Received: 25 September 2021 | Revised: 24 February 2022 | Accepted: 1 March 2022

DOI: 10.1111/1755-0998.13606

MOLECULAR ECOLOGY RESOURCES WILEY

RESOURCE ARTICLE

Information-based summary statistics for spatial genetic structure inference

Xinghu Qin | Oscar E. Gaggiotti

Centre for Biological Diversity, University of St Andrews, Fife, UK

Correspondence
Xinghu Qin, CAS Key Laboratory of Genomics and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences & China National Center for Bioinformation, Beijing, 10010, China.
Email: qin.xinghu@163.com

Oscar E. Gaggiotti, Centre for Biological Diversity, University of St Andrews, Fife, UK.
Email: oeg@st-andrews.ac.uk

Funding information
China Scholarship Council

Handling Editor: Kimberly Gilbert

Abstract

The measurement of biodiversity at all levels of organization is an essential first step to understand the ecological and evolutionary processes that drive spatial patterns of biodiversity. Ecologists have explored the use of a large range of different summary statistics and have come to the view that information-based summary statistics, and in particular so-called Hill numbers, are a useful tool to measure biodiversity. Population geneticists, on the other hand, have focused largely on summary statistics based on heterozygosity and measures of allelic richness. However, recent studies proposed the adoption of information-based summary statistics in population genetics studies. Here, we performed a comprehensive assessment of the power of this family of summary statistics to inform regarding spatial patterns of genetic diversity and we compared it with that of traditional population genetics approaches, namely measures based on allelic richness and heterozygosity. To give an unbiased evaluation, we used three machine learning methods to test the performance of different sets of summary statistics to discriminate between spatial scenarios. We defined three distinct sets, (i) one based on allelic richness measures which included the Jaccard index, (ii) a set based on heterozygosity that included F_{ST} and (iii) a set based on Hill numbers derived from Shannon entropy, which included the recently proposed Shannon differentiation, ΔD. The results showed that the last of these performed as well or, under some specific spatial scenarios, even better than the traditional population genetics measures. Interestingly, we found that a rarely or never used genetic differentiation measure based on allelic richness, Jaccard dissimilarity (J), showed the highest discriminatory power to discriminate among spatial scenarios, followed by Shannon differentiation ΔD. We concluded, therefore, that information-based measures as well as Jaccard dissimilarity represent excellent additions to the population genetics toolkit.

KEYWORDS
information-based statistics, population genetics, spatial structure

1 | INTRODUCTION

Spatial biodiversity patterns generated by different evolutionary and demographic processes can be observed at the ecological or species level and at the genetic or molecular level (Fortuna et al., 2009; Novembre & Stephens, 2008; Stotz et al., 2016; Van Tienderen, 1991; Wang et al., 2011). However, metrics and approaches to describe these spatial patterns and to infer the underlying processes differ greatly between these two biodiversity levels. The metrics used to study ecological variation (species) and genetic variation (alleles) are mainly dominated by the traditional indices in their own domains, such as species richness, Shannon index in ecology, and allelic richness, heterozygosity in population genetics. These indices comprise a spectrum of information measures (q profile, ${}^{q}H$; Hill, 1973; Jost, 2006), including richness ($q = 0$, S), Shannon entropy ($q = 1$, H) and heterozygosity (or Gini–Simpson index, $q = 2$, He). Each index of order q provides a different type of information, with the index of order $q = 0$ treating all elements equally regardless of their frequency, the index of order $q = 2$ disproportionately favours the most common elements, while the index of order $q = 1$ favours common elements in proportion to their population share (Gaggiotti et al., 2018). However, the $q = 1$ family have received only sporadic attention in population genetics.

The use of summary statistics has facilitated our understanding of ecology and evolution in terms of describing spatial biodiversity patterns (e.g., distance-decay; Nekola & White, 1999), and examining likely processes underlying them. Typically, this is done by decomposing total diversity (γ-diversity) into within-aggregate (α-diversity) and between aggregate (β-diversity) diversity based on species' or community spatial aggregation (Jost, 2007). The derived β-diversity is then used to examine the dissimilarity or differentiation between aggregates. Two main decomposition methods have been used to do this, multiplicative ($SS_\gamma = SS_\beta * SS_\alpha$) and additive ($SS_\gamma = SS_\beta + SS_\alpha$) decomposition (Ricotta, 2005). The desirable β-diversity should be additive when pooling or partitioning the aggregates and should represent the actual proportion of nonshared elements (true dissimilarity or differentiation) due to divergence or differentiation between aggregates (Chao et al., 2014).

For any given value of q, it is possible to obtain a similarity or differentiation measure. In principle, true measures of differentiation should vary between 0 (when assemblages are identical) and 1 (when assemblages are fully differentiated). In the case of $q = 0$ (allelic richness), an appropriate differentiation measure is given by the Jaccard dissimilarity index (but see Jost et al., 2011 for other measures), which measures the overall proportion of shared alleles in the combined assemblage and is a true differentiation measure (cf. Jost et al., 2011). Differentiation measures of order $q = 1$, (i.e., Shannon differentiation, ΔD) also measure true differentiation (cf. Gaggiotti et al., 2018). However, the traditional and most popular measures of differentiation of order $q = 2$ used by population geneticists, namely F_{ST} and G_{ST}, are not true differentiation measures as they can be much lower than 1 for multiallelic markers even in the case of two assemblages that share no alleles (Jost, 2008). An alternative true differentiation measure for $q = 2$ is Jost's (2008) D_{EST}, which is particularly appropriate for conservation genetic studies. Note, however, that although not a true differentiation measure, the fixation index F_{ST} is a fundamentally important parameter for the study of evolutionary processes (cf. Whitlock, 2011).

A common difficulty faced when measuring biodiversity with standard metrics is that, with the exception of richness, they do not have an intuitive interpretation in terms of the number of effective elements in the system (Jost, 2006). However, this problem is easily overcome by using Hill numbers (Hill, 1973), and this is the approach we used in the present study. Thus, allelic richness is represented by ${}^{0}D$ while the effective numbers of alleles based on Shannon entropy and heterozygosity are given by ${}^{1}D$ and ${}^{2}D$ respectively.

Diversity at one level of biological organization (community, species) may sustain the diversity at the other (Lankau & Strauss, 2007). Thus, in addition to describing diversity patterns, researchers have made substantial efforts to unify the two levels of biodiversity (species diversity of ecological communities and genetic diversity of populations) and to reveal ecological and evolutionary processes underpinning their spatial patterns (Vellend, 2005). However, these so-called species-genetic-diversity-correlation (SGDC) studies have rarely measured the two types of diversity consistently (Gaggiotti et al., 2018). Integrative studies of species and genetic diversity, and the ecological factors underlying their association or lack thereof using the same type of index, would contribute to a better understanding of eco-evolutionary dynamics. Although not focused on spatial scenarios, some recent studies (Gaggiotti et al., 2018; Luiselli et al., 2021; Overcast et al., 2019, 2021) have developed a community assembly model that makes predictions of genetic, species and functional diversity in terms of Hill numbers.

The use of informative diversity metrics is crucial, not only for detecting changes in biodiversity patterns but also for understanding the demographic and evolutionary history of species (Csilléry et al., 2010). The performance of population genetics summary statistics has been thoroughly evaluated in the context of spatial demographic inference (Alvarado-Serrano & Hickerson, 2016) and similar studies are needed for equivalent statistics based on Shannon entropy. The present study represents the first step in this direction by evaluating the power of the information-based diversity measures (represented by ${}^{1}D$ and Shannon differentiation, ΔD) and comparing it with that of traditional measures (represented by allelic richness, heterozygosity and their β-diversity measures) to discriminate between spatial scenarios using recent machine learning approaches.

We simulated microsatellite data under five spatial scenarios that include panmixia, the finite island model, hierarchical island model, stepping-stone model and hierarchical stepping-stone model, which are the typical spatial demographic models that have been used to describe the spatial structure of natural populations in fragmented landscapes. We employed three machine learning approaches, kernel local Fisher discriminant analysis (KLFDA), conditional random forest classification and deep neural networks, to characterize the behaviour of these diversity metrics for discriminating different spatial scenarios. Our results showed that information-based summary statistics can provide more power than traditional measures to make inferences about spatial genetic structure.

TABLE 1 Parameters used in the simulations. In the case of the panmixia scenario, we simulated a single population but generated 16 samples at random. In the case of the hierarchical models, we indicate the number of populations per region in parentheses

Scenarios	Regions	Number of populations	Population size	Sample size	Migration rate	Mutation rate	Number of loci
Panmixia	1	1 (16)[a]	U (1600, 16,000)	320		5×10^{-4}	10
Island model	1	16	U (100, 1000)	20	U (0.001, 0.1)	5×10^{-4}	10
Hierarchical island model	4 (4,4,4,4)	16	U (100, 1000)	20	m_{within}: U (0.001, 0.1) $m_{between}$: U (5E−5, 5E−3)	5×10^{-4}	10
Stepping-stone	1	16	U (100, 1000)	20	U (0.001, 0.1)	5×10^{-4}	10
Hierarchical stepping-stone	2 (8,8)	16	U (100, 1000)	20	m_{within}: U (0.001, 0.1)) $m_{between}$: U (5E−5, 5E−3)	5×10^{-4}	10

2 | METHODS

To evaluate the ability of the three families of summary statistics in discriminating different spatial scenarios, we simulated five spatial scenarios that encompass hierarchical and nonhierarchical population structures using coalescent simulations. More specifically, we considered populations without hierarchical structure and populations structured into three hierarchical levels, ecosystem, aggregate (e.g., region) and subaggregate (e.g., population) level.

We calculated the three families of summary statistics from these scenarios and then used the machine learning approaches to test their power to discriminate among spatial scenarios.

2.1 | Models and model parameters

We considered five spatial scenarios, panmixia, the island model, hierarchical island model, stepping-stone model and hierarchical stepping-stone model. Instead of using fixed values for the parameters, we sampled them from probability distributions. Table 1 presents all scenarios and the respective parameter distributions used in the simulations. For the island model, stepping-stone model, hierarchical island model and hierarchical stepping-stone model, each scenario consisted of 16 populations with population size sampled from U(100, 1000). For the panmixia model, we simulated one panmictic population, with population size drawn from U(1600, 16,000). The hierarchical island models consist of four regions with each region comprising four populations. In terms of the hierarchical stepping-stone models, we simulated two regions with each region comprising eight populations. We assume a stepwise mutation model with a constant mutation rate of 5×10^{-4} for all scenarios. In the case of the nonhierarchical scenarios (island model and stepping-stone model), the migration rate, m, was drawn from a uniform distribution U(0.001, 0.1). In the case of the hierarchical scenarios, migration rates between pairs of populations within regions were sampled from U(0.001, 0.1) and migration rates between populations from different regions were sampled from U(0.00005, 0.005).

2.2 | Simulations

The coalescent-based simulator FASTSIMCOAL2 (Excoffier et al., 2013; Excoffier & Foll, 2011) was used to generate microsatellite synthetic data under the five scenarios described above. For each of these five spatial scenarios, we simulated 10 independent microsatellite loci sharing the same mutation rate. One hundred sets of parameters (100 simulations) were randomly drawn from prior distributions, and each parameter set was used to generate 1000 replicate data sets. We sampled 20 individuals per population under each spatial model (standard and hierarchical versions of the island and stepping-stone models). In the case of the panmixia model, we sampled 320 individuals and then randomly partitioned them into 16 samples consisting of 20 individuals each to obtain a set of samples equivalent to those of the other four scenarios.

2.3 | Summary statistics

We chose the commonly used genetic diversity indices, allelic richness (*Ar*, noted ^{Ar}SS hereafter) and heterozygosity (*He*, noted ^{He}SS hereafter) as well as their corresponding differentiation measures (β-diversity measures), Jaccard dissimilarity (Jaccard, 1912) and fixation index (Weir & Cockerham, 1984), as the traditional summary statistics. The allelic richness and expected heterozygosity were partitioned into three hierarchical levels, population level, regional level and ecosystem level, with the corresponding measures being, A_r^P (allelic richness at the population level), A_r^R (allelic richness at the regional level) and A_r^T (total allelic richness in the ecosystem), and H_e^P (expected heterozygosity at the population level), H_e^R (expected heterozygosity at the regional level) and H_e^T (total heterozygosity in the ecosystem). Accordingly, the differentiation measures were partitioned into J_r^P (Jaccard dissimilarity among populations within a region) and J_r^R (Jaccard dissimilarity among regions within an ecosystem) for allelic richness, and F_{ST}^P (F_{ST} among populations within a region) and F_{ST}^R (F_{ST} among regions within an ecosystem) for expected heterozygosity.

We chose the diversity of order $q = 1$, the transformed Shannon "effective number" ^1D, as well as Shannon differentiation (ΔD) as the new summary statistics (^{1_D}SS). ^1D was also decomposed into the population level, regional level and ecosystem level, which were D_γ, D_α^R and D_α^P, respectively. The equivalent number of regions and the equivalent number of populations were thus D_β^R and D_β^P, respectively. In the same way, the allelic differentiation ΔD was decomposed into differentiation among populations within a region (ΔD^P) and differentiation among regions within an ecosystem (ΔD^R). Details regarding the equations for diversity decomposition can be found in Gaggiotti et al. (2018).

As Shannon entropy avoids undue emphasis on either rare or common alleles (Sherwin et al., 2017), it is increasingly being used in evolutionary biology and molecular ecology as a measure of genetic diversity and evolvability (Day, 2015; Hampe et al., 2003; Wagner, 2017). Therefore, we also use summary statistics based on Shannon entropy (^1H, ^HSS hereafter) for comparison with diversity measures (^{1_D}SS). Shannon entropy per population (H^P), per region (H^R) and total Shannon entropy (H^T) were calculated in line with the same hierarchies above. The additive decomposition of Shannon beta entropy ($H_\beta = H_\gamma - H_\alpha$), was estimated at the population level (H_β^P) and regional level (H_β^R) as well. Here, we also included Shannon differentiation (ΔD) to keep the number of statistics in ^HSS the same as with ^{1_D}SS.

In addition, we also calculated Mantel's *r*, the correlation coefficient between genetic distance and geographical distance for differentiation measures ($\rho_{J,d}$, $\rho_{\Delta D,d}$, $\rho_{F_{ST},d}$), with distance measured in terms of the number of steps (edges) separating any two populations (vertices).

Each set of summary statistics includes the mean and standard deviation (*SD*). For each measure at the population level, we calculated the value for each population and the mean across populations. The total number of summary statistics for ^{Ar}SS is 44, the same as for ^{He}SS. The total number of summary statistics for ^{1_D}SS is 48, the same as for ^HSS. The total number of different summary statistics is 178. The description of summary statistics is shown in Table S1.

2.4 | Data analysis

The pipelines (R functions) to calculate the summary statistics are wrapped in the R package *HierDpart* (Qin, 2019). We built nine subsets of summary statistics, ^{Ar}SS, ^HSS, ^{He}SS, ^{1_D}SS, $^{Ar+He}SS$, $^{H+^1D}SS$, $^{Ar+H+He}SS$, $^{Ar+He+^1D}SS$ and $^{Ar+H+He+^1D}SS$, for the discriminatory power test.

2.5 | The power of summary statistics to discriminate among spatial scenarios

The power assessed by various machine learning methods may differ. Thus, to ensure that our tests are as comprehensive as possible, we employed three machine learning approaches to evaluate the power of the different subsets of summary statistics to discriminate among spatial structure scenarios: KLFDA (Sugiyama, 2007), conditional random forest classification (CRFC; Strobl et al., 2007) and deep neural network (Ripley & Hjort, 1996).

2.5.1 | Kernel local fisher discriminant analysis

KLFDA is a supervised dimensionality reduction based on local Fisher discriminant analysis (LFDA, Sugiyama, 2007). As opposed to the standard Fisher discriminant analysis (LDA), LFDA can separate different classes (e.g., genetic clusters) while preserving the within-class structure (Sugiyama, 2007); in other words, it allows for genetic substructuring within clusters. KLFDA represents an extension of LFDA that considers nonlinear boundaries between classes (see Sugiyama, 2007 for a detailed explanation).

We carried out KLFDA on the nine subsets of summary statistics. The Gaussian kernel was chosen for kernel transformation. Three key hyperparameters impact the accuracy of KLFDA: *d*, the number of reduced features for discriminant analysis, σ, the radius (the standard deviation) of the Gaussian kernel, and *knn*, the number of nearest neighbours. We first determined the appropriate number of reduced features ranging from five to 50 based on classification accuracy during training. We then performed fine hyperparameter tuning on σ and *knn* via cross-validation with the best number of reduced features selected in the first step. The σ value was tuned considering values between 0.001 and 10 (0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 5, 10) and *knn* was tuned between 5 and 50 (5, 10, 15, 20, 25, 30, 35, 40, 45, 50). Analyses were implemented using the R packages *lfda* and *DA* (Qin et al., 2021; Tang & Li, 2016, 2017).

TABLE 2 The overall performance of different sets of summary statistics in discriminating five spatial scenarios using KLFDA

Summary statistics	^{Ar}SS	^{H}SS	^{He}SS	$^{1}{}_{D}$SS	$^{Ar+He}$SS	$^{H+1D}$SS	$^{Ar+H+He}$SS	$^{Ar+He+1D}$SS	$^{Ar+H+He+1D}$SS
Accuracy (Acc)	0.934	0.936	0.908	**0.946**	0.926	0.944	0.930	0.944	0.942
Acc 95% CI	(0.909, 0.954)	(0.911, 0.956)	(0.879, 0.932)	(0.922, 0.964)	(0.899, 0.947)	(0.920, 0.962)	(0.904, 0.951)	(0.920, 0.962)	(0.918, 0.961)
Kappa	0.917	0.920	0.885	0.932	0.907	0.930	0.912	0.930	0.927

Note: Acc, accuracy; Acc 95% CI, the 95% interval of accuracy; Kappa, Cohen's kappa coefficient (κ). The best performance (highest Acc) is highlighted in bold.

2.5.2 | Conditional random forest classification

We conducted the unbiased random forest classification based on conditional inference trees (*cforest*) that adopt the subsampling validation process with unbiased variable selection (bootstrap without replacement; Strobl et al., 2007). To avoid overfitting in random forest classification, we optimized the key parameter (*mtry*) that governs the number of features that are randomly chosen to grow each tree from the bootstrapped data. We tuned the parameter *mtry* [mtry ϵ (1: *n*), *n* is the number of variables] via leave-one-out validation with 1000 trees for each subset of summary statistics. The parameter with the highest classification accuracy was chosen as the optimal model for evaluating the performance of the various sets of summary statistics.

The standardized conditional importance of each variable, measured by the mean decrease in accuracy (MDA), was estimated from the optimum model based on bootstrapping without replacement according to Strobl et al. (2008). Analyses were implemented using the R package "*caret*" (Kuhn, 2015) calling the *cforest* function from the *party* package (Hothorn et al., 2010).

2.5.3 | Deep neural network

We conducted neural network (Baum, 1988; Guarnieri et al., 1999) classification using a Multilayer Perceptron (MLP) with three hidden layers and a weight decay to test the performance of the above subsets of summary statistics for spatial structure inference. The deep neural network training was carried out through a backpropagation with weighted decay optimization (a procedure to repeatedly adjust the weights to minimize the difference between true values and observed values) and a nonlinear activation function (logistic) at the output layer. We first made a grid search on the parameter space via cross-validation to minimize the parameter range, and then we tuned the parameters through dense parameter combinations via leave-one-out cross-validation. Finally, we tuned the number of neurons in each hidden layer using: layer1 = (1, 5, 10, 15), layer2 = (0, 5, 10, 15), layer3 = (0, 5, 10, 15), and the rate of decay using: decay = (0, 1e-5, 1e-4, 1e-3, 1e-2, 1e-1). The model with the highest accuracy was chosen as the optimal model for evaluating the performance of various sets of summary statistics.

The (overall) importance of summary statistics is determined based on Garson's algorithm (Garson, 1991; Gevrey et al., 2003), which uses combinations of the absolute values of the weights. We also used neural networks to assess the importance of summary statistics to identify a specific scenario. Deep neural network models were built using the *caret* package (Kuhn, 2008, 2012).

2.6 | Evaluating discriminatory power of different sets of summary statistics

In terms of KLFDA, random forest classification and neural network, we calculated the confusion matrix as well as overall

performance statistics for each set of summary statistics. Overall performance statistics included model accuracy and kappa. The methods for calculating these performance metrics based on Table S5 are presented in the Supporting Information. The detailed description of these statistics can be found in Kuhn and Johnson (2013).

We compared the performance of different sets of summary statistics in discriminating the five spatial scenarios using each of the above-mentioned methods separately to identify the best set of summary statistics.

3 | RESULTS

3.1 | KLFDA inference

Table 2 presents the overall performance of different sets of summary statistics in discriminating five spatial scenarios using KLFDA. The best performing statistics set should have the highest accuracy and the largest kappa value. The results indicated that ^{1D}SS presented the highest discriminatory power at discriminating among scenarios. Although ^{H}SS performed slightly better than ^{Ar}SS and ^{He}SS, it underperformed ^{1D}SS (Table 2). On the other hand, the set of summary statistics with the lowest discriminatory power corresponded to the most commonly used ^{He}SS in population genetics (Table 2).

Although the 95% confidence intervals (Cis) of different summary statistics sets overlap, Figure 1 shows that ^{1D}SS is the only set that allows us to clearly distinguish between all different scenarios. All other summary statistics sets, or the combination thereof, either failed to clearly distinguish between panmixia and the island model or failed to clearly distinguish between the standard stepping-stone model and hierarchical stepping-stone model (Figure 1a–i). ^{1D}SS did a better job at discriminating among all of them (Figure 1d).

The confusion matrix supported these results (Table S2). Specifically, ^{Ar}SS can correctly identify the island model, panmixia and stepping-stone model (100%). However, it did worse in identifying the hierarchical stepping-stone model (Figure 1b; Table S2). ^{H}SS did better at identifying hierarchical scenarios but performed less well in the case of the stepping-stone model (Figure 1c; Table S2). ^{1D}SS performed very well under all scenarios, with the exception of the hierarchical stepping-stone (Figure 1d; Table S2). ^{He}SS performed poorly in most scenarios with the exception of panmixia and hierarchical island scenarios (Figure 1e; Table S2). Combinations of ^{1D}SS with other summary statistics showed similar results to those obtained with ^{1D}SS alone except when including ^{He}SS, in which case discriminatory power was decreased (Table 2; Table S2). In fact, combining ^{He}SS with other summary statistics decreased the discriminatory power. Overall, the hierarchical stepping-stone scenario was the most difficult to identify correctly. ^{1D}SS and ^{H}SS did better at discriminating the hierarchical stepping-stone model from other scenarios (Figure 1; Table S2).

3.2 | Conditional random forest classification

As is the case for KLFDA, among all the sets of summary statistics, ^{He}SS had the lowest classification accuracy (Table 3). Slightly different from KLFDA results, ^{1D}SS and ^{H}SS, having the same discriminatory power, outclassed ^{Ar}SS and ^{He}SS in discriminating the five scenarios (Table 3). Note that conditional random forest did not show a power difference between ^{1D}SS and ^{H}SS, as well as between $^{Ar+H+He}SS$, $^{Ar+He+1D}SS$ and $^{Ar+H+He+1D}SS$ (Table 3).

Compared to KLFDA, the discriminatory power of all sets of summary statistics to discriminate spatial scenarios increased when using conditional random forest (Table 3). Moreover, as opposed to KLFDA results, combining different sets of summary statistics led to an increase in discriminatory power (Table 3). Again, the most difficult scenario to identify is the stepping-stone model. However, consistent with KLFDA results, ^{Ar}SS showed worse performance in distinguishing the stepping-stone scenario than ^{H}SS, ^{1D}SS and ^{He}SS (Table S3). On the other hand, ^{He}SS did a worse job at identifying the hierarchical stepping-stone model compared to ^{Ar}SS, ^{H}SS and ^{1D}SS (Table S3).

A particular advantage of random forest classification is that it allows us to rank individual summary statistics in terms of their discriminatory power. Figure 2 presents the top 30 ranked summary statistics out of a total number of 178 including ^{Ar}SS, ^{H}SS, ^{He}SS and ^{1D}SS. The best performing statistics in discriminating the spatial scenarios were the differentiation measures and their Mantel statistics (ρ). Among all the summary statistics, the standard deviation of the correlation between Jaccard dissimilarity and geographical distance, $SD(\rho_{J,d})$ (belonging to ^{Ar}SS), and the correlation between Shannon differentiation and geographical distance, $\rho_{\Delta D,d}$ (belonging to ^{1D}SS), were the two most important statistics contributing to the ability to discriminate among all spatial scenarios (Figure 2). Among the top 10 most informative statistics, the first (the standard deviation of the correlation between Jaccard dissimilarity and geographical distance, $SD(\rho_{J,d})$), the third (the correlation between Jaccard dissimilarity and geographical distance, $\rho_{J,d}$), the fourth (the standard deviation of Jaccard dissimilarity between regions, $SD\ (J_r^R)$) and the seventh (Jaccard dissimilarity between regions, J_r^R) best-performing statistics belong to ^{Ar}SS. The correlation between Shannon differentiation and geographical distance ($\rho_{\Delta D,d}$), and its standard deviation ($SD(\rho_{\Delta D,d})$), as well as the standard deviation of Shannon differentiation between regions ($SD(\Delta D^R)$), accounting for the second, fifth and eighth most important statistics, belong to ^{1D}SS and ^{H}SS. Only three out of the top 10 summary statistics, the correlation between F_{ST} and geographical distance ($\rho_{Fst,d}$), F_{ST} between regions (F_{ST}^R) and its standard deviation ($SD(F_{ST}^R)$), ranking as the sixth, ninth and tenth best-performing statistics respectively, belong to ^{He}SS (Figure 2).

3.3 | Deep neural network

The deep neural network analysis produced results similar to those of the two previous methods. Generally, the summary statistics can

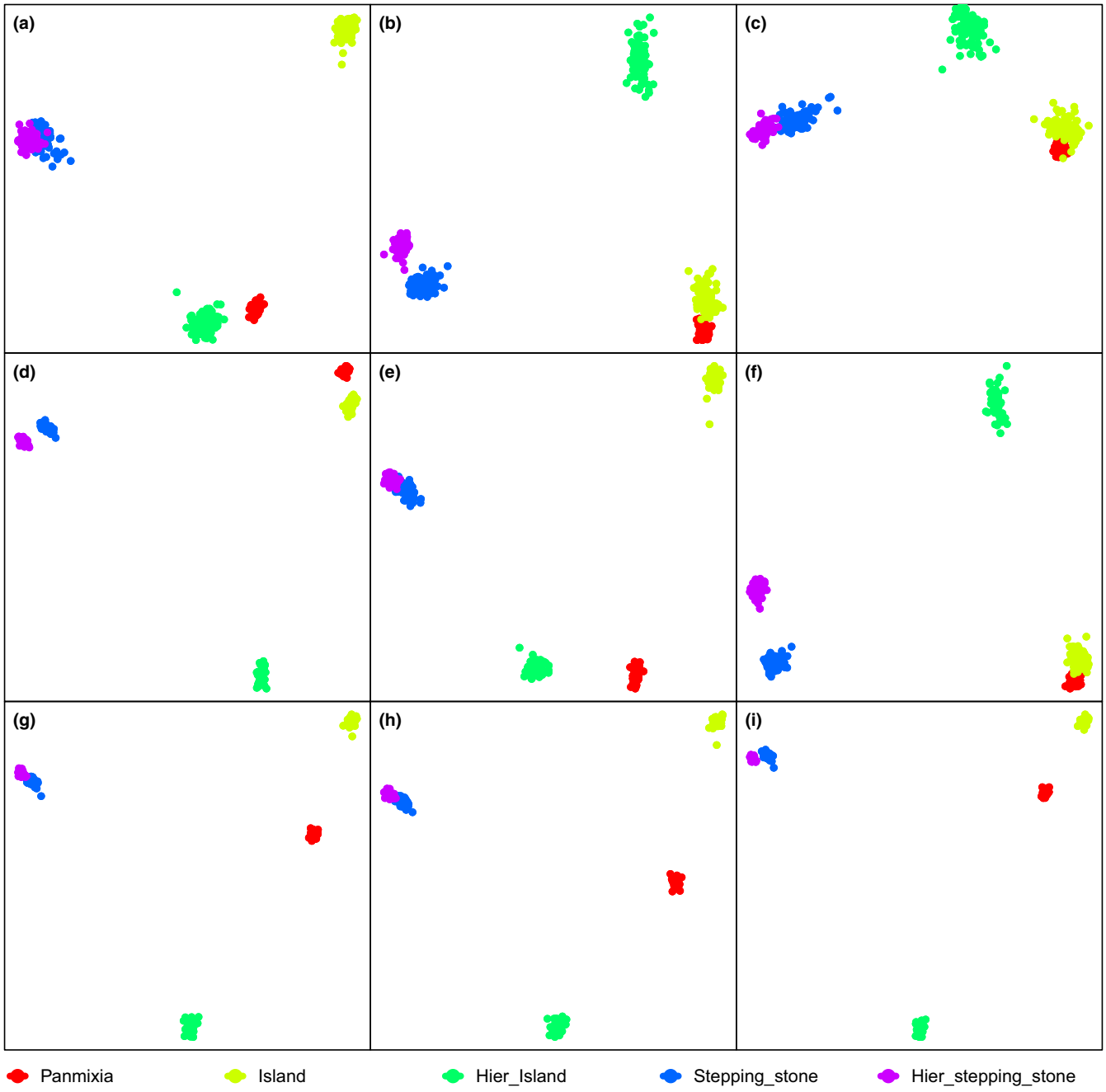

FIGURE 1 Projections of five spatial scenarios into two-dimensional subspaces using KLFDA based on different sets of summary statistics: (a) ^{Ar}SS, (b) ^{H}SS, (c) ^{He}SS, (d) ^{1_D}SS, (e) $^{Ar+He}SS$, (f) $^{H+^1D}SS$, (g) $^{Ar+H+He}SS$, (h) $^{Ar+He+^1D}SS$, (i) $^{Ar+H+He+^1D}SS$. Each dot represents a simulated data set

be categorized into four discriminatory sets based on discriminatory power. Again, ^{1_D}SS, the most powerful summary statistics, along with $^{Ar+He+^1D}SS$, performed best among all the sets of summary statistics (Table 4). ^{Ar}SS, ^{H}SS and $^{Ar+H+He}SS$ comprised the second most discriminatory sets of summary statistics, with their discriminatory accuracy being only slightly lower than that of ^{1_D}SS (0.988, Table 4). The third most discriminatory sets of summary statistics were $^{Ar+He}SS$, ^{H+D}DD and $^{Ar+H+He+^1D}SS$. Finally, the least discriminatory set of summary statistics was ^{He}SS (Table 4). As was the case with KLFDA, neural network results indicated that combining different sets of summary statistics (increasing the number of summary statistics) did not increase the discriminatory power (Table 4).

The discriminatory power of all sets of summary statistics using the neural network was higher than that of KLFDA and CRFC (Tables 2–4). This indicates that the neural network performed better than the two other machine learning (ML) methods. Unlike KLFDA and CRFC, the deep neural network did better at discriminating between the panmixia and island model, with most sets of summary statistics (except ^{He}SS) 100% successfully discriminating between these two scenarios (Table S4). $^{Ar+He+^1D}SS$ and $^{Ar+H+He}SS$ did a better job

TABLE 3 The performance of different sets of summary statistics in discriminating five spatial scenarios using conditional random forest classification

Summary statistics	^{Ar}SS	^{H}SS	^{He}SS	^{1_D}SS	$^{Ar+He}SS$	^{H+1D}SS	$^{Ar+H+He}SS$	$^{Ar+He+1D}SS$	$^{Ar+H+He+1D}SS$
Accuracy (Acc)	0.96	0.972	0.958	0.972	0.970	0.978	**0.986**	**0.986**	**0.986**
Acc 95% CI	(0.939, 0.975)	(0.953,0.985)	(0.936, 0.974)	(0.953, 0.985)	(0.951, 0.983)	(0.961, 0.989)	(0.971, 0.994)	(0.971, 0.994)	(0.971, 0.994)
Kappa	0.950	0.965	0.947	0.965	0.962	0.972	0.982	0.982	0.982

Note: Acc, accuracy; Acc 95% CI, the 95% interval of accuracy; Kappa, Cohen's kappa coefficient (κ). The best performance (highest Acc) is highlighted in bold.

(0.2% error rate) in differentiating the stepping-stone model and hierarchical stepping-stone model compared to other sets of summary statistics (Table S4).

Figure 3 presents the variable importance of the top 30 ranked summary statistics among the total 178 summary statistics according to their discriminatory power estimated from the deep neural network. ^{H}SS, ^{1_D}SS, ^{Ar}SS and ^{He}SS accounted for 11/30 (five overlapped statistics with ^{1_D}SS), 10/30, 8/30 and 6/30 of the top 30 ranked summary statistics respectively (Figure 3). The first three most informative summary statistics were the correlation between Jaccard dissimilarity and geographical distance ($\rho_{J,d}$), the standard deviation of Jaccard dissimilarity between regions (SD (J_r^R)) and the correlation between Shannon differentiation and geographical distance ($\rho_{\Delta D,d}$). They contributed equally toward the ability to discriminate among all spatial scenarios (importance values are all 100, Figure 3; Figure S1). Similar to CRFC results, among the top 10 most informative statistics, the first (correlation between Jaccard dissimilarity and geographical distance, $\rho_{J,d}$), second (the standard deviation of Jaccard dissimilarity between regions, SD (J_r^R)) and the seventh (the standard deviation of the correlation between Jaccard dissimilarity and geographical distance, $SD(\rho_{J,d})$) most important statistics belong to ^{Ar}SS. The correlation between Shannon differentiation and geographical distance, $\rho_{\Delta D,d}$, and its standard deviation, $SD(\rho_{\Delta D,d})$, the standard deviation of the Shannon differentiation between regions, $SD(\Delta D^R)$ and the equivalent number of regions, D_β^R, which were the third, fourth, sixth and tenth best-performing summary statistics respectively, belong to ^{1_D}SS. Only two out of the 10 best-performing statistics, the correlation between F_{ST} and geographical distance, $\rho_{Fst,d}$ and the standard deviation of F_{ST} between regions, $SD(F_{ST}^R)$, ranking as the fifth and the eighth most important summary statistic respectively, belong to ^{He}SS (Figure 3).

Figure S1 presents the scenario-specific variable importance ranked in accordance with their overall importance (cf. Figure 3). The 16 top summary statistics contributed almost equally to the panmixia model, stepping-stone model, hierarchical stepping-stone model and hierarchical island model (Figure S1). On the other hand, only the top five statistics, the correlation between Jaccard dissimilarity and geographical distance, $\rho_{J,d}$, the standard deviation of Jaccard dissimilarity between regions, SD (J_r^R), the correlation between Shannon differentiation and geographical distance, $\rho_{\Delta D,d}$, and its standard deviation, $SD(\rho_{\Delta D,d})$, as well as the correlation between F_{ST} and geographical distance, $\rho_{Fst,d}$, contributed most to the power of discriminating the island model from other models (Figure S1). Besides the top 16 most important statistics, Jaccard dissimilarity between regions (J_r^R) and F_{ST} between populations (F_{ST}^P) also contributed substantially to the power of discriminating the stepping-stone and hierarchical stepping-stone models (Figure S1).

4 | DISCUSSION

In this study, we performed a comprehensive assessment of the discriminatory power of nine sets of summary statistics, comprising

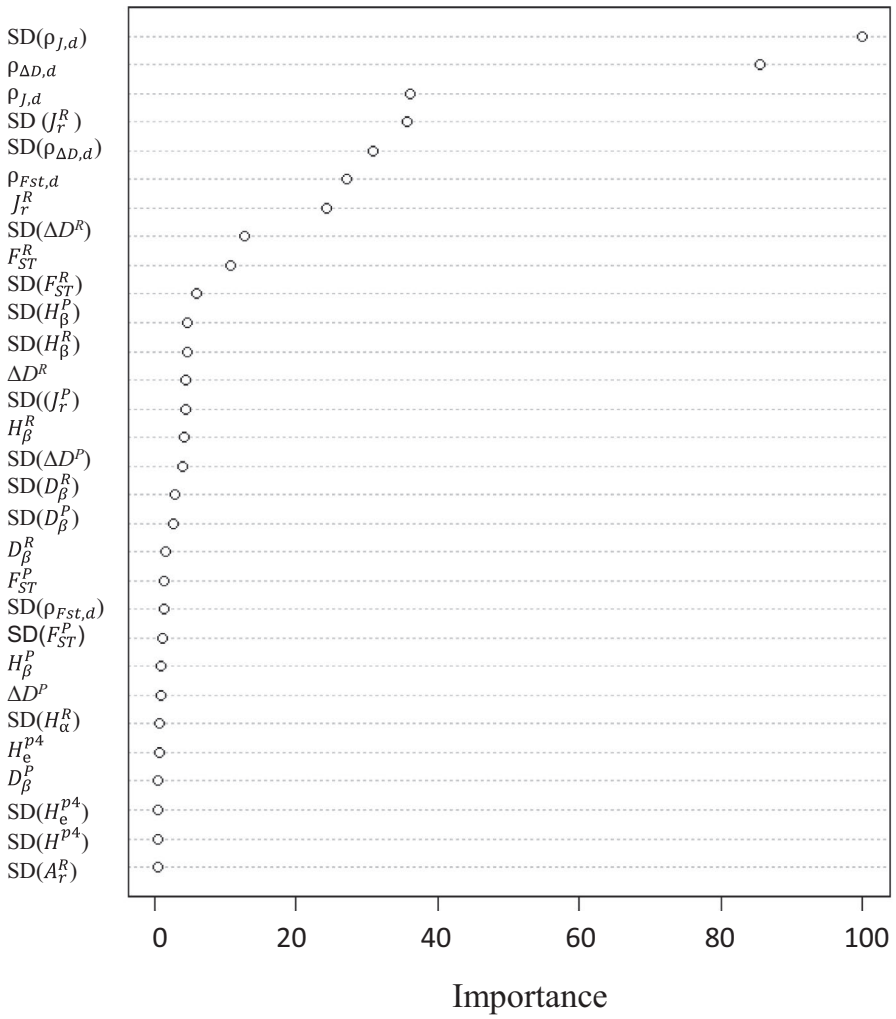

FIGURE 2 Ranked conditional variable importance for individual summary statistics estimated by conditional random forest classification. Results are shown only for the top 30 most important summary statistics (out of a total of 178). Statistical abbreviations are given in Table S1

^{Ar}SS, ^{H}SS, ^{He}SS and ^{1D}SS. Since different methods to estimate discriminatory power may lead to different results, we employed three powerful machine learning methods to compare the power of different sets of summary statistics. All results led to the same conclusion that ^{1D}SS performed the best among four sets of summary statistics in the discrimination of spatial scenarios. Although ^{1D}SS was overall the best set of diversity measures, ^{Ar}SS and ^{He}SS also provided complementary information that s did not capture.

Jaccard dissimilarity (*J*) and its Mantel's *r* ranked as the top summary statistics among all the summary statistics for differentiating spatial scenarios, followed by ΔD and then F_{ST} as well as their Mantel's *r*. In addition, we found that combining sets of summary statistics did not necessarily increase the discriminatory power (e.g., KLFDA and neural network models in Tables 2 and 4). Therefore, a more efficient strategy would be combining the most informative summary statistics in each set depending on the alternative spatial scenarios that could apply to each data set based on existing information.

During the past 20 years, evolutionary biologists and population geneticists have been using diversity metrics as the summary statistics to make inference on the evolutionary and demographic histories of populations via approximate Bayesian computation (ABC). Information theory offers a spectrum of summary statistics that can be used with ABC. However, the choice of summary statistics in population genetics has focused on the ^{He}SS family (i.e., heterozygosity, *He*, and fixation, F_{ST}). The use of ^{He}SS up-weights the signal provided by common alleles while down-weighting rare alleles, and thus it may miss important information under scenarios that involve bottlenecks or founder events. To avoid this problem, it is common to combine ^{Ar}SS and ^{He}SS, but our results indicate that the same or greater discriminatory power could be obtained using only the ^{1D}SS set. These results provide further support for the idea that simply increasing the number of summary statistics without considering their individual discriminatory power may decrease the inference accuracy.

Our systematic assessment of the power of these summary statistics showed that ^{He}SS, the most commonly used set of summary statistics in population genetics, performed worst in the discrimination of typical spatial scenarios tested by three different classification approaches. The differentiation measures, *J*, ΔD and F_{ST}, evaluate the extent of genetic differentiation between populations, with ΔD and F_{ST} being estimated based on allele frequency, and *J* being estimated based on allele presence/absence data. Usually, genetic differentiation is estimated using F_{ST} (Wright, 1949) and its

TABLE 4 Performance of different sets of summary statistics in discriminating five spatial scenarios using deep neural network

Summary statistics	^{Ar}SS	^{H}SS	^{He}SS	^{1_D}SS	$^{Ar+He}SS$	$^{H+^1D}SS$	$^{Ar+H+He}SS$	$^{Ar+He+^1D}SS$	$^{Ar+H+He+^1D}SS$
Accuracy (Acc)	0.988	0.988	0.974	**0.990**	0.986	0.986	0.988	**0.990**	0.986
Acc 95% CI	(0.974, 0.996)	(0.974, 0.996)	(0.956, 0.986)	(0.977, 0.997)	(0.971, 0.994)	(0.971, 0.994)	(0.974, 0.995)	(0.977, 0.997)	(0.971, 0.994)
Kappa	0.985	0.985	0.968	0.988	0.983	0.983	0.985	0.988	0.983

Note: Acc, accuracy; Acc 95% CI, the 95% interval of accuracy; Kappa, Cohen's kappa coefficient (κ). The best performance (highest Acc) is highlighted in bold.

variants (G_{ST}; Nei, 1973) calculated from heterozygosity (*He*) while *J* and ΔD, which are more informative according to our results, are rarely used as statistics to measure population genetic inference.

Our results indicate that ^{Ar}SS was very good at differentiating between panmixia and the other scenarios. ^{1_D}SS on the other hand, exhibits high accuracy in differentiating all scenarios, being especially good at discriminating between stepping-stone and hierarchical stepping-stone models. Therefore, ^{Ar}SS seems useful for detecting the scenarios that depart from panmixia, and ^{1_D}SS may be helpful to differentiate between more complex spatial scenarios. On the other hand, we did not observe advantageous properties for ^{He}SS in detecting the spatial structuring signals under the five spatial scenarios considered. Nevertheless, these results do not necessarily suggest that all differentiation measures of order $q = 2$ are problematic as we did not consider Jost's (2008) D_{EST} in this study because we wanted to focus on the most widely used measures of genetic differentiation, namely F_{ST}. More detailed studies would be needed to further explore the behaviour of the three families ($q = 0, 1, 2$) under more complex nonequilibrium scenarios, including population and range expansions as well as divergence and admixture events. Recent population genetic simulators such as SliM3 (Haller & Messer, 2019) would prove very useful in this context.

For a long time, important guidelines for species and genetic diversity conservation have been made using richness and Simpson's index for species diversity (Jost et al., 2010; Scott et al., 1987), and heterozygosity (derived from *F*-statistics framework) for genetic diversity (Aitken et al., 2012). The results of this study suggest that summary statistics based on Hill numbers are promising tools for detecting diversity changes in biological conservation studies.

In this study, we considered microsatellite markers instead of single nucleotide polymorphisms (SNPs) because they are multi-allelic and, therefore, the information content of a single microsatellite locus is much higher than that of a single SNP locus. The fact that SNPs are bi-allelic markers limits the range of values expected when using Hill numbers and makes them less sensitive to capture genetic signatures left by evolutionary processes. We note, however, that it is possible to generate multi-allelic markers from SNPs by focusing on chromosome windows containing two or more SNPs. Moreover, RADSeq loci containing more than one SNP with different patterns of segregation represents a readily available multi-allelic marker which is being used mainly by researchers working with nonmodel species, many of which are of conservation concern. We think this would be the best approach when using Hill numbers to study spatial genetic structure and the results of our study would be relevant in this type of application.

In summary, diversity of order $q = 1$ (^1D) and Shannon differentiation offer a unified approach integrating diversity across all levels of biological organizations. Our results suggest that ^{1_D}SS would perform well for the purpose of inference of population structure using inferential frameworks such as ABC. It is clear that no single set of diversity measures can capture all the information contained in raw population genetics data sets, and our study suggests that the type of summary statistics we may want to use depends on the specific question being asked.

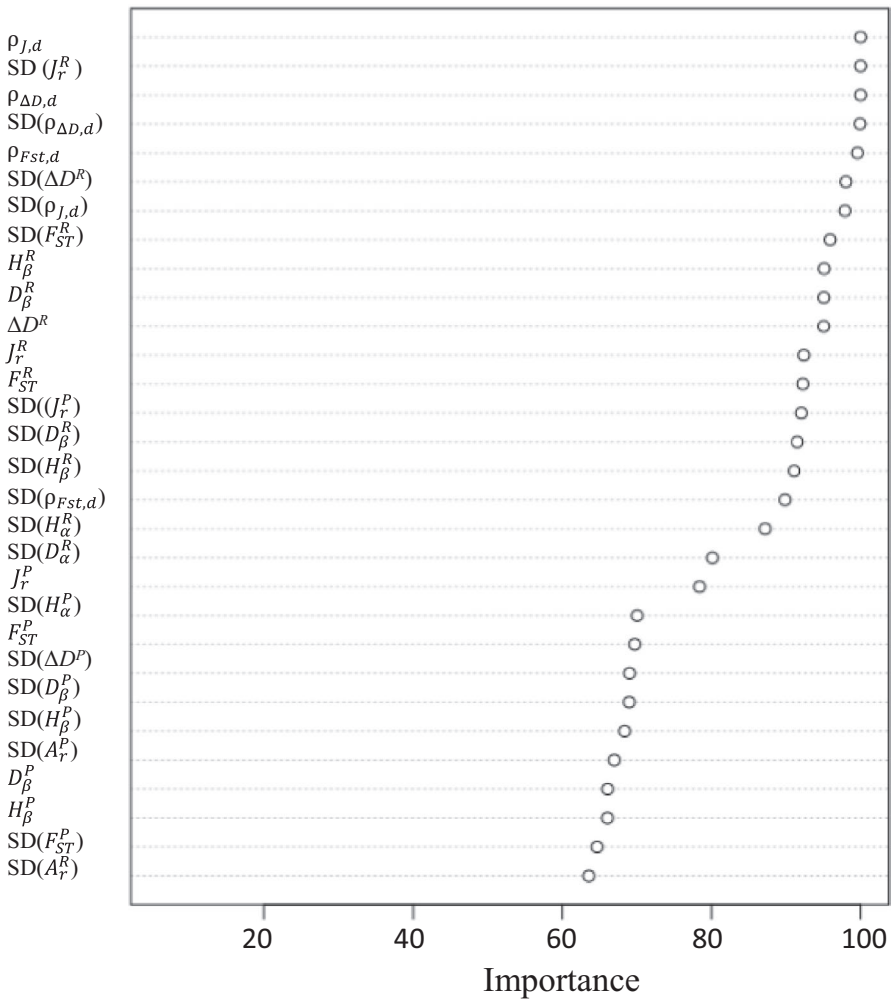

FIGURE 3 Variable importance estimated by the deep neural network. Results are shown only for the top 30 ranked summary statistics out of a total of 178. Statistical abbreviations are given in Table S1

Finally, we found different machine learning methods showed different performance in distinguishing spatial structure scenarios. KLFDA gave the lowest discriminatory accuracy while the deep neural network gave the highest discriminant accuracy among the three classification methods (Tables 2–4). In contrast, conditional random forest did not show the difference between the power of ^{H}SS and ^{1D}SS as well as other combinations of summary statistics (Table 3). The deep neural network showed more advantages than KLFDA and conditional random forest in this study, which provides additional support to recent assertions that machine learning methods represent promising tools to carry out inference in ecology and evolution (Schrider & Kern, 2018).

ACKNOWLEDGMENTS

X.H.Q. was supported by the China Scholarship Council.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interests.

AUTHOR CONTRIBUTIONS

X.Q. and O.E.G. designed the study. X.Q. carried out the analyses and interpreted results with the input from O.E.G. X.Q. wrote the manuscript with the input of O.E.G. Both authors contributed to editing and revising the manuscript.

BENEFIT SHARING

Benefits from this research accrue from the sharing of our data and results on public databases as described above.

DATA AVAILABILITY STATEMENT

The input files and scripts for generating simulation as well as the analyses of summary statistics are available at https://zenodo.org/record/6197846#.Yhdq6-hBxPY (Qin & Gaggiotti, 2022). The scripts for calculating different summary statistics have been wrapped in the R package *HierDpart* (Qin, 2019) available at https://cran.r-project.org/web/packages/HierDpart/index.html.

ORCID

Xinghu Qin https://orcid.org/0000-0003-2351-3610

REFERENCES

Aitken, S. N., Luikart, G., & Allendorf, F. W. (2012). *Conservation and the genetics of populations*. John Wiley & Sons.

Alvarado-Serrano, D. F., & Hickerson, M. J. (2016). Spatially explicit summary statistics for historical population genetic inference. *Methods in Ecology and Evolution*, *7*, 418–427. https://doi.org/10.1111/2041-210X.12489

Baum, E. B. (1988). On the capabilities of multilayer perceptrons. *Journal of Complexity*, *4*, 193–215. https://doi.org/10.1016/0885-064X(88)90020-9

Chao, A., Chiu, C.-H., & Jost, L. (2014). Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through Hill numbers. *Annual Review of Ecology, Evolution, and Systematics*, *45*, 297–324. https://doi.org/10.1146/annurev-ecolsys-120213-091540

Csilléry, K., Blum, M. G., Gaggiotti, O. E., & François, O. (2010). Approximate Bayesian computation (ABC) in practice. *Trends in Ecology & Evolution*, *25*, 410–418. https://doi.org/10.1016/j.tree.2010.04.001

Day, T. (2015). Information entropy as a measure of genetic diversity and evolvability in colonization. *Molecular Ecology*, *24*, 2073–2083. https://doi.org/10.1111/mec.13082

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, *9*, e1003905. https://doi.org/10.1371/journal.pgen.1003905

Excoffier, L., & Foll, M. (2011). Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*, *27*, 1332–1334. https://doi.org/10.1093/bioinformatics/btr124

Fortuna, M. A., Albaladejo, R. G., Fernández, L., Aparicio, A., & Bascompte, J. (2009). Networks of spatial genetic variation across species. *Proceedings of the National Academy of Sciences*, *106*, 19044–19049. https://doi.org/10.1073/pnas.0907704106

Gaggiotti, O. E., Chao, A., Peres-Neto, P., Chiu, C.-H., Edwards, C., Fortin, M.-J., Jost, L., Richards, C. M., & Selkoe, K. A. (2018). Diversity from genes to ecosystems: A unifying framework to study variation across biological metrics and scales. *Evolutionary Applications*, *11*(7), 1176–1193. https://doi.org/10.1111/eva.12593

Garson, D. G. (1991). Interpreting neural network connection weights. *Artificial Intelligence Expert*, *6*, 46–51.

Gevrey, M., Dimopoulos, I., & Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling*, *160*, 249–264. https://doi.org/10.1016/S0304-3800(02)00257-0

Guarnieri, S., Piazza, F., & Uncini, A. (1999). Multilayer feedforward networks with adaptive spline activation function. *IEEE Transactions on Neural Networks*, *10*, 672–683. https://doi.org/10.1109/72.761726

Haller, B. C., & Messer, P. W. (2019). SLiM 3: Forward genetic simulations beyond the Wright-Fisher model. *Molecular Biology and Evolution*, *36*, 632–637. https://doi.org/10.1093/molbev/msy228

Hampe, J., Schreiber, S., & Krawczak, M. (2003). Entropy-based SNP selection for genetic association studies. *Human Genetics*, *114*, 36–43. https://doi.org/10.1007/s00439-003-1017-2

Hill, M. O. (1973). Diversity and evenness: A unifying notation and its consequences. *Ecology*, *54*, 427–432. https://doi.org/10.2307/1934352

Hothorn, T., Hornik, K., Strobl, C., & Zeileis, A. (2010). *Party: A laboratory for recursive partytioning*.

Jaccard, P. (1912). The distribution of the flora in the alpine zone. 1. *New Phytologist*, *11*, 37–50. https://doi.org/10.1111/j.1469-8137.1912.tb05611.x

Jost, L. (2006). Entropy and diversity. *Oikos*, *113*, 363–375.

Jost, L. (2007). Partitioning diversity into independent alpha and beta components. *Ecology*, *88*, 2427–2439. https://doi.org/10.1890/06-1736.1

Jost, L. (2008). GST and its relatives do not measure differentiation. *Molecular Ecology*, *17*, 4015–4026.

Jost, L., Chao, A., & Chazdon, R. L. (2011). Compositional similarity and β (beta) diversity. *Biological diversity: frontiers in measurement and assessment*, 66–84. Oxford University Press.

Jost, L., DeVries, P., Walla, T., Greeney, H., Chao, A., & Ricotta, C. (2010). Partitioning diversity for conservation analyses. *Diversity and Distributions*, *16*, 65–76. https://doi.org/10.1111/j.1472-4642.2009.00626.x

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*, 1–26.

Kuhn, M. (2012). *The caret package*. R Foundation for Statistical Computing, . https://cran.r-project.org/package=caret

Kuhn, M. (2015). *Caret: classification and regression training*. Astrophysics Source Code Library.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modelling*. Springer.

Lankau, R. A., & Strauss, S. Y. (2007). Mutual feedbacks maintain both genetic and species diversity in a plant community. *Science*, *317*, 1561–1563. https://doi.org/10.1126/science.1147455

Luiselli, J., Overcast, I., Rominger, A., Ruffley, M., Morlon, H., & Rosindell, J. (2021). *Detecting the ecological footprint of selection. bioRxiv*.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, *70*, 3321–3323. https://doi.org/10.1073/pnas.70.12.3321

Nekola, J. C., & White, P. S. (1999). The distance decay of similarity in biogeography and ecology. *Journal of Biogeography*, *26*, 867–878. https://doi.org/10.1046/j.1365-2699.1999.00305.x

Novembre, J., & Stephens, M. (2008). Interpreting principal component analyses of spatial population genetic variation. *Nature Genetics*, *40*, 646–649. https://doi.org/10.1038/ng.139

Overcast, I., Emerson, B. C., & Hickerson, M. J. (2019). An integrated model of population genetics and community ecology. *Journal of Biogeography*, *46*, 816–829. https://doi.org/10.1111/jbi.13541

Overcast, I., Ruffley, M., Rosindell, J., Harmon, L., Borges, P. A. V., Emerson, B. C., Etienne, R. S., Gillespie, R., Krehenwinkel, H., Mahler, D. L., Massol, F., Parent, C. E., Patiño, J., Peter, B., Week, B., Wagner, C., Hickerson, M. J., & Rominger, A. (2021). A unified model of species abundance, genetic diversity, and functional diversity reveals the mechanisms structuring ecological communities. *Molecular Ecology Resources*, *21*, 2782–2800. https://doi.org/10.1111/1755-0998.13514

Qin, X. (2019). *HierDpart*, *v0*, 3.5.

Qin, X., & Gaggiotti, O. E. (2022). *Information-based summary statistics for spatial genetic structure inference (ed. Zenodo)*. Zenodo.

Qin, X., Lock, T. R., & Kallenbach, R. L. (2021). DA: Population structure inference using discriminant analysis. *Methods in Ecology and Evolution*, *13*(2), 485–499. https://doi.org/10.1111/2041-210X.13748

Ricotta, C. (2005). On hierarchical diversity decomposition. *Journal of Vegetation Science*, *16*, 223–226. https://doi.org/10.1111/j.1654-1103.2005.tb02359.x

Ripley, B. D., & Hjort, N. (1996). *Pattern recognition and neural networks*. Cambridge University Press.

Schrider, D. R., & Kern, A. D. (2018). Supervised machine learning for population genetics: A new paradigm. *Trends in Genetics*, *34*, 301–312. https://doi.org/10.1016/j.tig.2017.12.005

Scott, J. M., Csuti, B., Jacobi, J. D., & Estes, J. E. (1987). Species richness. *BioScience*, *37*, 782–788. https://doi.org/10.2307/1310544

Sherwin, W. B., Chao, A., Jost, L., & Smouse, P. E. (2017). Information theory broadens the spectrum of molecular ecology and evolution. *Trends in Ecology & Evolution*, *32*, 948–963. https://doi.org/10.1016/j.tree.2017.09.012

Stotz, G. C., Gianoli, E., & Cahill, J. F. (2016). Spatial pattern of invasion and the evolutionary responses of native plant species. *Evolutionary Applications*, *9*, 939–951. https://doi.org/10.1111/eva.12398

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*, 307. https://doi.org/10.1186/1471-2105-9-307

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*, 25. https://doi.org/10.1186/1471-2105-8-25

Sugiyama, M. (2007). Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *Journal of Machine Learning Research*, *8*, 1027–1061.

Tang, Y., & Li, W. (2016). *Lfda: An R package for local fisher discriminant analysis and visualization*.

Tang, Y., & Li, W. (2017). *lfda: Local fisher discriminant analysis in R*.

Van Tienderen, P. H. (1991). Evolution of generalists and specialist in spatially heterogeneous environments. *Evolution*, *45*(6), 1317–1331. https://doi.org/10.1111/j.1558-5646.1991.tb02638.x

Vellend, M. (2005). Species diversity and genetic diversity: Parallel processes and correlated patterns. *The American Naturalist*, *166*, 199–215. https://doi.org/10.1086/431318

Wagner, A. (2017). Information theory, evolutionary innovations and evolvability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*, 20160416. https://doi.org/10.1098/rstb.2016.0416

Wang, X., Wiegand, T., Wolf, A., Howe, R., Davies, S. J., & Hao, Z. (2011). Spatial patterns of tree species richness in two temperate forests. *Journal of Ecology*, *99*, 1382–1393. https://doi.org/10.1111/j.1365-2745.2011.01857.x

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*, 1358–1370.

Whitlock, M. C. (2011). G'ST and D do not replace FST. *Molecular Ecology*, *20*, 1083–1091. https://doi.org/10.1111/j.1365-294X.2010.04996.x

Wright, S. (1949). The genetical structure of populations. *Annals of Eugenics*, *15*, 323–354. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x

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: Qin, X., & Gaggiotti, O. E. (2022). Information-based summary statistics for spatial genetic structure inference. *Molecular Ecology Resources*, *00*, 1–13. https://doi.org/10.1111/1755-0998.13606