

Genogeographic clustering to identify cross-species concordance of spatial genetic patterns

Vanessa Arranz^{1,2}  | Rachel M. Fewster³ | Shane D. Lavery^{1,2}

¹School of Biological Sciences, University of Auckland, Auckland, New Zealand

²Institute of Marine Sciences, University of Auckland, Auckland, New Zealand

³Department of Statistics, University of Auckland, Auckland, New Zealand

Correspondence

Vanessa Arranz, School of Natural and Computational Sciences, Massey University Auckland, Albany, Auckland 0745, New Zealand.
Email: vanearranz@hotmail.com

Funding information

Beate Schuler Doctoral Scholarship

Editor: Chris Burridge

Abstract

Aim: While in recent years, there have been considerable advances in discerning spatial genetic patterns within species, the task of identifying common patterns across species is still challenging. Approaches using new data from co-sampled species permit rigorous statistical analysis but are often limited to a small number of species. Meta-analyses of published data can encompass a much broader range of species, but are usually restricted by uneven data properties. There is a need for new approaches that bring greater statistical rigour to meta-analyses and are also able to discern more than a single spatial pattern among species. We propose a new approach for comparative multi-species meta-analyses of published population genetic data that address many existing limitations.

Innovation: The proposed “genogeographic clustering” technique takes a three-stage approach: (i) use common genetic metrics to gain location-specific measures across the sampled range of each species; (ii) for each species, determine the spatial genetic pattern by fitting a smooth “genogeographic” trend curve to the genetic data; and (iii) quantitatively cluster species according to their similarity in spatial pattern. We apply this technique to 21 species of intertidal invertebrates from the New Zealand coastline, to resolve common spatial patterns from disparate profiles of genetic diversity.

Main conclusions: The genogeographic curves are shown to successfully capture the known spatial patterns within each intertidal species and readily permit statistical comparison of those patterns, regardless of sampling and marker inconsistencies. The species clustering technique is shown to discern groups of species that clearly share spatial patterns within groups but differ significantly among groups. Genogeographic species clustering provides a novel approach to discerning multiple common spatial patterns of diversity among a large number of species. It will permit more rigorous comparative studies from diverse published data and can be easily extended to a wide variety of alternative measures of genetic diversity or divergence.

KEYWORDS

clustering, comparative phylogeography, genetic connectivity, haplotype diversity, multi-species, spatial patterns

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Diversity and Distributions* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

It is a considerable challenge in conservation biology to understand the relationships among genetic, ecological and environmental variables (Selkoe et al., 2014; Taberlet et al., 2012). One relevant genetic measure is genetic diversity, which is a fundamental component of biodiversity and has crucial implications for fitness, productivity and resilience of species, and ultimately for ecosystem functionality (Aguirre & Marshall, 2012; Bernhardt & Leslie, 2013; Hughes et al., 2008). Patterns of genetic diversity within populations can provide insights about historical processes such as demographic change (Garrick et al., 2010; Nason et al., 2002). Information about genetic diversity plays an important role in conservation management (von der Heyden et al., 2014), and there is an increasing focus on spatial patterns of diversity in large-scale conservation initiatives (Beger et al., 2014; Coates et al., 2018).

For insight into genetic processes at the whole ecosystem level, a multi-species approach is required (Dawson, 2014; Hickerson & Meyer, 2008). Comparative phylogeography describes the study of multiple co-distributed species, often to identify common spatial genetic patterns (Dawson, 2014; Liggins et al., 2015; Moritz & Faith, 1998; Rocha et al., 2007). Although many such studies are focused on genetic divergence (Marko & Hart, 2011), there is also an increasing drive to understand patterns in genetic diversity (Lamy et al., 2017; Vellend et al., 2014). Within each species, genetic diversity is influenced by numerous mechanisms including mutation, migration, genetic drift, population demography and natural selection (Hewitt, 2004; Hughes et al., 2008; Vucetich & Waite, 2003). Therefore, different species will naturally exhibit different spatial genetic patterns due to their undoubtedly different histories. However, each species serves as a sample replicate of a system in which genetic diversity is shaped by the action of common mechanisms. Multi-species studies may provide insight into these common mechanisms by illuminating spatial patterns of genetic diversity that are shared across taxa (Dawson, 2014; Liggins et al., 2016; Toonen et al., 2011).

Although there are various methods available for analysing spatial genetic patterns within species, identifying a suite of key patterns underpinning a community of species remains challenging (Liggins et al., 2016). Riginos et al. (2011) outlined two main types of quantitative approaches to finding concordant patterns. The first is to collect congruent data by co-sampling multiple species at shared locations (Dawson et al., 2014; Hickerson et al., 2007; Liggins et al., 2016; Villamor et al., 2014), while the second is to perform a meta-analysis of previously published studies (Pelc et al., 2009; Riginos et al., 2011). Studies using the first approach are limited by the considerable logistics and time required to undertake extensive co-sampling and usually based on relatively few species. However, because they share uniform data properties, they are readily amenable to statistical analysis such as hypothesis testing (Liggins et al., 2016). Additionally, where studies of co-sampled species can incorporate temporal information about divergence events and identify pseudocongruence, they present an opportunity to understand factors driving patterns of genetic differentiation (Dawson, 2014).

By contrast, studies based on previously published analyses tend to be severely restricted by different sampling locations for each species and often by not having access to the raw genetic data. These aspects greatly hinder investigations of the factors underlying geographic patterns. However, studies based on pre-existing data have the advantage of being able to include a much larger number of species (Riginos et al., 2011; Selkoe et al., 2014).

Recently, some studies have merged the two approaches identified by Riginos et al. (2011), for instance by using a meta-analysis of synchronously diverging codistributed pairs of taxa to discern factors driving patterns of population differentiation (Schiebelhut & Dawson, 2018). Despite this, there are still clear benefits in taking advantage of the wealth of previously published studies in order to maximize the breadth of species analysed. It seems clear that in complex ecosystems such as marine assemblages, assessing a broad spectrum of species offers us the greatest chance of understanding community assembly and functioning. By examining the patterns of population genetic variation over a broad range of species, we can build a true overview of the most important factors responsible for community structure.

Here, we present and explore a new approach for meta-analysis of spatially explicit genetic patterns across multiple species that addresses some of the limitations of existing approaches. Difficulties encountered in previous meta-analyses include the following: (1) differences in sampling location and intensity, (2) differences in genetic markers used, (3) inaccessibility of the raw genetic data, (4) a limited number of methods to objectively compare spatial genetic patterns among species, and (5) no existing method to quantitatively group species by their spatially explicit genetic patterns. To address these problems, we present a method of capturing the spatial pattern for each species that is indifferent to precise sampling locations, enabling us to identify species whose spatial genetic patterns are similar regardless of differences in sampling location and intensity.

The analysis follows a three-stage approach: (1) compile location-specific measures of a genetic quantity of interest, such as diversity, for each species across the geographic range of the study, (2) fit a smooth spatial trend curve to the genetic data for each species, to enable comparisons between the spatial patterns of all species regardless of their individual sampling regimes, and (3) quantitatively cluster species according to their similarity in spatial pattern. As it is the shape of the genetic geographic patterns that are being compared, and not the measured diversity values themselves, this approach can integrate results from contributory studies using different genetic metrics to measure diversity. Because this encompasses possibilities beyond the phylogenetic relationships of DNA haplotypes, we term these “genogeographic” rather than “phylogeographic” patterns.

Our ultimate aim is to create a quantitative method to capture spatial genetic patterns using data from previously published studies, notwithstanding differences in sampling location and intensity, and to cluster species according to similarities in their corresponding genogeographic profiles. In this way, it becomes possible to evaluate the variables that unite species with similar patterns, gaining

insight into both the intrinsic (biological) and extrinsic (environmental) forces potentially governing the entire community. Our goal here is to find and compare genetic patterns and not investigate whether genetic indices give a good description of evolutionary processes. As our approach cannot tease apart the complex genetic and demographic forces driving the observed patterns, we view it as an exploratory method for investigating multi-species communities, suitable for generating hypotheses about the processes shaping biodiversity, which can then be addressed by more targeted analyses. We illustrate the approach using an example of within-population genetic data on allelic diversity, in a one-dimensional scenario corresponding to the intertidal community around the coastline of New Zealand. Such one-dimensional scenarios have been common in previous studies of marine comparative phylogeography (Cahill et al., 2017; Crandall et al., 2019; Matias & Riginos, 2018; Mertens et al., 2018; Nielsen et al., 2017; Stanley et al., 2018). Further developments, such as exploring genetic divergence rather than diversity, and extension to two-dimensional habitats, are considered in the Section 4.

2 | METHODS

2.1 | Focal species dataset

For an adequate test of our approach, we sought a dataset from species of one community type, with as broad taxonomic range as possible and with each species relatively well-sampled throughout the geographic range considered. We assembled a dataset of benthic intertidal marine species around the New Zealand coast. The study area comprised the North, South and Stewart Islands. The species of interest were endemic invertebrates inhabiting the intertidal and high subtidal zones, which have been relatively well examined previously. Measures of genetic diversity at each sampling location were obtained for each species from previously published studies and studies underway in our laboratory. Studies were included if they met a set of inclusion criteria based on Pelc et al. (2009) (see Appendix S1). Twenty-one species were included comprising all the main invertebrate phyla found along the New Zealand coast: Cnidaria, Arthropoda, Echinodermata and Mollusca (Table 1).

2.2 | Genetic measures

For each species, a genetic diversity measure was computed within each geographic location. The measure of genetic diversity used depended on the genetic marker available for each species. Haplotype diversity " H " was used for mitochondrial DNA, and the analogous allelic expected heterozygosity " H_e " for nuclear DNA microsatellites. Collectively, these values are henceforth referred to simply as allelic diversity, " H ". Values were either reported directly in the publications used, or alternatively calculated using ARLEQUIN v 3.5 (Excoffier & Lischer, 2010) and GENEALEx (Peakall & Smouse, 2012) from the

genetic data associated with the publication. Adjustments for small sample sizes were included in the calculation of all parameters, using unbiased estimators (Nei, 1972). The assembled dataset is available from the Dryad Digital Repository (<https://doi.org/10.5061/dryad.b2rbnzsbr>).

2.3 | Linearized coastline

The New Zealand coastline was linearized using coastal distances between sampling points. Sampling points within 50 km were synonymized to the same location. Linearization enabled us to plot genetic patterns along one spatial dimension, namely distance along coast. The starting point at 0 km was Cape Reinga, the northernmost point of the country, and the distance coordinate followed the coastline in a clockwise direction around New Zealand, through to Stewart Island in the south, before returning to Cape Reinga (Figure 1). The linearized coastline crossed Cook Strait between the North and South Islands at its narrowest point. The north-south route at this point best captured the known predominant spatial genetic patterns in the region of Cook Strait, which are north-south divergences, rather than east-west divergences.

2.4 | Genogeographic species curves

For each species, the data points corresponding to the measured H at each sampled location were plotted against the appropriate coordinate on the linearized coastline, and a curve was fitted to these points to create the spatial pattern for that species. The fitted curves aimed to capture all major trends in the data points while being smooth enough to allow the curves for different species to be compared. To achieve this, we used cubic spline curves, which are piecewise cubic polynomials that can fit very flexible shapes while remaining smooth (Hastie & Tibshirani, 1990).

In the case of our example data set, which formed a closed loop, each curve was constrained to have the same height at the start and end, because both ends corresponded to Cape Reinga. This was achieved by cycling the data three times over the entire coastline length, and using only the curve fitted to the middle cycle, so that the curve height at each endpoint was correctly influenced by data points to each side of the Cape Reinga divide and then imposing a penalty to force the curve heights to match at the start and end of the middle cycle. Spline knots were equally spaced, and the middle cycle was estimated with $n/3$ knots, where n is the number of data points for the corresponding species. This number of knots was selected to give the curve sufficient flexibility to expose key patterns in the spatial genetic profile, while maintaining sufficient smoothness to respect the spatial autocorrelation in each profile. The impact of choosing Cape Reinga as the start point was investigated using three alternative start points spaced at 1000 km intervals.

If the coastline followed a linear geometry rather than being a closed loop, the triple-cycling of data and the penalty steps would

TABLE 1 New Zealand species and studies included in the meta-analysis. Studies were selected according to the inclusion criteria in Appendix S1

Taxon	Species	Code name	Habitat type	Marker type	Number of locations (number of individuals)	Reference
Cnidaria	<i>Actinia tenebrosa</i>	Aten	Low intertidal	Microsatellites	23 (405)	Veale (2007)
Arthropoda	<i>Petrolisthes elongatus</i>	Pelo	Low intertidal	mtDNA	16 (440)	Arranz et al. (2021)
Arthropoda	<i>Hemigrapsus sexdentatus</i>	Hsex	Low intertidal	mtDNA	11 (249)	Hinnendael (2008)
Echinodermata	<i>Evechinus chloroticus</i>	Echlo	Subtidal	Microsatellites	11 (311)	Nagel et al. (2015)
Echinodermata	<i>Amphipholis squamata</i>	Asqu	Low intertidal	mtDNA	16 (173)	Sponer and Roy (2002)
Echinodermata	<i>Patiriella regularis</i>	Preg	Low intertidal	mtDNA	22 (284)	Ayers and Waters (2005)
Mollusca	<i>Austrovenus stutchburyi</i>	Astu	Subtidal	mtDNA	27 (372)	Ross et al. (2012)
Mollusca	<i>Cellana ornata</i>	Corn	High intertidal	mtDNA	30 (302)	Goldstien et al. (2006)
Mollusca	<i>Cellana radians</i>	Crad	Low intertidal	mtDNA	31 (321)	Goldstien et al. (2006)
Mollusca	<i>Chiton glaucus sp A</i>	CglaspA	Low intertidal	mtDNA	7 (100)	Lavery, unpublished data
Mollusca	<i>Chiton glaucus sp C</i>	CglaspC	Low intertidal	mtDNA	9 (86)	Lavery, unpublished data
Mollusca	<i>Haliotis iris</i>	Hiris	Subtidal	mtDNA	24 (534)	Will et al. (2011)
Mollusca	<i>Lunella smaragda</i>	Lsma	Low intertidal	mtDNA	26 (727)	Arranz et al. (2021)
Mollusca	<i>Paphies australis</i>	Paus	Subtidal	Microsatellites	13 (674)	Hannan (2014)
Mollusca	<i>Paphies subtriangulata</i>	Psub	Subtidal	Microsatellites	9 (571)	Hannan (2014)
Mollusca	<i>Perna canaliculus</i>	Pcan	Intertidal	mtDNA	22 (572)	Apte and Gardner (2002)
Mollusca	<i>Pecten novaezelandiae</i>	Pnov	Subtidal	Microsatellites	12 (952)	Silva and Gardner (2015)
Mollusca	<i>Scutus breviculus</i>	Sbre	Low intertidal	mtDNA	17 (112)	Waters et al. (2007)
Mollusca	<i>Sypharochiton pelliserpentis</i>	Spel	High intertidal	mtDNA	28 (472)	Veale and Lavery (2011)
Mollusca	<i>Zeacumantus lutulentus</i>	Zlut	Low intertidal	mtDNA	16 (173)	Keeney et al. (2013)
Mollusca	<i>Zeacumantus subcarinatus</i>	Zsub	Low intertidal	mtDNA	26 (551)	Keeney et al. (2013)

be unnecessary, and $n/3$ would be the number of knots used for the whole curve. When dealing with a linear coastline, ideally curves should not be extrapolated beyond the sampled range for any species. This might restrict the spatial scope of the study to the sampled extent common to all species of interest.

A beta distribution was used for the scatter of the measured H observations about the curve. This distribution confines the curves and their confidence intervals between 0 and 1 to match the range of the genetic H measure and allows a dispersion parameter to be estimated for each species. The curves were fitted by maximum likelihood using custom code written in R (R Core Team, 2020), adapted from the curve-fitting procedures of Cheeseman et al. (2017) and Antunes et al. (2016).

The smooth fitted curves define the spatial trend in the chosen genetic measure for each species. These curves are here termed “genogeographic” curves, as they capture similar patterns to those detected in standard phylogeographic studies (Avice, 2000), but are not dependent on phylogenetic data. Here, the curves summarize the broad spatial pattern of genetic diversity within each species. However, in general, they are applicable to a wide range of genetic and geographic scenarios, as long as there are location-specific genetic measures available and the topography can be adequately represented by a linear spatial coordinate.

2.5 | Clustering the spatial genetic curves: genogeographic clustering

The fitted curves were scaled and centred before clustering across species, such that the mean of values along each scaled-and-centred curve was 0 and the standard deviation was 1. The scaling forced the same overall variability in all curves before clustering, to allow the clustering procedures to identify species that showed similarities in their geographic patterns rather than in their absolute genetic measures. This step was appropriate for our study objective of comparing spatial genetic patterns among species, irrespective of the degree of genetic diversity within each species, but could be omitted if comparisons of absolute genetic diversity were of primary interest. With the scale-and-centre step, the methodology will identify species that have similar locations for higher and lower genetic diversity relative to their own mean diversity. A bootstrap method described below is used to take account of the level of statistical support for each curve, to ensure the scaling step does not create spurious patterns. Without the scaling step, species will cluster according to the absolute level of their genetic diversity and the extent and locations of geographical differences. Without scaling, comparisons should therefore be restricted to species for which a common genetic marker is available.

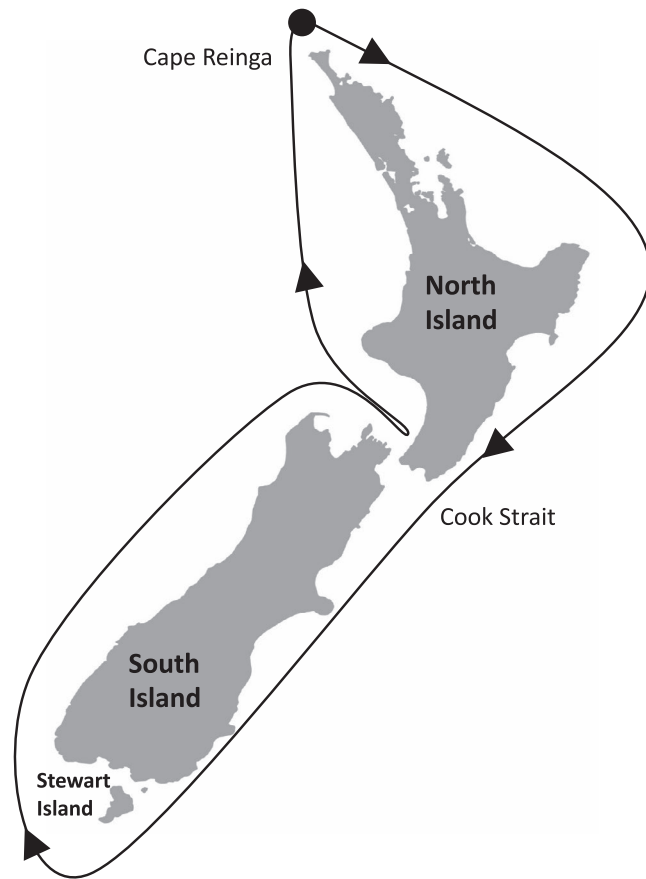


FIGURE 1 New Zealand map showing the linearized coastline concept. The black dot indicates the start and end of the linearized coastal profile. Arrows indicate the direction of the linearized coordinate

Following the scale-and-centre step, we used a cluster dendrogram to cluster species with similar genogeographic curves. We took the pairwise distance between two curves to be the Euclidean distance between the heights of the curves evaluated at 1000 equally spaced points along the linearized coastline. We chose 1000 points as a number that was well over ten times the maximum number of knots per curve and checked that results were consistent between alternative choices ranging from 500 to 2000.

In the dendrogram chart, species with the most similar genogeographic curves, corresponding to the smallest pairwise distances, are assigned to the first clusters at the bottom of the chart. Those species that are least similar only link together in the final, largest clusters that connect at the top of the chart. For our dataset of 21 species, the dendrogram contains a total of 20 clusters. The dendrogram was obtained using R function “hclust” with method “ward.D2” to decide the order of clusters. This method chooses the link at each step that minimizes the total within-cluster variance and can be thought of as the clustering equivalent of principal components analysis (Murtagh & Legendre, 2014).

A colour map was also generated for each species, to display the species-specific genogeographic pattern for the diversity measure. Low values of the scaled-and-centred curves were depicted in

yellow, and high values in red. Colour maps were also generated for each of the clusters obtained, using the average of the scaled-and-centred curves within the cluster.

2.6 | Measuring significance of genogeographic species clusters: parametric bootstrap

The data matrix we used to generate the dendrogram was not of the traditional species-by-site configuration, so to assess the significance of the clusters we used a parametric bootstrap. Each bootstrap replicate consisted of new data drawn from the fitted curves at the same linearized coastline coordinates as the real data, created by generating random draws from the fitted beta distributions underlying the trend curves. The percentile intervals shown on the trend curves mark the precision with which the trend was estimated for each species, based on the estimated dispersion parameter for that species, and they indicate the interval in which 95% of the randomly drawn bootstrap points will fall at any location. Having drawn a set of new data points for every species, the genogeographic curves were re-fitted to the new random data. For species with high scatter of the data points about the curve in the real data, the shape of the fitted curve changed substantially across bootstrap replicates, whereas for species with low dispersion, the curve tended to repeat the same shape for each random dataset. This procedure enabled us to allow for uneven sampling intensity or precision among species in our data. From each set of bootstrap replicate data across the 21 species, a new clustering of the replicate trend curves was undertaken, and a new dendrogram constructed. The whole process was repeated for 1000 bootstrap replicates to obtain 1000 different dendrograms.

To determine significant joins or splits, each pair of species was examined. For each of the 1000 dendrograms, if two species joined together before the final link, they were considered to belong to the same major cluster, or the same “half” of the dendrogram for that bootstrap replicate, whereas if they did not join until the final link, they were considered to have appeared in opposite halves of the dendrogram for that replicate. The two species were considered significantly joined at the 5% level if they appeared in the same half of the dendrogram for at least 95% of replicates, and they were considered significantly split at the 5% level if they appeared in opposite halves of the dendrogram for at least 95% of replicates. The significance of pairwise joins or splits therefore provided a measure of the robustness of these joins or splits to the variability encountered in the bootstrap procedure, which reflected the different levels of uncertainty for different species due to uneven sampling intensities and differences in the intrinsic spatial variability of the genetic measure.

2.7 | Principal coordinates analysis (PCO) for genogeographic curves

We used a principal coordinates analysis (PCO) to display species according to similarity of their scaled-and-centred genogeographic

curves and visualize species with significant joins or splits. PCO takes the matrix of distances between the genogeographic curves for each pair of species, as described above, and depicts these distances on a two-dimensional projection plot. We plotted lines between pairs of species that were significantly joined or split according to the bootstrap analysis.

3 | RESULTS

3.1 | Spatial genetic patterns from genogeographic curves and maps

The fitted genogeographic curves for the 21 species listed in Table 1 are plotted in Figure 2 and Figure S1. The plots show how comparable curves can be obtained despite different sampling locations for each species. Colour maps for the scaled-and-centred curves are shown in Figure 2 and Figure S2, highlighting the regions of elevated or reduced diversity.

Examination of the individual curves revealed a variety of spatial patterns among species (Figure 2). For example, the curve for *Hemigrapsus sexdentatus* showed higher H values in most North Island populations (far left and right of the curve) and lower H values in most South Island populations (centre of the curve). This geographic pattern is more clearly shown in the corresponding colour map (Figure 2). For species with a large number of sampling points, the fitted curves captured the main geographic trends without overfitting the spatial pattern, for example, *Sypharochiton pelliserpentis* and *Perna canaliculus* (Figure 2). Examination of the full suite of genogeographic curves (Figures S1 and S2) revealed

that there were several common patterns evident within subgroups of species.

3.2 | Similarity among species curves: genogeographic species clustering

Figure 3a shows the cluster dendrogram based on the 21 genogeographic H curves. The scaled-and-centred curves contributing to each cluster are shown in Figure 3b and Figure S3, showcasing the similarity of curves within clusters. For example, Cluster 11 contains five genogeographic curves with similar trends, in which the analysis detected a group of species with a common pattern of lower genetic diversity in southern populations compared to northern populations. This spatial pattern is more readily observed in the average heat map for Cluster 11 (Figure 3c). Species patterns within each cluster remained relatively similar for Clusters 1 to 17, while clusters beyond this combined some relatively different genogeographic curves. Clusters in which most links were identified as significant using the bootstrap approach are indicated by asterisks on the dendrograms, using a 10% significance level to allow for the stringency of the test where clusters contain species with few sampling points.

3.3 | Cluster membership and significance

Based on the cluster dendrogram in Figure 3a, the 21 species were mainly clustered into three well-defined groups: Clusters 15, 17, and 18. Cluster 17 grouped species with overall highest genetic diversity in the North Island, whereas Cluster 18 grouped species with highest

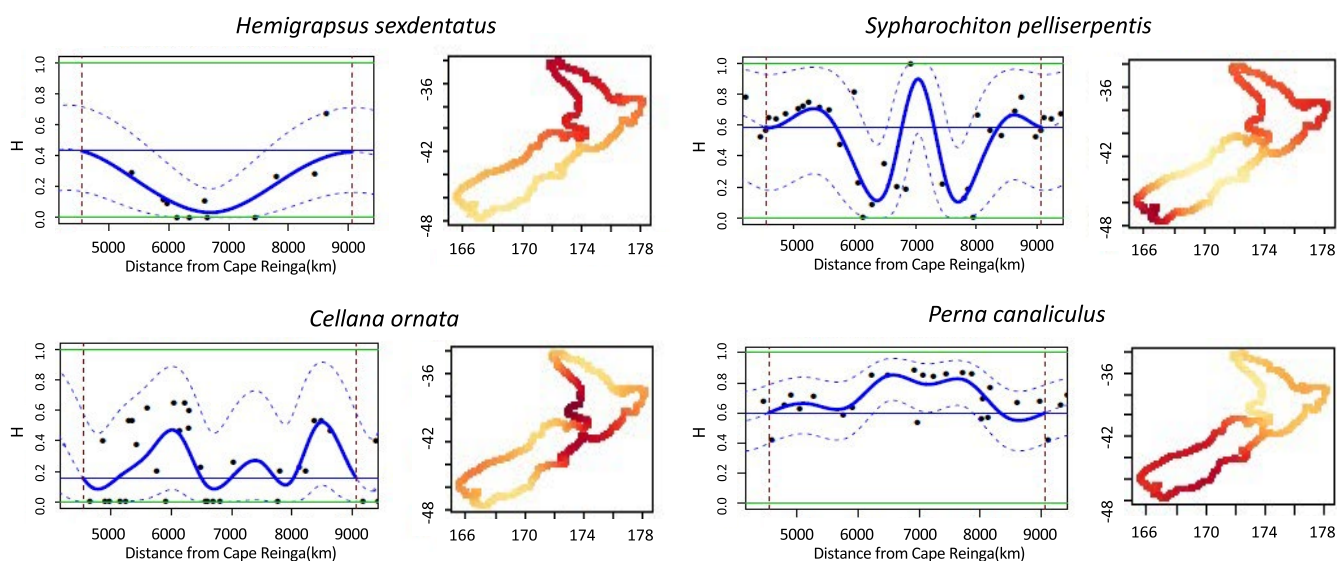


FIGURE 2 Genogeographic curves and colour maps for four species based on the genetic diversity measure H . The horizontal axis corresponds to distance from Cape Reinga. The black dots are the raw data points measuring H at each location. The thick blue curves are the fitted genogeographic curves. The blue horizontal lines demonstrate that the height of each curve is the same at the start and end points each corresponding to Cape Reinga. The dotted lines show the central 95% percentile intervals of the fitted beta distributions about each curve, indicating the estimated dispersion for each species. Horizontal green lines are plotted at 0 and 1, the range of the genetic measure. The colour maps depict the value of H after the curves have been scaled and centred, with increasing red intensity denoting higher H values

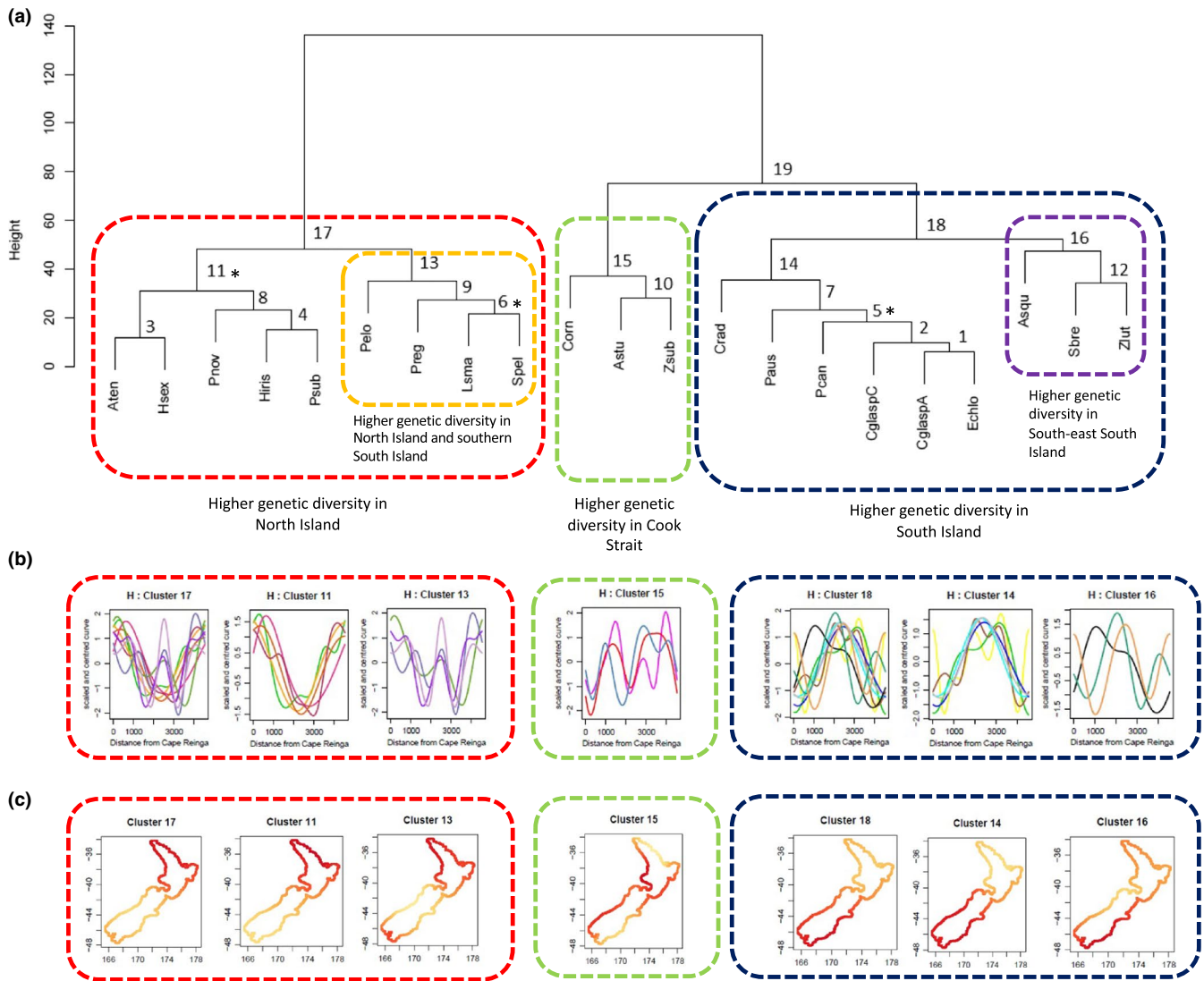


FIGURE 3 (a) Genogeographic species clustering dendrogram based on the genetic diversity measure H , clustering together those species with similar spatial patterns. Cluster number is marked on each branch. Clusters predominantly joined at the 10% significance level are marked with (*). (b) Scaled-and-centred genogeographic curves within selected clusters. (c) Average colour maps of selected clusters

genetic diversity in the South Island, and Cluster 15 grouped species with highest diversity in the central Cook Strait region (Figure 3c).

Cluster 17 comprised two further aggregations, Clusters 11 and 13. Cluster 11 was defined by species with the highest genetic diversity occurring strictly in the North Island, whereas Cluster 13 displayed higher diversity in the North Island and, in addition, a peak of genetic diversity in the southern South Island: see the average map for Cluster 13 (Figure 3c). Cluster 11 had strong significant support at the 5% level (see PCO plot, Figure 4). Within Cluster 13, Cluster 6 joining *L. smaragda* and *S. pelliserpentis* was significantly supported at the 10% level.

The cluster with higher diversity in the South Island, Cluster 18, was further differentiated into two well-defined clusters: 14 and 16. Cluster 14 included species with the overall highest genetic diversity throughout the whole South Island. Within this Cluster 14, several of the species in Cluster 5 were significantly joined, especially at the 10% level (Figure 3). Furthermore, Cluster 5 exhibited several

significant splits from both of the best-supported northern clusters 11 and 6 (Figure 4), indicating strong statistical support for the distinctiveness of the southern versus northern profiles. The three species grouped in the southern sub-cluster 16 were joined together higher in the dendrogram, indicating that their genogeographic patterns had lower levels of similarity, but overall, these species shared two regions of higher genetic diversity: the highest peak in the south-eastern region of the South Island, and a secondary region of high diversity around the northern tip of New Zealand.

Changing the start point of the analysis from Cape Reinga had minimal impact. Only four of the 21 species ever swapped major cluster among all four trials. None of these four species were involved in significant joins or splits, so these swaps were consistent with the findings from our bootstrap method. This emphasizes the importance of the bootstrap significance test in our framework, in terms of exposing cluster memberships that might be sensitive to sampling variability.

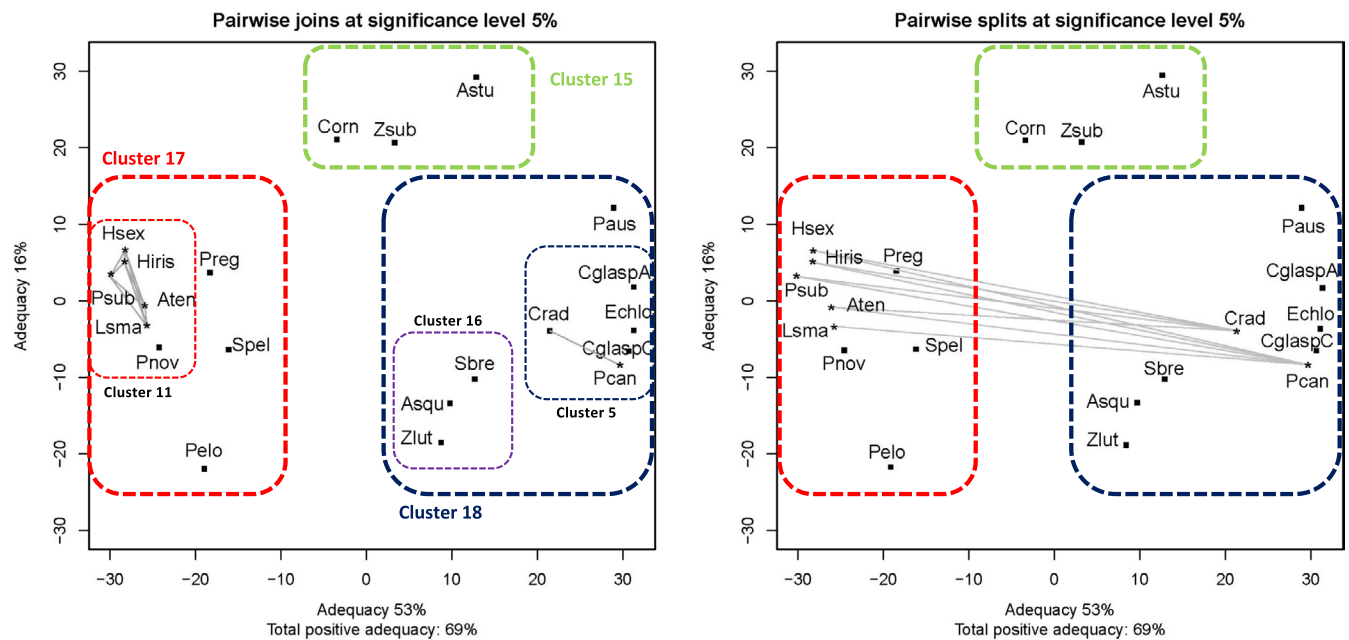


FIGURE 4 PCO plot based on similarity between scaled-and-centred genogeographic curves fitted to the genetic diversity measure H . Grey lines on the left figure show significant pairwise joins at the 5% level, and those on the right figure denote significant pairwise splits at the 5% level. Major clusters are marked on each panel

4 | DISCUSSION

4.1 | Evaluating the statistical approach

We have presented an innovative approach to comparative multi-species analysis of spatial genetic patterns that accommodates species sampled at different locations and analysed using different genetic markers. In general, the fitted genogeographic curves appeared realistic, reflecting known major patterns of spatial genetic variation without noisy detail. The curves illuminated several distinctive patterns of spatial variation, enabling species to be clustered according to their similarity in pattern.

4.1.1 | Use of original genetic data

Analyses that are solely based on a compilation of conclusions from individual studies are likely to be impacted by the questions that were being investigated in the contributory studies. In New Zealand, well-documented biogeographic boundaries are likely to have predisposed authors to test for specific concordance with the resulting bioregions (Shears et al., 2008), potentially overlooking other interesting patterns in the spatial profile of genetic diversity. By using the raw within-population genetic diversities extracted from the original studies, our analysis retained all the geographic information available, so the patterns exposed were not confined to those reported by the original authors.

4.1.2 | No limitations on the genetic measures used

Our methodology can accommodate any measure of genetic diversity or divergence, depending on the questions to be addressed. In this initial study, a within-population genetic diversity measure (H) was chosen for illustration. However, other metrics of genetic properties could equally well be employed. For the purposes of identifying priority regions for conservation, a measure of phylogeographic endemism (Rosauer et al., 2009) may be appropriate. Where the goal is to identify regions of common past demographic expansion, metrics of departure from neutrality may be useful, such as Tajima's D (Tajima, 1989) or Fu's F (Fu, 1997).

A direction of particular interest would be to incorporate measures of genetic divergence among sampled locations, such as F_{ST} , Φ_{ST} or D_{Jost} . These metrics compare two or more sampled locations, whereas diversity measures such as H are associated with a single sample location, so attention would need to be given to associating a population divergence metric with specific locations. Illuminating regions of high and low genetic divergence that are common across species could bring substantial benefits for defining discrete geographic units for conservation and management (Pikitch et al., 2004; Villamor et al., 2014). However, although our method can detect common hotspots of genetic divergence at a regional level, it will not reveal the precise locations of sudden discontinuities due to the smooth nature of the trend curves. If discontinuities were hypothesized at particular locations, we could fit curves that were piecewise smooth, which could then be clustered as usual.

4.1.3 | Comparisons across a wide range of species

A principal advantage of this methodology was to allow comparison across a wide range of species, permitting the inclusion of many large datasets from previous publications with disparate sampling localities. It is usually recommended that a common sampling design be used to compare spatial genetic patterns among species, especially to uncover the underlying processes driving common patterns (Dawson et al., 2014; Liggins et al., 2016; Villamor et al., 2014). However, by creating a statistical framework that does not rely upon a common sampling design, we can fully exploit the wealth of existing genetic data. Given the success of this trial, a much larger data set from New Zealand waters could be analysed to provide a more comprehensive assessment of marine biodiversity patterns around the country.

4.1.4 | Limitations and extensions

The framework we have presented is an exploratory one, with various options that can be adjusted according to the context and goals of the study. One such option concerns the scaling of genogeographic curves to a common mean and standard deviation before clustering. This is a pragmatic strategy that enabled us to illuminate commonalities in spatial genetic patterns across species, regardless of the patterns manifesting with different magnitudes in different species. It is expected that genetic measures will manifest with different absolute value and variability in different species due to different population sizes and mutation rates. By applying the scaling step, we were able to focus on spatial pattern irrespective of absolute value. The bootstrap method is an important part of this strategy, because spatial patterns for some species may be falsely exaggerated in the scaling step so it is important to identify species for which the shape of the trend curve is not robust to sampling variability. These species will fail to return significant cluster memberships in the bootstrap analysis. Omission of the scaling step would be appropriate if a comparison of absolute diversity patterns were of primary interest and can readily be achieved by toggling a switch in the code provided. This option would require comparisons to be restricted to species that were analysed with a common genetic marker.

The methodology as presented is most readily applied to closed-loop coastline geometries that can be readily linearized. Application to linear coastline geometries is analytically straightforward, but there might be practical difficulties if the species of interest are sampled over different spatial ranges. Additionally, there will be a loss of precision in the estimated trend curves towards the extremes of the range, so it might be useful to apply buffering to ensure the clustering of trends across species is not influenced by edge effects. This can be accomplished by omitting a minor section of each trend curve at the start and end of the range when clustering the curves.

Although many comparative studies focus on one-dimensional coastlines and river systems that can be naturally linearized (Cahill et al., 2017; Crandall et al., 2019; Matias & Riginos, 2018; Mertens

et al., 2018; Nielsen et al., 2017; Stanley et al., 2018), more complex geometries may be of interest. The linearization step might be hard to implement over complex coastlines such as archipelagos. This difficulty could potentially be addressed by undertaking the analysis at both broad and fine scales. Expansion of the methodology to two-dimensional systems would involve the calculation of "genogeographic surfaces," which could then be clustered across species. Reduction of edge effects through buffering would likely become more important for such higher dimensional analyses.

Our approach appears to be successful at discerning the similarity of broad regional patterns, but might miss fine-scale geographic differences due to the smoothing procedures. Areas where fine-scale patterns are of interest could be examined over a smaller geographic range, permitting more detailed curve-fitting. Preliminary work suggests our framework is promising for detecting major spatial patterns of genetic divergence, as opposed to diversity, but there may be greater difficulty in using it to discern patterns of chaotic patchiness (Johnson & Black, 1982) or isolation-by-distance (Wright, 1943). However, with respect to fitting the genogeographic curves used here, these processes would tend only to add noise rather than to create bias.

Finally, as with any meta-analysis, the conclusions are only as robust as the data supplied. Our methodology allows multiple studies with disparate sampling regimes to be included in the same analysis, and it estimates the uncertainty separately around each genogeographic curve to ensure that only those curves with strong statistical support will generate significant outcomes. However, if a predominance of contributory studies missed some important feature of a spatial genetic profile, perhaps due to low sampling coverage or a common gap arising from shared sampling designs, the same limitation will apply to the meta-analysis.

4.2 | Common genetic diversity patterns for New Zealand coastal species

Spatial genetic diversity patterns have been observed across the geographic distribution of many species and have been used in the past to infer range expansion and other historical processes affecting the population demography of a species (Templeton et al., 1995). The results here suggested a high variation in patterns of genetic diversity among the species analysed, but some strong commonalities in spatial patterns were also highlighted by the species clusters detected.

Importantly, if genetic diversity patterns were simply averaged across all species, as is commonly undertaken in comparative phylogeography studies (Miraldo et al., 2016; Selkoe et al., 2016), no geographic patterns would have been evident. This is best exemplified by comparing the maps of genetic diversity within each major cluster (Figure 3c) with those for all species combined (cluster 20; Figure S4). It was only when groups of similar species were discriminated through clustering that patterns became evident. Our analysis revealed at least three major geographic patterns, each shared by a number of species; but because the regions of high and low

diversity differed between clusters, these counteracted each other and caused the spatial trend to disappear altogether when all species were combined (Figure S4). Notably, the clusters of species identified were not obvious groupings by any biological criteria such as taxonomy or life history and would have been unlikely to have been discerned through *a priori* clustering within any traditional analysis of diversity.

The principal divide between spatial diversity patterns in our data occurred between Cluster 17, corresponding to species with highest genetic diversity in the North Island, and Cluster 19, which grouped species with the overall highest genetic diversity in Cook Strait (Cluster 15) and the South Island (Cluster 18) (Figure 3). The spatial genetic pattern displayed by the species of Cluster 17 corresponds to an inverse correlation of genetic diversity with latitude. This pattern has already been documented in several of the species included in this cluster, such as *P. novaeseelandiae* (Silva & Gardner, 2015), *H. iris* (Will et al., 2011), *P. regularis* (Ayers & Waters, 2005) and *L. smaragda* (Arranz et al., 2021). This spatial diversity distribution agrees with global patterns of genetic diversity, where high levels of genetic diversity are often expected at lower latitudes (Adams & Hadly, 2013), together with evidence that speciation rates are also higher in the tropics (Jablonski et al., 2006).

The species grouped in Cluster 18, with higher diversity in the south, presumably had a different ancestral origin or were affected by alternative evolutionary processes. It is possible that this genetic pattern was driven by higher population sizes in the south for these species. Human actions such as harvesting or protection may also have impacted population sizes and thus genetic diversity for some species (Allendorf et al., 2014; Silva & Gardner, 2015).

Species in Cluster 15 had highest diversity in the central Cook Strait region. A possible explanation for this is the mid-domain effect, commonly referred to as the “abundant centre” model, whereby populations at the edges of their distribution are expected to have lower abundance (Colwell & Lees, 2000). This pattern is also one component of the “core-periphery” or “central-marginal” hypothesis (Eckert et al., 2008). Most of the species examined in this study were endemic, with range confined to the North, South and Stewart Islands, such that the Cook Strait region corresponds to the central part of their distribution. However, other species have a wider distribution range, including areas such as East Australia, Chatham Islands and Kermadec Islands, among others, that were not covered in this analysis. Studies that encompass the entire distribution range of a species can be particularly informative about the specific factors that are instrumental in shaping the spatial genetic diversity (Hasselman et al., 2013; Silva & Gardner, 2015).

Overall, genogeographic analysis of our study data illuminated several groups of species within the New Zealand intertidal marine community with considerable similarity in their spatial genetic patterns. These findings will facilitate study of the variables that unite species with similar patterns, leading to a better understanding of the biological and environmental factors governing the community assemblage.

5 | CONCLUSIONS

A major limitation of previous comparative phylogeographic approaches has been the need to group either populations or species prior to conducting quantitative analysis (Kelly and Palumbi, 2010; Riginos et al., 2011). The methodology used here provides an objective assessment of the genetic patterns present, therefore avoiding the possible biases introduced by *a priori* groupings (Pelc et al., 2009).

The analytical approach developed here provides a powerful exploratory method that takes full advantage of existing published data to better understand the predominant spatial patterns of genetic diversity across a community. These findings can be used to generate further investigations of the processes shaping biodiversity, by finding characteristics in common among the species clustered together. Such investigations may then drive more specific tests of hypotheses and population models that can foster deeper understanding of the historical, ecological and biological processes underlying the spatial patterns revealed.

ACKNOWLEDGMENTS

We thank the many authors who previously published valuable genetic data that contributed to this study. Special thanks to Martin Hingston (University of Auckland) and Clinton Duffy (New Zealand Department of Conservation) for setting the foundations of this study. We acknowledge Mat Pawley for the helpful discussions. This research was supported financially by the New Zealand Department of Conservation, the Beate Schuler Doctoral Scholarship (to VA) and the University of Auckland.

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

DATA AVAILABILITY STATEMENT

The R code, instructions and example data to use “Genogeographic species clustering” approach are available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.b2rbnzsb>.

ORCID

Vanessa Arranz  <https://orcid.org/0000-0002-3404-5317>

REFERENCES

- Adams, R. I., & Hadly, E. A. (2013). Genetic diversity within vertebrate species is greater at lower latitudes. *Evolutionary Ecology*, 27(1), 133–143. <https://doi.org/10.1007/s10682-012-9587-x>
- Aguirre, J. D., & Marshall, D. J. (2012). Genetic diversity increases population productivity in a sessile marine invertebrate. *Ecology*, 93(5), 1134–1142. <https://doi.org/10.1890/11-1448.1>
- Allendorf, F. W., Berry, O., & Ryman, N. (2014). So long to genetic diversity, and thanks for all the fish. *Molecular Ecology*, 23(1), 23–25. <https://doi.org/10.1111/mec.12574>
- Antunes, A. P., Fewster, R. M., Venticinque, E. M., Peres, C. A., Levi, T., Rohe, F., & Shepard, G. H. (2016). Empty forest or empty rivers? A century of commercial hunting in Amazonia. *Science Advances*, 2(10), e1600936. <https://doi.org/10.1126/sciadv.1600936>

- Apte, S., & Gardner, J. P. A. (2002). Population genetic subdivision in the New Zealand greenshell mussel (*Perna canaliculus*) inferred from single-strand conformation polymorphism analysis of mitochondrial DNA. *Molecular Ecology*, 11(9), 1617–1628. <https://doi.org/10.1046/j.1365-294X.2002.01554.x>
- Arranz, V., Thakur, V., & Lavery, S. D. (2021). Demographic history, not larval dispersal potential, explains differences in population structure of two New Zealand intertidal species. *Marine Biology*, 168(7), 1–14. <https://doi.org/10.1007/s00227-021-03891-2>
- Avise, J. C. (2000). *Phylogeography: The history and formation of species*. Harvard University Press.
- Ayers, K., & Waters, J. (2005). Marine biogeographic disjunction in central New Zealand. *Marine Biology*, 147(4), 1045–1052. <https://doi.org/10.1007/s00227-005-1632-7>
- Beger, M., Selkoe, K. A., Treml, E., Barber, P. H., von der Heyden, S., Crandall, E. D., Toonen, R. J., & Riginos, C. (2014). Evolving coral reef conservation with genetic information. *Bulletin of Marine Science*, 90(1), 159–185. <https://doi.org/10.5343/bms.2012.1106>
- Bernhardt, J. R., & Leslie, H. M. (2013). Resilience to climate change in coastal marine ecosystems. *Annual Review of Marine Science*, 5, 371–392. <https://doi.org/10.1146/annurev-marine-121211-172411>
- Cahill, A. E., De Jode, A., Dubois, S., Bouzaza, Z., Aurelle, D., Boissin, E., Chabrol, O., David, R., Egea, E., Ledoux, J.-B., Mérigot, B., Weber, A.-T., & Chenuil, A. (2017). A multispecies approach reveals hot spots and cold spots of diversity and connectivity in invertebrate species with contrasting dispersal modes. *Molecular Ecology*, 26(23), 6563–6577. <https://doi.org/10.1111/mec.14389>
- Cheeseman, J. F., Fewster, R. M., & Walker, M. M. (2017). Circadian and circatidal clocks control the mechanism of semilunar foraging behaviour. *Scientific Reports*, 7, 3780. <https://doi.org/10.1038/s41598-017-03245-3>
- Coates, D. J., Byrne, M., & Moritz, C. (2018). Genetic diversity and conservation units: Dealing with the species-population continuum in the age of genomics. *Frontiers in Ecology and Evolution*, 6, 165. <https://doi.org/10.3389/fevo.2018.00165>
- Colwell, R. K., & Lees, D. C. (2000). The mid-domain effect: Geometric constraints on the geography of species richness. *Trends in Ecology & Evolution*, 15(2), 70–76. [https://doi.org/10.1016/S0169-5347\(99\)01767-X](https://doi.org/10.1016/S0169-5347(99)01767-X)
- Crandall, E. D., Riginos, C., Bird, C. E., Liggins, L., Treml, E., Beger, M., Barber, P. H., Connolly, S. R., Cowman, P. F., DiBattista, J. D., Eble, J. A., Magnuson, S. F., Horne, J. B., Kochzius, M., Lessios, H. A., Liu, S. Y. V., Ludt, W. B., Madduppa, H., Pandolfi, J. M., ... Gaither, M. R. (2019). The molecular biogeography of the Indo-Pacific: Testing hypotheses with multispecies genetic patterns. *Global Ecology and Biogeography*, 28(7), 943–960. <https://doi.org/10.1111/geb.12905>
- Dawson, M. N. (2014). Natural experiments and meta-analyses in comparative phylogeography. *Journal of Biogeography*, 41(1), 52–65. <https://doi.org/10.1111/jbi.12190>
- Dawson, M. N., Hays, C. G., Grosberg, R. K., & Raimondi, P. T. (2014). Dispersal potential and population genetic structure in the marine intertidal of the eastern North Pacific. *Ecological Monographs*, 84(3), 435–456. <https://doi.org/10.1890/13-0871.1>
- Eckert, C., Samis, K., & Loughheed, S. (2008). Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular Ecology*, 17(5), 1170–1188. <https://doi.org/10.1111/j.1365-294X.2007.03659.x>
- Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3), 564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
- Fu, Y.-X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147(2), 915–925. <https://doi.org/10.1093/genetics/147.2.915>
- Garrick, R. C., Caccione, A., & Sunnucks, P. (2010). Inference of population history by coupling exploratory and model-driven phylogeographic analyses. *International Journal of Molecular Sciences*, 11(4), 1190. <https://doi.org/10.3390/ijms11041190>
- Goldstien, S. J., Schiel, D. R., & Gemmill, N. J. (2006). Comparative phylogeography of coastal limpets across a marine disjunction in New Zealand. *Molecular Ecology*, 15(11), 3259–3268. <https://doi.org/10.1111/j.1365-294X.2006.02977.x>
- Hannan, D. A. (2014). Population genetics and connectivity in Paphies subtriangulata and Paphies australis (Bivalvia: Mesodesmatidae). Victoria University, Wellington.
- Hasselmann, D. J., Ricard, D., & Bentzen, P. (2013). Genetic diversity and differentiation in a wide ranging anadromous fish, American shad (*Alosa sapidissima*), is correlated with latitude. *Molecular Ecology*, 22(6), 1558–1573. <https://doi.org/10.1111/mec.12197>
- Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models, volume 43 of Monographs on statistics and applied probability*. Chapman & Hall.
- Hewitt, G. M. (2004). The structure of biodiversity—insights from molecular phylogeography. *Frontiers in Zoology*, 1(1), 4. <https://doi.org/10.1186/1742-9994-1-4>
- Hickerson, M. J., & Meyer, C. P. (2008). Testing comparative phylogeographic models of marine vicariance and dispersal using a hierarchical Bayesian approach. *BMC Evolutionary Biology*, 8(1), 322. <https://doi.org/10.1186/1471-2148-8-322>
- Hickerson, M. J., Stahl, E., & Takebayashi, N. (2007). msBayes: Pipeline for testing comparative phylogeographic histories using hierarchical approximate Bayesian computation. *BMC Bioinformatics*, 8(1), 1–7. <https://doi.org/10.1186/1471-2105-8-268>
- Hinnendael, F. (2008). *Phylogeny of worldwide Varunidae and phylogeography of New Zealand Hemigrapsus*. MSc, University of Auckland, Auckland.
- Hughes, A. R., Inouye, B. D., Johnson, M. T., Underwood, N., & Vellend, M. (2008). Ecological consequences of genetic diversity. *Ecology Letters*, 11(6), 609–623. <https://doi.org/10.1111/j.1461-0248.2008.01179.x>
- Jablonski, D., Roy, K., & Valentine, J. W. (2006). Out of the tropics: Evolutionary dynamics of the latitudinal diversity gradient. *Science*, 314(5796), 102–106. <https://doi.org/10.1126/science.1130880>
- Johnson, M., & Black, R. (1982). Chaotic genetic patchiness in an intertidal limpet. *Siphonaria* sp. *Marine Biology*, 70(2), 157–164. <https://doi.org/10.1007/BF00397680>
- Keeney, D. B., Szymaniak, A. D., & Poulin, R. (2013). Complex genetic patterns and a phylogeographic disjunction among New Zealand mud snails *Zeacumantus subcarinatus* and *Z-lutulentus*. *Marine Biology*, 160(6), 1477–1488. <https://doi.org/10.1007/s00227-013-2201-0>
- Kelly, R. P., & Palumbi, S. R. (2010). Genetic structure among 50 species of the northeastern Pacific rocky intertidal community. *PLoS One*, 5(1), e8594. <https://doi.org/10.1371/journal.pone.0008594>
- Lamy, T., Laroche, F., David, P., Massol, F., & Jarne, P. (2017). The contribution of species–genetic diversity correlations to the understanding of community assembly rules. *Oikos*, 126(6), 759–771. <https://doi.org/10.1111/oik.03997>
- Liggins, L., Booth, D. J., Figueira, W. F., Treml, E. A., Tonk, L., Ridgway, T., Harris, D. A., & Riginos, C. (2015). Latitude-wide genetic patterns reveal historical effects and contrasting patterns of turnover and nestedness at the range peripheries of a tropical marine fish. *Ecography*, 38(12), 1212–1224. <https://doi.org/10.1111/ecog.01398>
- Liggins, L., Treml, E. A., Possingham, H. P., & Riginos, C. (2016). Seascape features, rather than dispersal traits, predict spatial genetic patterns in co-distributed reef fishes. *Journal of Biogeography*, 43(2), 256–267. <https://doi.org/10.1111/jbi.12647>
- Marko, P. B., & Hart, M. W. (2011). The complex analytical landscape of gene flow inference. *Trends in Ecology & Evolution*, 26(9), 448–456. <https://doi.org/10.1016/j.tree.2011.05.007>
- Matias, A. M. A., & Riginos, C. (2018). Revisiting the “Centre Hypotheses” of the Indo-West Pacific: Idiosyncratic genetic diversity of nine reef species offers weak support for the Coral Triangle as a centre of

- genetic biodiversity. *Journal of Biogeography*, 45(8), 1806–1817. <https://doi.org/10.1111/jbi.13376>
- Mertens, L. E., Treml, E. A., & Von der Heyden, S. (2018). Genetic and biophysical models help define marine conservation focus areas. *Frontiers in Marine Science*, 5, 268. <https://doi.org/10.3389/fmars.2018.00268>
- Miraldo, A., Li, S., Borregaard, M. K., Flórez-Rodríguez, A., Gopalakrishnan, S., Rizvanovic, M., Wang, Z., Rahbek, C., Marske, K. A., & Nogués-Bravo, D. (2016). An Anthropocene map of genetic diversity. *Science*, 353(6307), 1532–1535. <https://doi.org/10.1126/science.aaf4381>
- Moritz, C., & Faith, D. P. (1998). Comparative phylogeography and the identification of genetically divergent areas for conservation. *Molecular Ecology*, 7(4), 419–429. <https://doi.org/10.1046/j.1365-294x.1998.00317.x>
- Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of Classification*, 31(3), 274–295. <https://doi.org/10.1007/s00357-014-9161-z>
- Nagel, M. M., Sewell, M. A., & Lavery, S. D. (2015). Differences in population connectivity of a benthic marine invertebrate *Evechinus chloroticus* (Echinodermata: Echinoidea) across large and small spatial scales. *Conservation Genetics*, 16(4), 965–978. <https://doi.org/10.1007/s10592-015-0716-2>
- Nason, J. D., Hamrick, J., & Fleming, T. H. (2002). Historical vicariance and postglacial colonization effects on the evolution of genetic structure in *Lophocereus*, a Sonoran Desert columnar cactus. *Evolution*, 56(11), 2214–2226. <https://doi.org/10.1111/j.0014-3820.2002.tb00146.x>
- Nei, M. (1972). Genetic distance between populations. *The American Naturalist*, 106(949), 283–292. <https://doi.org/10.1086/282771>
- Nielsen, E. S., Beger, M., Henriques, R., Selkoe, K. A., & von der Heyden, S. (2017). Multispecies genetic objectives in spatial conservation planning. *Conservation Biology*, 31(4), 872–882. <https://doi.org/10.1111/cobi.12875>
- Peakall, P., & Smouse, R. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, 28, 2537–2539. <https://doi.org/10.1111/j.1471-8286.2005.01155.x>
- Pelc, R. A., Warner, R. R., & Gaines, S. D. (2009). Geographical patterns of genetic structure in marine species with contrasting life histories. *Journal of Biogeography*, 36, 1881–1890. <https://doi.org/10.1111/j.1365-2699.2009.02138.x>
- Pikitch, E. K., Santora, C., Babcock, E. A., Bakun, A., Bonfil, R., Conover, D. O., Dayton, P., Doukakakis, P., Fluharty, D., Heneman, B., Houde, E. D., Link, J., Livingston, P. A., Mangel, M., McAllister, M. K., Pope, J., & Sainsbury, K. J. (2004). Ecosystem-based fishery management. *Science*, 305(5682), 346–347. <https://doi.org/10.1126/science.1098222>
- R Core Team (2020). *R: A language and environment for statistical computing*. (Version 4.0.1). R Foundation for Statistical Computing. <http://www.R-project.org>
- Riginos, C., Douglas, K. E., Jin, Y., Shanahan, D. F., & Treml, E. A. (2011). Effects of geography and life history traits on genetic differentiation in benthic marine fishes. *Ecography*, 34, 566–575. <https://doi.org/10.1111/j.1600-0587.2010.06511.x>
- Rocha, L., Craig, M., & Bowen, B. (2007). Phylogeography and the conservation of coral reef fishes. *Coral Reefs*, 26(3), 501–512. <https://doi.org/10.1007/s00338-007-0261-7>
- Rosauer, D., Laffan, S. W., Crisp, M. D., Donnellan, S. C., & Cook, L. G. (2009). Phylogenetic endemism: A new approach for identifying geographical concentrations of evolutionary history. *Molecular Ecology*, 18(19), 4061–4072. <https://doi.org/10.1111/j.1365-294X.2009.04311.x>
- Ross, P., Hogg, I., Pilditch, C., Lundquist, C., & Wilkins, R. (2012). Population genetic structure of the New Zealand Estuarine Clam *Austrovenus stutchburyi* (Bivalvia: Veneridae) reveals population subdivision and partial congruence with biogeographic boundaries. *Estuaries and Coasts*, 35(1), 143–154. <https://doi.org/10.1007/s12237-011-9429-z>
- Schiebelhut, L. M., & Dawson, M. N. (2018). Correlates of population genetic differentiation in marine and terrestrial environments. *Journal of Biogeography*, 45(11), 2427–2441. <https://doi.org/10.1111/jbi.13437>
- Selkoe, K. A., Gaggiotti, O. E., ToBo, L., Bowen, B. W., & Toonen, R. J. (2014). Emergent patterns of population genetic structure for a coral reef community. *Molecular Ecology*, 23, 3064–3079. <https://doi.org/10.1111/mec.12804>
- Selkoe, K. A., Gaggiotti, O. E., Treml, E. A., Wren, J. L., Donovan, M. K., & Toonen, R. J. (2016). The DNA of coral reef biodiversity: predicting and protecting genetic diversity of reef assemblages. *Proceedings of the Royal Society B: Biological Sciences*, 283(1829), 20160354. <https://doi.org/10.1098/rspb.2016.0354>
- Shears, N. T., Smith, F., Babcock, R. C., Duffy, C. A. J., & Villouta, E. (2008). Evaluation of biogeographic classification schemes for conservation planning: Application to New Zealand's Coastal Marine Environment. *Conservation Biology*, 22(2), 467–481. <https://doi.org/10.1111/j.1523-1739.2008.00882.x>
- Silva, C. N., & Gardner, J. (2015). Emerging patterns of genetic variation in the New Zealand endemic scallop *Pecten novaezelandiae*. *Molecular Ecology*, 24(21), 5379–5393. <https://doi.org/10.1111/mec.13404>
- Sponer, R., & Roy, M. S. (2002). Phylogeographic analysis of the brooding brittle star *Amphipholis squamata* (Echinodermata) along the coast of New Zealand reveals high cryptic genetic variation and cryptic dispersal potential. *Evolution*, 56(10), 1954–1967. <https://doi.org/10.1111/j.0014-3820.2002.tb00121.x>
- Stanley, R. R. E., DiBacco, C., Lowen, B., Beiko, R. G., Jeffery, N. W., Van Wyngaarden, M., Bentzen, P., Brickman, D., Benestan, L., Bernatchez, L., Johnson, C., Snelgrove, P. V. R., Wang, Z., Wringe, B. F., & Bradbury, I. R. (2018). A climate-associated multispecies cryptic cline in the northwest Atlantic. *Science Advances*, 4(3), eaaq0929. <https://doi.org/10.1126/sciadv.aaq0929>
- Taberlet, P., Zimmermann, N. E., Englisch, T., Tribsch, A., Holderegger, R., Alvarez, N., Niklfeld, H., Coldea, G., Mirek, Z., Moilanen, A., Ahlmer, W., Marsan, P. A., Bona, E., Bovio, M., Choler, P., Cieślak, E., Colli, L., Cristea, V., Dalmás, J.-P., ... Gugerli, F. (2012). Genetic diversity in widespread species is not congruent with species richness in alpine plant communities. *Ecology Letters*, 15(12), 1439–1448. <https://doi.org/10.1111/ele.12004>
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), 585–595. <https://doi.org/10.1093/genetics/123.3.585>
- Templeton, A. R., Routman, E., & Phillips, C. A. (1995). Separating population structure from population history: A cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics*, 140(2), 767–782. <https://doi.org/10.1093/genetics/140.2.767>
- Toonen, R. J., Andrews, K. R., Baums, I. B., Bird, C. E., Concepcion, G. T., Daly-Engel, T. S., Eble, J. A., Faucci, A., Gaither, M. R., Iacchei, M., Puritz, J. B., Schultz, J. K., Skillings, D. J., Timmers, M. A., & Bowen, B. W. (2011). Defining boundaries for ecosystem-based management: A multispecies case study of marine connectivity across the Hawaiian Archipelago. *Journal of Marine Biology*, 2011, 1–13. <https://doi.org/10.1155/2011/460173>
- Veale, A. (2007). *Population genetics of Actinia tenebrosa and Sypharochiton pelliserpentis*. MSc, University of Auckland, Auckland.
- Veale, A. J., & Lavery, S. D. (2011). Phylogeography of the snakeskin chiton *Sypharochiton pelliserpentis* (Mollusca: Polyplacophora) around New Zealand: Are seasonal near-shore upwelling events a dynamic barrier to gene flow? *Biological Journal of the Linnean Society*, 104(3), 552–563. <https://doi.org/10.1111/j.1095-8312.2011.01743.x>

- Vellend, M., Lajoie, G., Bourret, A., Múrria, C., Kembel, S. W., & Garant, D. (2014). Drawing ecological inferences from coincident patterns of population- and community-level biodiversity. *Molecular Ecology*, 23(12), 2890–2901. <https://doi.org/10.1111/mec.12756>
- Villamor, A., Costantini, F., & Abbiati, M. (2014). Genetic structuring across marine biogeographic boundaries in Rocky Shore Invertebrates. *PLoS One*, 9, e101135. <https://doi.org/10.1371/journal.pone.0101135>
- von der Heyden, S., Beger, M., Toonen, R. J., van Herwerden, L., Juinio-Meñez, M. A., Ravago-Gotanco, R., Fauvelot, C., & Bernardi, G. (2014). The application of genetics to marine management and conservation: examples from the Indo-Pacific. *Bulletin of Marine Science*, 90(1), 123–158. <https://doi.org/10.5343/bms.2012.1079>
- Vucetich, J. A., & Waite, T. A. (2003). Spatial patterns of demography and genetic processes across the species' range: Null hypotheses for landscape conservation genetics. *Conservation Genetics*, 4(5), 639–645.
- Waters, J. M., McCulloch, G. A., & Eason, J. A. (2007). Marine biogeographical structure in two highly dispersive gastropods: Implications for trans-Tasman dispersal. *Journal of Biogeography*, 34(4), 678–687. <https://doi.org/10.1111/j.1365-2699.2006.01615.x>
- Will, M., Hale, M. L., Schiel, D. R., & Gemmill, N. J. (2011). Low to moderate levels of genetic differentiation detected across the distribution of the New Zealand abalone, *Haliotis iris*. *Marine Biology*, 158(6), 1417–1429. <https://doi.org/10.1007/s00227-011-1659-x>
- Wright, S. (1943). Isolation by distance. *Genetics*, 28(2), 114. <https://doi.org/10.1093/genetics/28.2.114>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Arranz, V., Fewster, R. M., & Lavery, S. D. (2022). Genogeographic clustering to identify cross-species concordance of spatial genetic patterns. *Diversity and Distributions*, 28, 611–623. <https://doi.org/10.1111/ddi.13474>

BIOSKETCH

Vanessa Arranz is a Lecturer at Massey University, New Zealand. Her research expertise is in various fields as marine biology, molecular ecology, genetics and bioinformatics to disentangle the ecology, diversity and distributions of organisms. From single-species studies to broader multiple-species views for assessing the patterns and processes driving the genetic connectivity among populations. Recently immerse in the genomics field, using Next Generation Sequencing (NGS) technologies and bioinformatics in various projects including environmental DNA metabarcoding and whole-genome sequencing (WGS) applied to kelp forests and coral reef ecosystems.

Author contribution: V.A. collected and analysed the data and wrote the first draft of the manuscript. R.M.F designed and performed the statistical analysis, and edited the manuscript. S.L secured project funding, designed the analysis and analysed the data. V.A, R.M.F and S.L discussed the data and designed and prepared figures and tables.