough, or at great enough numbers to have the same level of genetic diversity as northern populations. This would only be possible if glacial cycles did not completely remove them from southern and central North Island. (This would also mean that *I. hudsoni* and *I. bicolor* would have come close to interacting if we assume that *I. bicolor* had the same trend of northern populations, like other species). The decrease in genetic diversity at higher latitudes of the North Island could be due to the southern north island being relatively young, only emerging from the sea around 1.5 Ma (Trewick & Bland, 2012, Taylor-Smith et al., 2019). If genetic diversity is related to a populations long term stability, there is the possibility that these populations have not had as much time to diversify. Therefore, *I hudsoni* may have been restricted to the north of the North Island until the southern land surfaced, then moved south until the LGM was reached when population numbers in the south declined leaving some refugia. Ecological Niche Modelling using climatic variables suggests that large areas of the North Island may have provided suitable habitat for *I hudsoni* when it was cooler, but this is not supported by the pattern of population genetic diversity. There is no evidence of lower population genetic diversity in population samples from the central North Island when compared to the north and south, so it is not clear if volcanic activity (specifically the Taupo eruptions) has influenced population history and genetic structure. *I. hudsoni* do occur in this area (see ENM chapter) but no specimens were available for DNA sequencing.

3.4.3. Is the population genetic structure compatible with inferences from ENM?

Ichthybotus hudsoni showed more genetic diversity than *I. bicolor* which would not be inferred from current distributions but could be predicted by ecological niche models projected into the past. The lower levels of mtDNA divergence within *I. bicolor* suggests the southern species may have had a more restricted range in the past than its sister species *I. hudsoni*. Larger population samples of *I. bicolor* are needed to determine the pattern and extent of population genetic diversity but current data suggests there is spatial genetic structure within this species. Whether the pattern inferred from ecological niche model projections for the LGM exists within *I. bicolor* requires more sampling but current analysis of mtDNA does not suggest range expansion from the east towards the west.

The genetic data presented here provides no evidence that either of the two mayfly species have ever occurred in the alternative main island even though ENM suggested the climate was suitable. Further study to investigate whether competitive exclusion is the explanation would be interesting. ENM

indicated that during the LGM there was potential habitat in most of the North Island and much of the South Island for *I. hudsoni*. This contrasts with the signature of range expansion from the north seen in the genetic data. It is likely that climate variables are not a full representation of the factors limiting the distribution of these mayflies. Incorporating other features of the landscape during the LGM might improve estimates of potential past distribution.

Supplementary Table 3.1. North Island *I. hudsoni* CO1 haplotypes (73) and their locations. The row h represents the number of unique haplotypes found in the given location.

Region	ND					AK				CL		TK			WI	WN	
Site	WH	MG	RF	WR	MS	PS	CA	HU	HR	FB	TP	OT	PK	HD	KA	OW	WA
n	7	12	14	10	8	12	1	8	7	12	7	1	1	11	12	8	5
Ih 1							1										
Ih 2						1											
Ih 3						1											
Ih 4						1											
Ih 5					1												
Ih 6	1																
Ih 7	1																
Ih 8				1													
Ih 9						3											
Ih 10						1											
Ih 11	2	5	3	2	1			2									
Ih 12		1															
Ih 13		1	1														
Ih 14					1												
Ih 15					1												
Ih 16					1												
Ih 17			1	3													
Ih 18			1														
Ih 19			1														
Ih 20			2	1													
Ih 21	1																
Ih 22	1																
Ih 23	1																
Ih 24				1													
Ih 25			1														
Ih 26			1														
Ih 27			1														
Ih 28			1														
Ih 29				1													
Ih 30						2											
Ih 31			1														
Ih 32					1												
Ih 33		1		1	1			1									
Ih 34					1												
Ih 35		1															
Ih 36		2															
Ih 37		1															
Ih 38						3											
Ih 39											2						
Ih 40									1								
Ih 41										1							
Ih 42									1								

lh 43									2								
lh 44											1						
lh 45								1	1	6	1						
lh 46										1							
lh 47										1							
lh 48											1						
lh 49								1									
lh 50											1						
lh 51									1								
lh 52														1			
lh 53															2	1	
lh 54									1								
lh 55												1		1			
lh 56										1			1	5			
lh 57										2							
lh 58														1			
lh 59											1						
lh 60								1									
lh 61								1									
lh 62								1									
lh 63															1		
lh 64																	1
lh 65																	1
lh 66															1		
lh 67																1	
lh 68																	2
lh 69														3	7	4	
lh 70																	1
lh 71																1	
lh 72																1	
lh 73															1		
h	6	7	11	7	8	7	1	7	6	6	6	1	1	5	5	5	4

Supplementary Table 3.2. South Island *I. bicolor* CO1 haplotypes (22) and their locations. The row h represents the number of unique haplotypes found in the given location.

Region	NN		SD		MB	BR	KA		MC		OL
Site	CW	LY	HB	RR	RM	RE	MO	HW	OK	HN	TA
n	9	1	12	12	6	1	1	3	1	2	4
Ib_1						1					
Ib_2				1						1	1
Ib_4								1			
Ib_5										1	
Ib_6							1	1	1		3
Ib_7								1			
Ib_8		1	6	10	3						
Ib_9					1						
Ib_10					1						
Ib_11				1							
Ib_12			1								
Ib_13			1								
Ib_14					1						
Ib_15			2								
Ib_16			1								
Ib_17			1								
Ib_18	3										
Ib_19	2										
Ib_20	1										
Ib_21	1										
Ib_22	1										
Ib_23	1										
h	6	1	6	3	4	1	1	3	1	2	2

Chapter 4: General discussion

Ecological Niche models have not been applied to many aquatic insects and the current dataset may not have been ideal. The total number of presence locations used in this study was below recommendations for *Ichthybotus bicolor*. Although true absence data was used, the likelihood that competition as well as climate influences the presence/absence of these two closely related mayflies may have impacted the models inferred. The inferred LGM potential distribution for *I. hudsoni* covered a large climatic range. The model suggested suitable habitat for *I. hudsoni* was widespread in northern and western North Island, with potential habitat over the land bridge, and even at the southern end of the South Island. Interestingly, current models for this species are narrower, with most likely habitat in northern North Island, and some other potentially suitable habitat in the rest of lowland North Island and just the northern tip of South Island. The potential LGM niche for *I. bicolor* was inferred to be most likely along the east coast of what is now the South Island, in the Whanganui Basin, and the northern tip of the north island. This is also predicted to have decreased since the LGM, with fragmented habitat remaining along the east coast of the South Island and the lower north island. Both current and LGM niche models suggest that there was (and is) climatically suitable habitat in both islands and the potential for overlap of the two species, especially around the Cook Strait land bridge/Whanganui Basin.

If the current predicted distributions were realised some genetic crossover would be expected in North Island, New Zealand. This was not observed. There were no shared haplotypes found between the two species in the analysis of mtDNA haplotypes. This discrepancy in results is likely due to the ENM not being impacted by landscape features that are barriers to dispersal such as Cook Strait, and the models being an indicator of suitable climate with no biotic factors considered. The positive outcome from the current ENM is the suggestion that a focus on competitive exclusion interactions would be a fruitful avenue of research.

If the projections of ENM for the two mayfly species onto the climate of the LGM are realistic estimates of past distributions for these two *Ichthybotus* species, then it would be expected that *I. hudsoni* would have retained high genetic diversity throughout its North Island range with potential for greater diversity in the northwest of North Island. This is partly supported by the phylogeographic data, it seems that genetic diversity was highest in the north and decreased toward the south. This species had average to just above average genetic divergence compared to insects in other studies, possibly due to long term populations being maintained across the whole North Island (populations being larger in the north). Current projected niche space of the two *Ichthybotus* species was similar. However, *I. hudsoni* had a much larger projected niche space during the LGM. The link between long-term population and distribution size, and genetic diversity (Frankham, 1996) is that large distributions usually contain higher genetic diversity. The lower genetic diversity would also lead one to infer a narrower distribution for *I. bicolor* compared to *I. hudsoni*.

The interspecific divergence timeframe inferred from the divergence rate was around 1 Mya. This does not line up with the end of the LGM and the formation of the Cook Strait. If these species distributions possibly crossed over during the most recent LGM we might expect to see some shared haplotypes around the southern North Island. This was not the case. This might suggest that the niche model for the LGM applies to the distribution during previous glacial cycles i.e., ancestral Ichthybotus may have been able to spread across both islands (perhaps in a distribution similar the *I. hudsoni's* LGM predicted distribution), getting separated by interglacial periods. Following connections may have allowed for potential gene flow, but not enough to withstand getting swamped by the already established population.

Are Ecological Niche Models and phylogeography useful tools for aquatic insects?

Despite the limitation of the variables used in these ecological models, the distribution of current potential habitat inferred from the models were moderately successful. The difficulty incorporating biotic interactions such as competition might explain the inference of both species currently occurring in both main New Zealand islands. The phylogeographic inferences made were also useful for comparison with ENM's and exploration of spatial diversity. The COI mtDNA sequences were able to distinguish between the two *Ichthybotus* species accurately and allowed inferences to be made about within species genetic diversity.

References

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology, 43*(6), 1223-1232.

Araujo, M. B., Anderson, R. P., Marcia Barbosa, A., Beale, C. M., Dormann, C. F., Early, R., . . . Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Scientific Advances, 5*(1), 1-11.

Araujo, M. B., & New, M. (2006). Ensemble forecasting of species distributions. *Trends in Ecology and Evolution, 22*(1), 42-47.

Atkinson, E. (2006). *Phylogenetics, divergence and morphology of New Zealand Eleotridae (Gobiomorphus Gill)*. (Master of Science in Ecology). Massey University,

Avise, J. C. (2001). Phylogeography: The history and formation of species. *American Zoologist, 41*(1), 133-135.

Avise, J. C. (2009). Phylogeography: retrospect and prospect. *Journal of Biogeography, 36*(1), 3-15.

Ball, S. L., Hebert, P. D. N., Burian, S. K., & Webb, J. M. (2005). Biological identifications of mayflies (Ephemeroptera) using DNA barcodes. *Journal of the North American Benthological Society, 24*(3), 508-524.

Bandelt, H.-J., Forster, P., & Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution, 16*(1), 37-48.

Barber-James, H. M., Gattolliat, J.-L., Sartori, M., & Hubbard, M. D. (2008). Global diversity of mayflies (Ephemeroptera, Insecta) in freshwater. *Hydrobiologia, 595*(1), 339-350.

Beaumont, L. J., Graham, E., Duursma, D. E., Wilson, P. D., Cabrelli, A., Baumgartner, J. B., . . . Van Der Wal, J. (2016). Which species distribution models are more (or less) likely to project broad-scale, climate-induced shifts in species ranges? *Ecological Modelling, 342*, 135-146.

Beet, C. R. (2016). *Assessing the Diversity of Antarctic and New Zealand Arthropods through DNA Barcoding.* (Masters). University of Otago,

Beheregaray, L. B. (2008). Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology, 17*(17), 3754-3774.

Bisconti, R., Canestrelli, D., Tenchini, R., Belfiore, C., Buffagni, A., & Nascetti, G. (2016). Cryptic diversity and multiple origins of the widespread mayfly species group *Baetis rhodani* (Ephemeroptera: Baetidae) on northwestern Mediterranean islands. *Ecology and Evolution, 6*(21), 7901-7910.

Bohonak, A. J. (2002). IBD (Isolation By Distance): a program for analyses of isolation by distance. *Journal of Heredity, 93*, 153-154.

Boulangeat, I., Gravel, D., & Thuiller, W. (2012). Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. *Ecology Letters, 15*(6), 584-593.

Bouska, K. L., Whitledge, G. W., & Lant, C. (2015). Development and evaluation of species distribution models for fourteen native central U.S. fish species. *Hydrobiologia, 747*(1), 159-176.

Breiner, F. T., Guisan, A., Bergamini, A., Nobis, M. P., & Anderson, B. (2015). Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution, 6*(10), 1210-1218.

Buckley, T. R., James, S., Allwood, J., Bartlam, S., Howitt, R., & Prada, D. (2011). Phylogenetic analysis of New Zealand earthworms (Oligochaeta: Megascolecidae) reveals ancient clades and cryptic taxonomic diversity. *Molecular Phylogenetics and Evolution, 58*(1), 85-96.

Buckley, T. R., Marske, K., & Attanayake, D. (2010). Phylogeography and ecological niche modelling of the New Zealand stick insect *Clitarchus hookeri* (White) support survival in multiple coastal refugia. *Journal of Biogeography, 37*(4), 682-695.

Buckley, T. R., Marske, K. A., & Attanayake, D. (2009). Identifying glacial refugia in a geographic parthenogen using palaeoclimate modelling and phylogeography: the New Zealand stick insect *Argosarchus horridus* (White). *Molecular Ecology, 18*(22), 4650-4663.

Bulgarella, M., Trewick, S. A., Minards, N. A., Jacobson, M. J., Morgan-Richards, M., & Kissling, W. D. (2013). Shifting ranges of two tree weta species (*Hemideina* spp.): competitive exclusion and changing climate. *Journal of Biogeography, 41*(3), 524-535.

Carmelet-Rescan, D., Morgan-Richards, M., Koot, E. M., & Trewick, S. A. (2021). Climate and ice in the last glacial maximum explain patterns of isolation by distance inferred for alpine grasshoppers. *Insect Conservation and Diversity, 14*(5), 568-581.

Cengic, M., Rost, J., Remenska, D., Janse, J. H., Huijbregts, M. A. J., & Schipper, A. M. (2020). On the importance of predictor choice, modelling technique, and number of pseudo-absences for bioclimatic envelope model performance. *Ecology and Evolution, 10*(21), 12307-12317.

Chappell, E. M., Trewick, S. A., & Morgan-Richards, M. (2012). Shape and sound reveal genetic cohesion not speciation in the New Zealand orthopteran, *Hemiandrus pallitarsis*, despite high mitochondrial DNA divergence. *Biological Journal of the Linnean Society, 105*, 169-186.

Chefaoui, R. M., & Lobo, J. M. (2008). Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling, 210*(4), 478-486.

Chen, I.-C., Hill, J. K., Ohlemuller, R., Roy, D. B., & Thomas, C. D. (2011). Rapid Range Shifts of Species Associated with High Levels of Climate Warming. *Science, 333*(6045), 1024-1026.

Cheng, J., Sun, J., Yao, K., Xu, M., & Cao, Y. (2022). A variable selection method based on mutual information and variance inflation factor. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 268*, 120652.

Chinn, W. G., & Gemmell, N. J. (2004). Adaptive radiation within New Zealand endemic species of the cockroach genus *Celatoblatta* Johns (Blattidae): a response to Plio-Pleistocene mountain building and climate change. *Molecular Ecology, 13*(6), 1507-1518.

Chrisholm, W. P. (1984). *Aspects of the biology of Ichthybotus hudsoni (McLachlan) (Ephemeroptera: Ephemeridae)*. (Master of Science in Zoology). Victoria University of Wellington, Wellington, New Zealand.

Crimmins, S. M., Dobrowski, S. Z., & Mynsberge, A. R. (2013). Evaluating ensemble forecasts of plant species distributions under climate change. *Ecological Modelling, 266*, 126-130.

Crosby, T. K., Dugdale, J. S., & Watt, J. C. (1998). Area codes for recording specimen localities in the New Zealand subregion. *New Zealand Journal of Zoology, 25*(2), 175-183.

Cummins, K. W. (1973). Trophic relations of aquatic insects. *Annual Review of Entomology, 13*, 183-206.

Daufresne, M., & Boet, P. (2007). Climate change impacts on structure and diversity of fish communities in rivers. *Global Change Biology, 13*(12), 2467-2478.

Dawson, T. P., Jackson, S. T., House, J. I., Prentice, I. C., & Mace, G. M. (2011). Beyond Predictions: Biodiversity conservation in a changing climate. *Science, 332*(6025), 53-58.

Death, R. G., & Winterbourn, M. J. (1995). Diversity patterns in stream benthic invertebrate communities: The influence of habitat stability. *Ecology, 76*(5), 1446-1460.

Death, R. G. (2002). Predicting invertebrate diversity from disturbance regimes in forest streams. *Oikos, 97*, 18-30.

Death, R. G. (2003). Spatial patters in lotic invertebrate community composition: is substrate disturbance actually important? *Canadian Journal of Fisheries and Aquatic Sciences, 50*(5), 603-611.

Demoulin, G. (1957). Remarques critiques sur la position systématique des *Ichthybotus* Eaton, Éphéméroptères de Nouvelle-Zélande. *Bulletin et Annales de la Société Royale Entomologique de Belgique, 93*, 335-337.

Diniz-Filho, J. A. F., Mauricio Bini, L., Fernando Rangel, T., Loyola, R. D., Hof, C., Nogués-Bravo, D., & Araújo, M. B. (2009). Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography, 32*(6), 897-906.

Diniz-Filho, J. A. F., Soares, T. N., Lima, J. S., Dobrovolski, R., Landeiro, V. L., Telles, M. P. C., . . . Bini, L. M. (2013). Mantel test in population genetics. *Genetics and Molecular Biology, 36*(4), 475-485.

Dodds, W. K., Gido, K., Whiles, M. R., Fritz, K. M., & Matthews, W. J. (2004). Life on the edge: the ecology of Great Plains prairie streams. *BioScience, 54*(3).

Domisch, S., Araujo, M. B., Bonada, N., Pauls, S. U., Jahnig, S. C., & Haase, P. (2013). Modelling distribution in European stream macroinvertebrates under future climates. *Global Change Biology, 19*(3), 752-762.

Dormann, C. F., Bobrowski, M., Dehling, D. M., Harris, D. J., Hartig, F., Lischke, H., . . . Kraan, C. (2018). Biotic interactions in species distribution modelling: 10 questions to guide interpretation and avoid false conclusions. *Global Ecology and Biogeography, 27*(9), 1004-1016.

Drinan, T. J., Grainger, N. P. J., Harding, J. S., Collier, K. J., Smith, B. J., Death, R. G., . . . Rolfe, J. R. (2020). Analysis of the conservation status of New Zealand freshwater invertebrates: temporal changes, knowledge gaps, impediments, and management implications. *New Zealand Journal of Zoology, 48*(1), 81-96.

Dumbleton, L. J. (1970). Presidential address: Pleistocene climates and insect distributions. *New Zealand Entomologist 4*(3), 3-23.

Durance, I., & Ormerod, S. J. (2009). Trends in water quality and discharge confound long-term warming effects on river macroinvertebrates. *Freshwater Biology, 54*(2), 388-405.

Eaton, A. E. (1899). An annotated list of the Ephemeridae of New Zealand. *Transactions of the Entomological Society of London, 47*, 285-293.

Efford, M., Howitt, R., & Gleeson, D. (2002). Phylogenetic relationships of *Wainuia* (Mollusca: Pulmonata)—biogeography and conservation implications. *Journal of the Royal Society of New Zealand, 32*(3), 445-456.

Elith, J., Graham, C. H., Anderson, R. P., Dudik, M., Ferrier, S., Guisan, A., . . . Zimmermann, N. E. (2006). Novel methods improve prediction of species distributions from occurrence data. *Ecography, 29*, 129-151.

Elith, J., & Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, *Evolution and Systematics, 40*, 677-697.

Ellis, E. A., Marshall, D. C., Hill, K. B. R., Owen, C. L., Kamp, P. J. J., & Simon, C. (2015). Phylogeography of six codistributed New Zealand cicadas and their relationship to multiple biogeographical boundaries suggest a re-evaluation of the Taupo Line. *Journal of Biogeography, 42*(9), 1761-1775.

Ersts, P. J. (2006). Geographic Distance Matrix Generator (Version 1.2.3). American Museum of Natural History: Center for Biodiversity and Conservation. Retrieved from http://biodiversityinformatics.amnh.org/open_source/gdmg

Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic Consequences of Range Expansions. Annual Review of Ecology, *Evolution, and Systematics, 40*(1), 481-501.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861-874.

Fletcher, R. J., McCleery, R. A., Greene, D. U., & Tye, C. A. (2016). Integrated models that unite local and regional data reveal larger-scale environmental relationships and improve predictions of species distributions. *Landscape Ecology, 31*(6), 1369-1382.

Folmer, O., Black, M., Hoeh, W., Lutz, R., & 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.

Frankham, R. (1996). Relationship of genetic variation to population size in wildlife. *Conservation Biology, 10*(6), 1500-1508.

Franklin, J. (2010). Moving beyond static species distribution models in support of conservation biogeography. *Diversity and Distributions, 16*(3), 321-330.

Garcia, R. A., Cabeza, M., Rahbek, C., & Araujo, M. B. (2014). Multiple dimensions of climate change and their implications for biodiversity. *Science, 344*(6183), 1247579.

Gattolliat, J. L. (2004). First reports of the genus *Nigrobaetis Novikova* and *Kluge* (Ephemeroptera: Baetidae) from Madagascar and La Reunion with observations on afrotropical biogeography. *Revue suisse de Zoologie, 111*(3), 657-669.

Girden, E. R. (1992). *ANOVA: Repeated measures.* Newbury Park, CA: Sage Publications, Inc.

Greaves, S. N., Chapple, D. G., Gleeson, D. M., Daugherty, C. H., & Ritchie, P. A. (2007). Phylogeography of the spotted skink (*Oligosoma lineoocellatum*) and green skink (*O. chloronoton*) species complex (Lacertilia: Scincidae) in New Zealand reveals pre-Pleistocene divergence. *Molecular Phylogenetics and Evolution, 45*(2), 729-739.

Grinnell, J. (1917a). Field tests of theories concerning distributional control. *The American Naturalist, 51*(602), 115-128.

Grinnell, J. (1917b). The niche-relationships of the California thrasher. *The Auk, 34*(4), 427-433.

Guillera-Arroita, G., Lahoz-Monfort, J. J., Rooyen, A. R., Weeks, A. R., Tingley, R., & McCrea, R. (2017). Dealing with false-positive and false-negative errors about species occurrence at multiple levels. *Methods in Ecology and Evolution, 8*(9), 1081-1091.

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J. M. C., Aspinall, R., & Hastie, T. (2006). Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology, 43*(3), 386-392.

Hao, T., Elith, J., Guillera-Arroita, G., Lahoz-Monfort, J. J., & Serra-Diaz, J. (2018). A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. *Diversity and Distributions, 25*(5), 839-852.

Hebert, P. D., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings: Biological Sciences, 270*(1512), 313-321.

Heikkinen, R. K., Luoto, M., Araujo, M. B., Virkkala, R., Thuiller, W., & Sykes, M. T. (2006). Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography, 30*, 751-777.

Hernandez, P. A., Graham, C. H., Master, L. L., & Albert, D. L. (2006). The effect of sample size and species characteristics on performance of different species. *Ecography, 29*, 773-785.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology, 25*(15), 1965-1978.

Hitchings, T. R. (2008). The post glacial distribution of New Zealand Mayflies. In F. R. Hauer, J. A. Stanford, & R. L. Newell (Eds.), *International Advances in the Ecology, Zoogeography, and Systematics of Mayflies and Stoneflies* (Vol. 128, pp. 89-101). Berkley and Los Angeles, California: University of California Press.

Hopkins, C. L. (1976). Estimate of biological production in some stream invertebrates. *New Zealand Journal of Marine and Freshwater Research, 10*, 629-640.

Hudson, G.V. (1904). *New Zealand Neuroptera: a popular introduction to the life histories and habits of mayflies, dragonflies and allied insects inhabiting New Zealand, including notes on their relation to angling.* London, West Newman & Co.

Hutchinson, G. E. (1957). Concluding Remarks. Population Studies: *Animal Ecology and Demography. Cold Springs Harbor Symposia on Quantitative Biology, 22*, 415-427.

Jackson, S. T., & Overpeck, J. T. (2000). Responses of plant populations and communities to environmental changes of the late Quaternary. *Paleobiology, 26*(4) 194-220.

Jacobus, L. M., Macadam, C. R., & Sartori, M. (2019). Mayflies (Ephemeroptera) and Their Contributions to Ecosystem Services. *Insects, 10*(6).

Jähnig, S. C., Kuemmerlen, M., Kiesel, J., Domisch, S., Cai, Q., Schmalz, B., . . . Peter, L. (2012). Modelling of riverine ecosystems by integrating models: conceptual approach, a case study and research agenda. *Journal of Biogeography, 39*(12), 2253-2263.

Jiménez-Valverde, A., Lobo, J. M., & Hortal, J. (2008). Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions, 14*(6), 885-890.

Kaneko, H., Ishiwata, S. i., Bae, Y. J., & Takamura-Enya, T. (2021). Genetic characteristics and phylogeography of the habitat generalist mayfly *Ecdyonurus yoshidae* (Ephemeroptera: Heptageniidae) in the Japanese archipelago. *Entomological Research, 51*(5), 238-250.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., . . . Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics, 28*(12), 1647-1649.

Koot, E. M., Morgan-Richards, M., & Trewick, S. A. (2022). Climate change and alpine-adapted insects: modelling environmental envelopes of a grasshopper radiation. *Royal Society Open Science, 9*(3), 211596.

Kumar, R., & Kumar, V. (2018). A review of phylogeography: biotic and abiotic factors. *Geology, Ecology, and Landscapes, 2*(4), 268-274.

Kylafis, G., & Loreau, M. (2011). Niche construction in the light of niche theory. *Ecology Letters, 14*(2), 82-90.

Lavergne, S., Mouquet, N., Thuiller, W., & Ronce, O. (2010). Biodiversity and Climate Change: Integrating Evolutionary and Ecological Responses of Species and Communities. *Annual Review of Ecology, Evolution, and Systematics, 41*(1), 321-350.

Leigh, J. W., Bryant, D., & Nakagawa, S. (2015). popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution, 6*(9), 1110-1116.

Lepori, F., & Hjerdt, N. (2006). Disturbance and aquatic biodiversity: reconciling contrasting views. *BioScience, 56*(10), 809-818.

Mantel, N. (1967). The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Research, 27*, 209-220.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., & Thuiller, W. (2009). Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions, 15*(1), 59-69.

Marshall, D. C., Hill, K. B., Cooley, J. R., & Simon, C. (2011). Hybridization, mitochondrial DNA phylogeography, and prediction of the early stages of reproductive isolation: lessons from New Zealand cicadas (genus *Kikihia*). *Systematic Biology, 60*(4), 482-502.

Marshall, D. C., Hill, K. B., Fontaine, K. M., Buckley, T. R., & Simon, C. (2009). Glacial refugia in a maritime temperate climate: cicada (*Kikihia subalpina*) mtDNA phylogeography in New Zealand. *Molecular Ecology, 18*(9), 1995-2009.

Marshall, D. C., Slon, K., Cooley, J. R., Hill, K. B., & Simon, C. (2008). Steady Plio-Pleistocene diversification and a 2-million-year sympatry threshold in a New Zealand cicada radiation. *Molecular Phylogenetics and Evolution, 48*(3), 1054-1066.

Marske, K. A., Leschen, R. A., & Buckley, T. R. (2011). Reconciling phylogeography and ecological niche models for New Zealand beetles: Looking beyond glacial refugia. *Molecular Phylogenetics and Evolution, 59*(1), 89-102.

McCafferty, W. P. (1973). Commentary on the genus *Ichthybotus* (Ephemeridae) and the misplaced I. *Dodecus* Dubey. *Oriental Insects, 7*(3), 351-352.

McCulloch, G. A. (2010). *Evolutionary genetics of southern stoneflies.* (Doctor of Philosophy). University of Otago.

McLachlan, R. (1894). Some additions to the neuropterous fauna of New Zealand, with notes on certain described species. *Entomologist's Monthly Magazine, 30*, 270-272.

McPherson, J. M., Jetz, W., & Reogers, D. J. (2004). The effects of species range sizes on the accuracy of distribution models. *Journal of Applied Ecology, 41*, 811-823.

Moilanen, A., Leathwick, J., & Elith, J. (2008). A method for spatial freshwater conservation prioritization. *Freshwater Biology, 53*(3), 577-592.

Monaghan, M. T., Gattolliat, J. L., Sartori, M., Elouard, J.-M., James, H., Derleth, P., . . . de Moor, F. C. (2005). Trans-Oceanic and endemic origins of the small minnow mayflies (Ephemeroptera, Baetidae) of Madagascar. *Proceedings: Biological Sciences, 272*(1574), 1829-1836.

Monaghan, M. T., & Sartori, M. (2009). Genetic contributions to the study of taxonomy, ecology, and evolution of mayflies (Ephemeroptera): review and future perspectives. *Aquatic Insects, 31*, 19-39.

Morgan-Richards, M., Bulgarella, M., Sivyer, L., Dowle, E. J., Hale, M., McKean, N. E., & Trewick, S. A. (2017). Explaining large mitochondrial sequence differences within a population sample. *Royal Society Open Science, 4*(11), 170730.

Morgan-Richards, M., Langton-Myers, S. S., & Trewick, S. A. (2019). Loss and gain of sexual reproduction in the same stick insect. *Molecular Ecology, 28*(17), 3929-3941.

Morgan-Richards, M., Trewick, S. A., & Stringer, I. A. (2010). Geographic parthenogenesis and the common tea-tree stick insect of New Zealand. *Molecular Ecology, 19*(6), 1227-1238.

Morin, X., & Thuiller, W. (2009). Comparing niche- and process-based models to reduce prediction uncertainty in species range shifts under climate change. *Ecology, 90*(5), 1301-1313.

Newnham, R., McGlone, M., Moar, N., Wilmshurst, J., & Vandergoes, M. (2013). The vegetation cover of New Zealand at the Last Glacial Maximum. *Quaternary Science Reviews, 74*, 202-214.

Ogden, T. H., & Whiting, M. F. (2005). Phylogeny of Ephemeroptera (mayflies) based on molecular evidence. *Molecular Phylogenetics and Evolution, 37*(3), 625-643.

Olding-Smee, F. J., Laland, K. N., & Feldman, M. W. (2003). Chapter 1 Introduction. In *The Niche Contruction: The Neglected Process in Evolution* (MPB-37).

Ormerod, S. J., Dobson, M., Hildrew, A. G., & Townsend, C. R. (2010). Multiple stressors in freshwater ecosystems. *Freshwater Biology, 55*, 1-4.

Papadopoulou, A., Anastasiou, I., & Vogler, A. P. (2010). Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Molecular Biology and Evolution, 27*(7), 1659-1672.

Pearman, P. B., Guisan, A., Broennimann, O., & Randin, C. F. (2007). Niche dynamics in space and time. *Trends in Ecology & Evolution, 23*(3).

Pearson, R. G., & Dawson, T. P. (2003). Predicting the impacts of climate change on the distribution of species. *Global Ecology & Biogeography, 12*, 361-371.

Pearson, R. G., Thuiller, W., Araújo, M. B., Martinez-Meyer, E., Brotons, L., McClean, C., . . . Lees, D. C. (2006). Model-based uncertainty in species range prediction. *Journal of Biogeography, 33*(10), 1704-1711.

Petersen, I., Masters, Z., Hildrew, A. G., & Ormerod, S. J. (2004). Dispersal of adult aquatic insects in catchments of differing land use. *Journal of Applied Ecology, 41*, 934-950.

Phillips, S. J. (1930). A revision of New Zealand Ephemeroptera. *Transactions and Proceedings of the New Zealand Institute, 61*, 271-390.

Pohe, S. R. (2018). An annotated checklist of New Zealand mayflies. *New Zealand Natural Sciences, 43*, 1-20.

Pohe, S. R. (2019). *Macroecology of New Zealand Ephemeroptera*. (Doctor of Philosophy in Ecology). University of Canterbury.

Pohe, S. R., Winterbourn, M. J., & Harding, J. S. (2017). Comparison of fluorescent lights with differing spectral properties on catches of adult aquatic and terrestrial insects. *New Zealand Entomologist, 41*(1), 1-11.

Pulliam, R. H. (2000). On the relationship between niche and distribution. *Ecology Letters, 3*, 349-361.

Resh, V. H., Brown, A. V., Covich, A. P., Gurtz, M. E., Li, H. W., Minshall, W., . . . Wissmar, R. C. (1988). The role of disturbance in stream ecology. *Journal of the North American Benthological Society, 7*(4), 433-555.

Rozas, J. (2009). DNA sequence polymorphism analysis using DnaSP. *Methods in Molecular Biology, 537*, 337-350.

Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sanchez-Gracia, A. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Molecular Biology and Evolution, 34*(12), 3299-3302.

Sales, L. P., Hayward, M. W., & Loyola, R. (2021). What do you mean by niche? Modern ecological theories are not coherent on rhetoric about the niche concept. *Acta Oecologica, 110*, 1-9.

Sartori, M., & Brittain, J. E. (2015). Chapter 23 - Order Ephemeroptera. In J. H. Thorp & D. C. Rogers (Eds.), *Thorp and Covich's Freshwater Invertebrates* (4 ed., pp. 873-891): Academic Press.

Segurado, P., & Araujo, M. B. (2004). An evaluation of methods for modelling species distributions. *Journal of Biogeography, 31*, 1555-1568.

Sillero, N., & Barbosa, A. M. (2021). Common mistakes in ecological niche models. *International Journal of Geographical Information Science, 35*(2), 213-226.

Silvertown, J. (2004). Plant coexistence and the niche. *Trends in Ecology & Evolution, 19*(11), 605-611.

Stevens, R. D. (2021). Reflections of Grinnellian and Eltonian niches on the distribution of phyllostomid bats in the Atlantic Forest. *Journal of Biogeography, 49*(1), 94-103.

Stockwell, D. R. B., & Peterson, A. T. (2002). Effects of sample size on accuracy of species distribution models. *Ecological Modelling, 148*, 1-13.

Suggate, R. P., & Almond, P. C. (2005). The Last Glacial Maximum (LGM) in western South Island, New Zealand: implications for the global LGM and MIS 2. *Quaternary Science Reviews, 24*, 1923-1940.

Syphard, A. D., & Franklin, J. (2010). Species traits affect the performance of species distribution models for plants in southern California. *Journal of Vegetation Science, 21*(1), 177-189.

Taubmann, J., Theissinger, K., Feldheim, K. A., Laube, I., Graf, W., Haase, P., . . . Pauls, S. U. (2010). Modelling range shifts and assessing genetic diversity distribution of the montane aquatic mayfly *Ameletus inopinatus* in Europe under climate change scenarios. *Conservation Genetics, 12*(2), 503-515.

Taylor-Smith, B., Morgan-Richards, M., & Trewick, S. A. (2019). Patterns of regional endemism among New Zealand invertebrates. *New Zealand Journal of Zoology, 47*(1), 1-19.

Team, R. C. (2021). R: A language and environment for statistical computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Thuiller, W. (2003). BIOMOD optimizing predictions of species distributions and projecting potential. *Global Change Biology, 9*, 1353-1362.

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography, 32*(3), 369-373.

Tillyard, R. J. (1923). Descriptions of two new species of may-flies (order Plectoptera) from New Zealand. *Transactions and Proceedings of the New Zealand Institute, 54*(226-230).

Trewick, S. A., & Bland, K. J. (2012). Fire and slice: palaeogeography for biogeography at New Zealand's North Island/South Island juncture. *Journal of the Royal Society of New Zealand, 42*(3), 153-183.

Trewick, S. A., Henderson, I. M., Pohe, S. R., & Morgan-Richards, M. (2022). Spatial variation of *Acanthophlebia cruentata* (Ephemeroptera), a mayfly endemic to Te Ika-a-Maui-North Island of Aotearoa, New Zealand. *Insects, 13*(7).

Trewick, S. A., Pilkington, S., Shepherd, L. D., Gibb, G. C., & Morgan-Richards, M. (2017). Closing the gap: Avian lineage splits at a young, narrow seaway imply a protracted history of mixed population response. *Molecular Ecology, 26*(20), 5752-5772.

Trewick, S. A., Wallis, G. P., & Morgan-Richards, M. (2011). The invertebrate life of New Zealand: A phylogeographic approach. *Insects, 2*(3), 297-325.

Valencia-Rodriguez, D., Jimenez-Segura, L., Rogeliz, C. A., & Parra, J. L. (2021). Ecological niche modeling as an effective tool to predict the distribution of freshwater organisms: The case of the Sabaleta *Brycon henni* (Eigenmann, 1913). *PLoS One, 16*(3), 1-17.

Van Strien, M. J., Holderegger, R., & Van Heck, H. J. (2015). Isolation-by-distance in landscapes: considerations for landscape genetics. *Heredity, 114*(1), 27-37.

Wallis, G. P., & Trewick, S. A. (2009). New Zealand phylogeography: evolution on a small continent. *Molecular Ecology, 18*(17), 3548-3580.

Ward, R. D. (2009). DNA barcode divergence among species and genera of birds and fishes. *Molecular Ecology Resources, 9*(4), 1077-1085.

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., & Guisan, A. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions, 14*(5), 763-773.

Ye, Z., Zhu, G., Chen, P., Zhang, D., & Bu, W. (2014). Molecular data and ecological niche modelling reveal the Pleistocene history of a semi-aquatic bug (*Microvelia douglasi douglasi*) in East Asia. *Molecular Ecology, 23*(12), 3080-3096.

Zhao, M., Wang, C. Y., Sun, L., He, Z., Yang, P. L., Liao, H. J., & Feng, Y. (2021). Edible Aquatic Insects: Diversities, Nutrition, and Safety. *Foods, 10*(12).

Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., . . . Merow, C. (2020). A standard protocol for reporting species distribution models. *Ecography, 43*(9), 1261-1277.